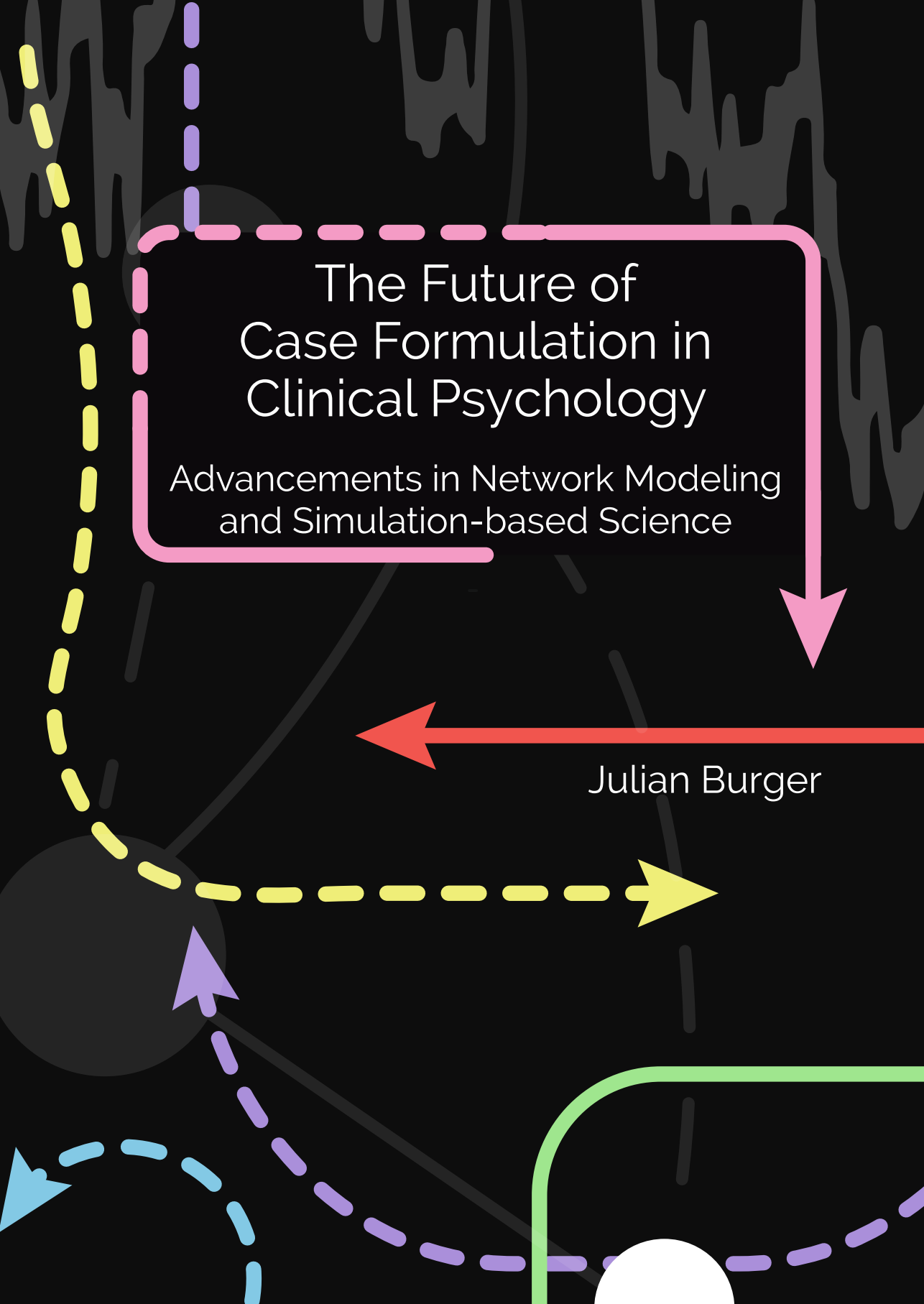

The Future of
Case Formulation in
Clinical Psychology
Advancements in Network Modeling
and Simulation-based Science
Julian Burger

The Future of Case Formulation in Clinical Psychology

Advancements in Network Modeling and Simulation-based Science

—

Colofon

The Future of Case Formulation in Clinical Psychology

Ph.D. thesis Julian Burger – University of Groningen

Layout & Cover design: Robin Weijland
Printing: Printing: Ridderprint, www.ridderprint.nl

university of groningen

The Future of Case Formulation in Clinical Psychology

Advancements in Network Modeling and Simulation-based Science

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. J.M.A. Scherpen
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Monday 11 December 2023 at 11.00 hours

by

Julian Burger

born on 29 November 1992
in Friedrichshafen, Germany

Supervisors

Dr. H. Riese
Prof. R.A. Schoevers

Co-supervisors

Dr. S. Epskamp
Dr. D.C. van der Veen

Assessment Committee

Prof. A.T.M. Jansen, Maastricht University
Prof. C.J. Albers, University of Groningen
Prof. D.C. Cath, University of Groningen

Contents

1	**General Introduction**	**9**
1.1	One size fits few: Issues with manualized treatment	10
1.2	The past and present of case formulation	12
1.3	The future of case formulation: Gaps in research and aims of this thesis	13
1.4	Chapter outline	14

Part I: Methodological Background

2	**Longitudinal Data and Research Design Choices**	**21**
2.1	Introduction	23
2.2	Data designs	23
2.3	Analysis designs	24
2.4	Differences between data and analysis	26
2.5	Separating contemporaneous and temporal effects	30
2.6	Conclusion	32
3	**Network Estimation from Time Series and Panel Data**	**35**
3.1	Introduction	37
3.2	Graphical vector auto-regression	37
3.3	$N = 1$ estimation: Personalized network models	39
3.4	$N > 1$ estimation: Multi-level estimation	43
3.5	Challenges to GVAR estimation	49
3.6	Conclusion	54
4	**Using Network Analysis to Evaluate Mental Health Interventions**	**57**
4.1	Introduction	59
4.2	Methods	60
4.3	Results	61
4.4	Discussion	74
5	**Reporting Standards for Psychological Network Analysis in Cross-sectional Data**	**81**
5.1	Introduction	83
5.2	Reporting standards for the 'Methods' section	86
5.3	Reporting standards for the 'Results' section	91
5.4	Illustrative examples	97
5.5	Conclusion	102

Part II: *Exploration* – Statistical Networks Based on Empirical Data

6	**Network Analysis of Anxiety Symptoms During the COVID-19 Pandemic**	**107**
6.1	Introduction	109
6.2	Methods	111
6.3	Results	114
6.4	Discussion	118
7	**Network Analysis of Depression Symptoms During the COVID-19 Pandemic**	**123**
7.1	Introduction	125
7.2	Methods	127
7.3	Results	135
7.4	Discussion	139
7.5	Conclusions	146
8	**Network Analysis of Depression Symptoms as a Consequence of Spousal Loss and Separation**	**149**
8.1	Introduction	151
8.2	Methods	153
7.3	Results	157
8.4	Discussion	161
8.5	Conclusions	163

Part III: *Integration* – Combining Clinical Prior Information with Statistical Networks

9	**PREMISE: The Prior Elicitation Module for Idiographic System Estimation**	**167**
9.1	Introduction	169
9.2	A formal integration of case formulation and personalized networks	171
9.3	The Prior Elicitation Module for Idiographic System Estimation (PREMISE)	173
9.4	Clinical example: Client with obsessive-compulsive disorder	174
9.5	Discussion	177
9.6	Conclusion	181
10	**Combining Case Formulations and Longitudinal Data to Estimate Networks for Eating Disorders**	**183**
10.1	Introduction	185
10.2	Methods	187
10.3	Results	196
10.4	Discussion	202
10.5	Conclusion	205

11 Longitudinal Perceived Causal Relation Networks 207

11.1 Introduction 209
11.2 Methods 210
11.3 Results 216
11.4 Discussion 221
11.5 Conclusions 225

Part IV: *Formalization* – Computational Models of Case Formulations

12 Formalizing Case Formulations 229

12.1 Introduction 231
12.2 Methods 236
12.3 Results 239
12.4 Discussion 245
12.5 Conclusion 250

Part V: Conclusion

13 General Discussion 255

13.1 Thesis summary and answers to research questions 256
13.2 Future research 263
13.3 Conclusion 266

Supplements
Supplement A 270
Supplement B 286
Supplement C 292
Supplement D 298
Supplement E 304

References 309

Appendices
Summary 341
Nederlandse Samenvatting 347
Publications 353
Acknowledgements 357

CHAPTER 1

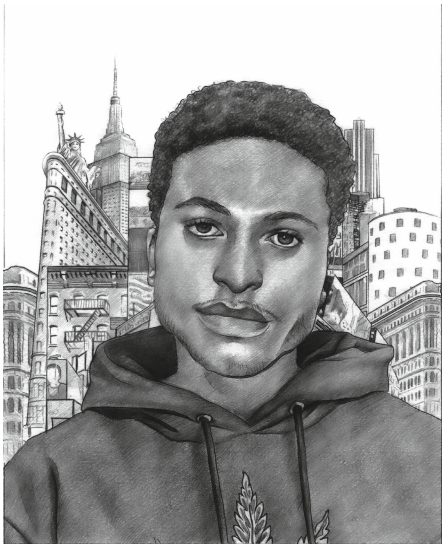
GENERAL INTRODUCTION

At some point in their life, about one in two people will develop psychological problems that are considered a mental disorder (Kessler et al., 2005) as defined in the diagnostic and statistical manual for mental disorders (American Psychiatric Association, 2013). This is a staggering number, seen as mental disorders can be debilitating, sometimes life-threatening conditions, with long lasting consequences for individuals, the direct environment of the individual, and connected to costs for society as a whole (World Health Organization, 2023). To address these problems, clinicians are developing treatments that improve the lives of individuals who suffer from mental disorders.

In this context, there is a long tradition of research that identifies recurring patterns of psychological symptoms, and arrange these patterns in a *diagnostic classification system*. For example, individuals who experience a clinically significant loss of interest or pleasure in their daily activities or who experience depressed mood, amongst other potential problems such as feelings of worthlessness, concentration problems, and recurrent thoughts of death, would be ascribed the diagnosis of *depression*. An important function of such diagnostic classifications is that they can be connected to evidence-based treatments, explicated in so-called treatment manuals (American Psychiatric Association, 2013). In cognitive-behavioral therapy (CBT), these manuals provide clinicians with a highly structured approach to therapy. For example, there are treatment manuals for depression that aim to change a person's negative thoughts about themselves and the negative evaluation they perceive from their environment (cognitive therapy for depression; Beck, 1979), and to encourage the person to engage in activities that help them make rewarding experiences (behavioral activation for depression; Barlow, 2021).

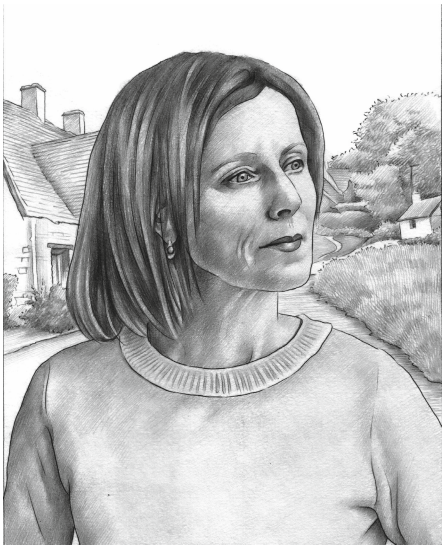
1.1 One size fits few: Issues with manualized treatment

The practice of CBT has been traditionally confined to manualized treatments. Over the past decades, however, clinicians have challenged this traditional approach. One reason for this is that not all clients respond to protocols. For example, a meta-analysis showed that only about 50% of clients with diagnosed anxiety disorder respond to CBT treatment (Loerinc et al., 2015) although anxiety disorders are among the diagnoses that have the strongest support for CBT generally (Hofmann et al., 2012). A possible explanation for this finding is that the general processes that are described in the diagnostic protocols do not necessarily match the psychological processes observed in *specific* clients (Persons, 2012). This mismatch stems from the problem that people with the same diagnosis usually have very different experiences of psychological symptoms, a phenomenon referred to as *heterogeneity*. Individuals may have different symptom profiles that are summarized under the same umbrella diagnosis, they may experience different intensities of shared symptoms, and their symptoms may have developed for very different reasons, and interact differently with one another. Consider the two hypothetical examples of Nathan and Maria below. Both of them have been diagnosed with Major Depressive Disorder (MDD), however, the psychological symptoms, as well as their origins and maintaining factors are very different from one another, and are therefore not captured very well by one general explanation of depression.



Nathan from New York, US[1]

Nathan is a 24-year-old gay Black man living in New York City. He has experienced discrimination and rejection from both his family and society. Even in the queer community, Nathan has struggled to find support and acceptance, and often experiences overt racism. Over the past few years, Nathan has felt increasingly sad, hopeless, and worthless. He cannot find pleasure in things he used to be passionate about, and more and more withdraws himself from social activities. In addition, Nathan engages in risky sexual behaviors, such as condomless sex, often under the influence of alcohol and marijuana. Nathan feels a lot of shame about these behaviors. He is also wary of the stigma that comes with seeing a therapist, and is therefore hesitant to seek help.



Maria from near Portsmouth, UK

Maria is a 42-year-old white woman living in a rural area in the south of England. She is a single mother of two children, and has to provide for her family on her own. She is also the primary caregiver of her elderly mother, who has been diagnosed with Parkinson in the previous year. Most of the time, Maria is overwhelmed with taking care of her children and her mother, next to working full time. She barely has any time for herself, and does not have many friends. The isolation and lack of social support makes Maria feel very hopeless, and she worries about the future which frequently keeps her from sleeping. Maria has been admitted to a clinic two times in the past due to suicidal thoughts, and is now seeing a therapist in an outpatient practice.

1 Sketches drawn by Rick: https://www.fiverr.com/freelancers/artist_canvas?source=order_page_summary_seller_link

Another problem of treatments based on manuals is that corresponding diagnoses often co-occur, referred to as the *comorbidity* of mental disorders. Comorbidity between diagnoses is the rule rather than the exception. A prime example is depression which often co-occurs with anxiety (Judd et al., 1998; Starr & Davila, 2012). Another example is illustrated in the case descriptions above as Nathan also received a diagnosis for substance use disorder. In addition to the challenge of treating multiple disorders, there are also specific problems that are not referenced in treatment manuals. For example, only recently did clinicians pay attention to developing manuals targeting problems connected to minority stress (Pachankis, 2015), which Nathan would likely benefit from. Finally, CBT manuals have clearly outlined treatment methods and goals. In practice, these methods and goals do not always align with the client's preferred outcomes and approaches, potentially resulting in non-adherence to the protocol and homework (Persons, 2012). For example, while Maria's preferred outcomes likely align with the goals of routine CBT for depression, it may not be Nathan's primary goal to become more socially proactive, as long as he is exposed to racist or homophobic stereotypes in his communities.

1.2 The past and present of case formulation

Treatment manuals offer a certain degree of flexibility in capturing the client-specific context. For example, they assess and integrate the client's specific situation, and the cognitive and behavioral responses these elicit (Barlow, 2021). However, some clinicians called for even more flexibility in acknowledging the different ways in which biological, psychological, and social factors of one particular person interact with one another (Hayes et al., 2019; Hofmann & Hayes, 2019; Schiepek, 2003). One of the first frameworks to acknowledge these personalized processes is the *case formulation* approach (Eells, 2022; Kuyken et al., 2009; Persons, 2012). A case formulation is a comprehensive personalized explanation of the problems a specific individual is experiencing. It describes the individual's problems and resources, as well as a hypothesis regarding the origin and maintenance of the symptoms. One of the main goals of a case formulation is to guide personalized treatment. In contrast to the manualized approach, these treatments are based on the specific formulation, and therefore may be different for individuals with the same diagnosis. For example, Nathan's treatment plan based on the case formulation approach may focus specifically on interventions that relate to his experiences of minority stress, whereas Maria's treatment plan may rather focus on more general behavioral and cognitive interventions.

Constructing case formulations is a challenging endeavor, requiring collaboration with clients and other clinicians, and using diverse sources of information. Currently, there are no gold standards for constructing case formulations, and clinicians may differ in the specific approach they take to arrive at a formulation (Riese et al., 2021). Given these challenges, valid assessment and data-driven inferences are key components in advancing case formulations. Ever since its introduction, the case formulation approach has developed in parallel with technological and methodological advances. These advances have equipped clinicians with a series of assessment tools and personalized models (Wright & Woods, 2020) that can help clinicians constructing case formulations. Three developments stand out in this context:

First, it is possible to collect data on personalized items via smartphones, using data collected via the *Experience Sampling Method* (Mestdagh & Dejonckheere, 2021; Myin-Germeys et al., 2018; Shiffman et al., 2008). ESM data have the potential to provide relevant insights for case formulation, because they are collected in the real-life context of the client, capturing experiences and contexts related to mental health if and when they occur. During a diagnostic session, both Nathan and Maria may find it challenging, for example, to recollect all relevant experiences from several weeks or even months ago. In contrast, ESM assessment allows to embed the specific psychological experiences in their respective (social) context, reducing recall bias and illustrating patterns that may be relevant for case formulation (Riese et al., 2021). For example, ESM data could illustrate that Nathan has feelings of shame specifically in the context of impulsive sexual behavior, and Maria worries a lot particularly in the evenings, keeping her from sleeping. Another particular strength of ESM is that items can be personalized such that they represent the specific thoughts and feelings of a client, such as the specific shame-related thoughts Nathan has, or the specific types of events Maria is worried about.

Second, new statistical techniques have been implemented that identify relationships between personalized items as *statistical network models* (Epskamp, Van Borkulo, et al., 2018; Epskamp, Waldorp, et al., 2018). Personalized networks can be estimated from ESM data, and they align conceptually well with the way clinician think and reason in constructing case formulations (Zuidersma et al., 2020). As such, they can be used to guide conversations between clinician and client, and to explore specific personalized dynamics that may be relevant for case formulation (von Klipstein et al., 2020). Nathan's personalized network, for example, may include a relationship between the ESM items 'shame' and 'impulsive sexual behavior'. Maria's personalized network may include a relationship between the ESM items 'worry' and sleep-related symptoms such as 'concentration problems'.

Third, there has been increasing interest in building computational models for theories that represent them as systems of mathematical equations, sometimes referred to as *formal theories* (Borsboom, van der Maas, et al., 2021; Fried, 2020; Guest & Martin, 2021; Haslbeck, Ryan, et al., 2021; Robinaugh et al., 2021). Formal theories are appealing for clinical psychology, because they can be used to conduct computer simulations that illustrate precisely what effects can be expected from any given (formalized) clinical intervention. In the context of case formulations, formalization opens doors for creating more accurate explanations, as well as more precise treatment planning. For example, there is a set of evidence-based treatments that Nathan and Maria could benefit from, and formalized case formulations could allow their clinicians to explore the impact that can be expected from these interventions, and to base decision-making on these simulated outcomes.

1.3 The future of case formulation: Gaps in research and aims of this thesis

The developments discussed above promise to move case formulations to a new era of personalized therapy, one that is informed by novel assessment strategies, sophisticated statistical models, and simulation-based science. Despite this promise, these tools have not yet found their way into routine clinical practice. There are many factors that hamper implementation efforts, such as a mismatch between data characteristics and model requirements, a disconnect between models and clinical

reasoning, or the fact that some of these approaches are simply not developed enough for producing valid inferences.

The main aim of this thesis is to contribute to bridging the gap between personalized modeling approaches and the clinical practice of case formulation. More specifically, I discuss how the use of different types of models depends on the particular strength of the respective approach, and how clinicians can match these strengths with specific aspects of case formulation. Part I of the thesis will discuss general aspects of statistical models in this context, such as model estimation and reporting. The following three parts are the main body of the thesis (Part II–IV), and will focus on three ways in which different types of models can inform case formulations: *Exploration* – using statistical networks to generate insights given contextual factors, such as loss experiences; *integration* – using initial case formulations as prior information for estimating personalized networks, therefore systematically combining case formulations with statistical networks; and finally, *formalization* – using case formulations to construct a formalized theory from the ground up, allowing for computer simulations on expected system behavior and effects of evidence-based interventions. In Part V, I discuss specific implementation barriers for each of these pathways, and reflect on their clinical utility.

1.4 Chapter outline

1.4.1 Part I: Methodological background

The first part will set the stage for the remaining sections by introducing statistical basics, methodological considerations, and current practices of network analysis. The chapters within this first part are organized in the chronological order of the standard research process: Starting with the research design and data collection, moving to the statistical estimation, using statistical inferences to answer specific research questions, to writing up results in a scientific report.

Chapter 2 discusses the fundamentals of *research designs* in the context of longitudinal network analysis, a statistical model that zooms in on temporal and contemporaneous relationships between items in time series and panel data. As is generally the case in empirical research, the research design constrains the types of questions one can answer in a study. This is no different for research using longitudinal data to estimate statistical network models. It is therefore crucial to map the specific research interests onto the appropriate design, and chapter 2 therefore provides a roadmap for navigating through the data, design, and analysis landscape in the context of longitudinal network analysis. **Chapter 3** discusses statistical estimation, and provide a technical account of network analysis using ecological momentary assessment and panel data. This chapter covers estimation techniques for the designs outlined in chapter 2. The discussed techniques are the basis for the wider literature on network estimation of longitudinal data as well as for most empirical chapters of this thesis (chapter 6 and 7). **Chapter 4** then zooms in on one specific research aim that is common for many clinical network contributions: Evaluating the effect of clinical interventions. The chapter provides a review of current practices, and compares different analysis choices to evaluate treatment. **Chapter 5** wraps up the methodological notes by introducing reporting standards for psychological networks. These guidelines are primarily developed for reporting cross-sectional networks, but can to a large extent also be used for the reporting of longitudinal network analysis. The goal of this

chapter is to guide researchers in communicating their results in a scientific report by identifying the elements of network analysis that are relevant in the context of good scientific practice.

1.4.2 Part II: Exploration – Statistical networks based on empirical data

The second part consists of a collection of empirical network contributions that investigate distinct contextual factors in relation to anxiety disorders and depression. This part will focus on the research question: "*How can statistical networks be used for exploration of symptom relationships, providing supporting insights for case formulations?*".

Chapter 6 highlights findings from a large longitudinal study (N = 1,706; 40 daily assessments) on anxiety symptoms in the context of the COVID-19 pandemic. **Chapter 7** uses data from the same population to analyze relationships between depressive symptoms. Both chapters provide insights that can be relevant for case formulations in the specific context of the COVID-19 pandemic. **Chapter 8** investigates the specific context of spousal loss and separation experiences in relation to depressive symptoms in a cross-sectional dataset. These empirical network contributions can offer clinicians relevant building blocks for constructing case formulations with clients who have developed anxiety and depression symptoms during the COVID-19 pandemic, or in response to spousal loss or separation from their partner. As mentioned above, the generated insights in **chapters 6-8** are purely exploratory, and stem from different types of models (longitudinal and cross-sectional networks). Models estimated in these studies do not draw on any particular clinical theory, prior empirical findings, or case specific knowledge.

1.4.3 Part III: Integration – Prior information in statistical networks

The third part moves beyond using network analysis to generate exploratory insights, and instead focuses on the *integration* of statistical networks and case formulations. This part will focus on the research question: "*How can statistical networks systematically be combined and integrated with the case formulation approach?*".

Chapter 9 introduces the Prior Elicitation Module for Idiographic System Estimation (PREMISE). The PREMISE approach centers the estimation of personalized networks around an initial case formulation established by clinician and client. In this chapter, I highlight how PREMISE may overcome several of the current barriers in personalized network analysis, such as low accuracy of parameter estimates, and the lack of opportunities to incorporate clinically relevant information into network structures. **Chapter 10** builds on the conceptual introduction of PREMISE in the previous chapter by introducing an empirical example of two clients suffering from Anorexia Nervosa. This chapter illustrates differences in networks across the two clients, as well as differences between the PREMISE and the traditional approach. In **chapter 11**, I introduce a second approach to constructing personalized networks based on daily assessments of the client's perceived relations between symptoms, the longitudinal perceived causal relations (L-PCR) approach. The L-PCR approach provides rich psychometric data for answering clinically relevant research questions on the stability of networks over time.

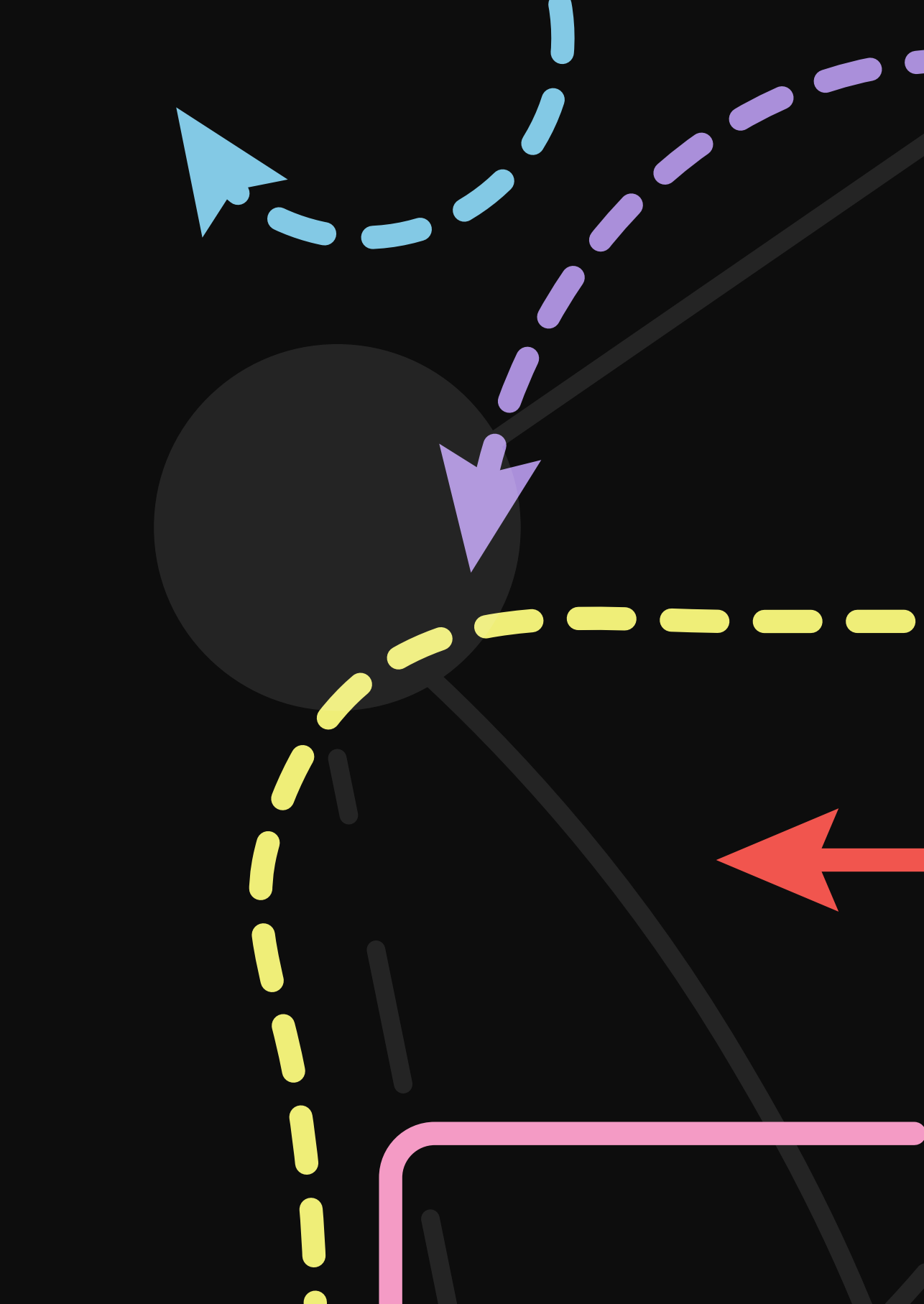
1.4.4 Part IV: Formalization – Computational models of case formulations

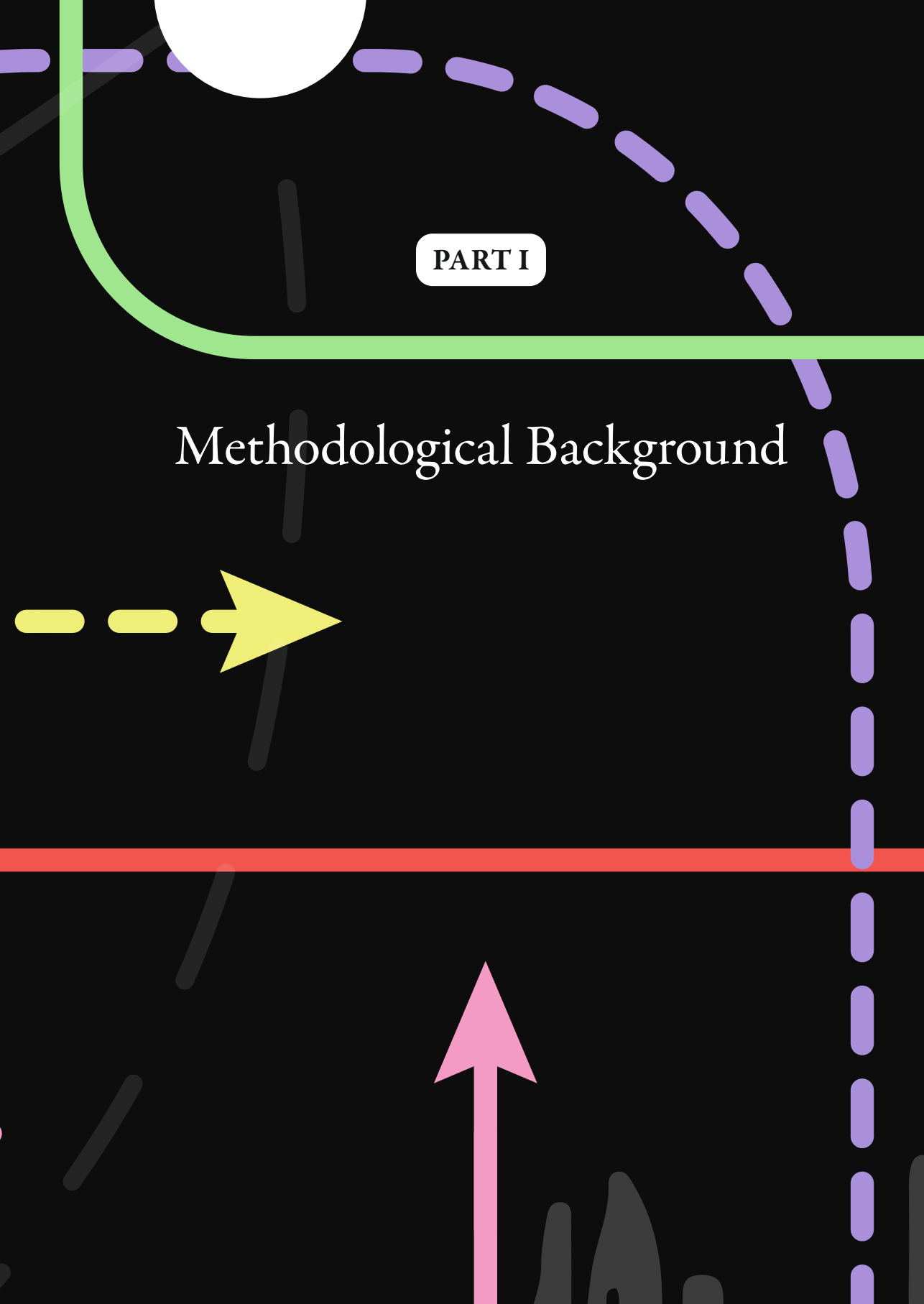
The fourth part of this thesis focuses on mathematical models and computer simulations, the *formalization* of case formulations. This part will focus on the research question: "*How can case formulations be advanced as idiographic theories and computational models using simulations?*".

Chapter 12 introduces this approach, and provides an example of a formalized case formulation for a client diagnosed with panic disorder. In this example, I illustrate the utility of formalization for case formulations, and discuss open questions for future research, such as considerations impacting the feasibility of implementing formalized case formulations in clinical practice.

1.4.5 Part V: Conclusion

In the final part of this thesis, **chapter 13**, I integrate and discuss the perspectives outlined in the previous chapters in regard to the specific types of questions that can be addressed to advance case formulations. I summarize the main findings of thesis, and sketch a roadmap for future research for using personalized models to inform case formulations.




PART I
Methodological Background

CHAPTER 2

LONGITUDINAL DATA AND RESEARCH DESIGN CHOICES

CHAPTER 2

Abstract

This chapter discusses challenges that emerge when repeated measures are introduced to the sampling plan. We distinguish between four types of data: single measurement data (one measure per person), panel data (many people measured a few times), $N = 1$ time series data (one person measured many times), and $N > 1$ time series data (several people measured many times). We also distinguish between four different types of analysis: cross-sectional analysis, between-persons analysis, within-persons analysis, and fixed-effects analysis. Likely, results from one type of analysis will not generalize to results from another type of analysis, and the interpretation of results strongly relies on the type of data analyzed as well as the content of the variables included. We discuss the interpretation of relationships estimated through these four types of analyses and tie these to the data types that can be used.

This chapter has been adapted from: Epskamp, S., Hoekstra, R. H. A., Burger, J., & Waldorp, L. J. (2022). Chapter 9. Longitudinal design choices: Relating data to analysis. In Isvoranu, A. M., Epskamp, S., Waldorp, L. J., & Borsboom, D. (Eds.). *Network psychometrics with R: A guide for behavioral and social scientists.* Routledge, Taylor & Francis Group.

2.1 Introduction

This chapter discusses several concepts and explains seemingly paradoxical differences on different levels of analysis. To simplify matters, we will mainly discuss two-variable models in which we only model a single correlation between variables.[2] First, we discuss different types of *data* that can be used for multivariate correlational analysis such as pairwise Markov random field (PMRF) estimation or factor analysis. Second, we discuss different types of *analysis* that can be used on these types of data. Finally, we discuss differences between within-person and between-person effects, especially in relation to the time frame of measurement. Chapter 3 will continue this discussion and also introduce temporal dynamics through vector auto-regressive modeling, and discuss differences in temporal and contemporaneous results.

2.2 Data designs

Figure 2.1 shows several types of data that can be used for network analysis.[3] Data sets in which we have one particular observation per person can be termed *single measurement data*, shown in the top left panel of Figure 2.1. Such data sets are often also termed *cross-sectional* data, but this term is incorrect and may potentially be confusing; the term *cross-sectional* refers to a cross-section at a particular moment in time, one which could have been different at another moment in time. Single measurement data may also refer to other types of data for which only one observation is available per person. For example, data sets in clinical psychology often involve clinician-rated scores on symptoms that cover a range of time (e.g., depressed mood over a period of several months), and variables that do not vary over time could be included in an analysis (such as experiencing childhood trauma in adults).

All other three types of data introduce a second dimension in the sampling setup: time. The top-right panel of Figure 2.1 shows $N = 1$ data, in which a single individual is measured over time. Such data sets are increasingly common, as technological advances made it possible to store these large data sets, and statistical advancements facilitated the analysis of this type of data (Hamaker, 2012). For example, we may measure a client in clinical practice through the use of the Experience Sampling Method (ESM) several times per day over a period of several weeks, leading to many observations of the same person over time.

Single measurement data only involves the people dimension, and $N = 1$ time series data only involves the time dimension. These two dimensions can also both be used in the sampling plan in a setup where multiple people are measured on multiple occasions. While any combination of number of people and number of measurements per person is possible, typically, data sets can be classified as one of the two bottom panels of Figure 2.1. In panel data designs, many people are measured on

2 A two-variable pairwise Markov random field (PMRF) is simply a model with two variables with one correlation that is not conditional on any other variable (because there are none).

3 While in this thesis I predominantly focus on network analysis, the general description of types of data and types of analyses in this chapter equally applies to other statistical routines, such as regression analysis or factor analysis.

relatively few occasions with large differences of time in between. For example, a large sample can be measured yearly for a period of 5 years. Finally, in $N > 1$ time series data, a smaller set of people (perhaps 50 – 100, rather than hundreds to thousands often seen in panel data sets) are measured on many measurement occasions, usually with small windows of time in between. As such, panel data could be considered a collection of single measurement data sets, whereas $N > 1$ time series can be considered a collection of $N = 1$ time series data sets. Both types of data, however, allow for some form of separation of within-person and between-person effects, something that is impossible to do in single measurement data and $N = 1$ time series.

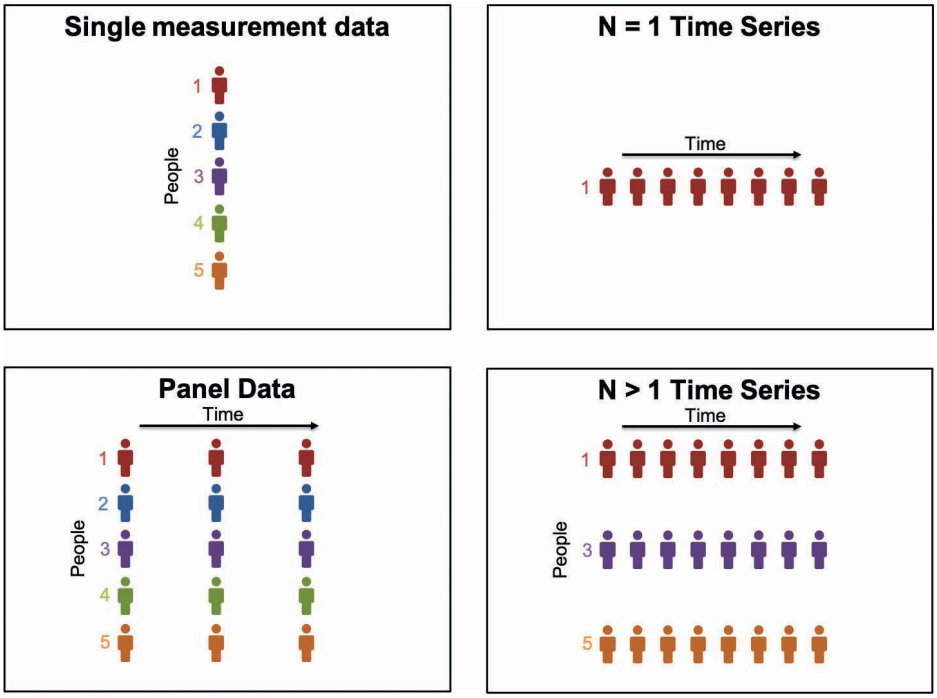


Figure 2.1. Different types of data that can be used for multivariate correlational analyses such as network analysis. In single measurement data only one observation is present per person, in $N = 1$ time series repeated measures of one individual over time are available, in panel data some repeated observations of many people are available, and in $N > 1$ time series many repeated measures are available from, usually, a relatively small set of people.

2.3 Analysis designs

Figure 2.2 shows examples of different types of analysis that could be performed. Such analyses are often confused with the types of data shown in Figure 2.2. While some types of analysis require certain types of data (e.g., within-person analysis requires many observations of one particular individual), the relationship between types of data and types of analysis is often not transparent.

The top left panel of Figure 2.2 shows an example of *cross-sectional analysis*. Here, several people are analyzed on data from one particular moment in time, and, importantly, these people could

have responded differently in other moments in time. We can obtain a relationship of interest in such an analysis, such as a correlation between 'drinking coffee' and 'being productive,' or more advanced network structures. Suppose we assessed several researchers on a particular day of writing, whether they drank much coffee and whether they wrote many words, and we find a positive correlation between these variables. This correlation is a blend of between- and within-person effects (Molenaar, 2004; Hamaker, 2012): These variables could be correlated because researchers who on average drink a lot of coffee on average also write many words (a between-person effect), or because when researchers drink more coffee than they normally do, they write more words than they normally do (a within-person effect). This downside of cross-sectional analysis is important to recognize: cross-sectional analysis cannot distinguish between these types of variances.

While cross-sectional analysis is often termed *between-person analysis*, this generalization is not correct. *Between-person analysis* refers to the analysis of relationships between aspects that are stable in the studied people over time. As such, we can interpret between-person analysis to be collapsed over the time dimension, investigating only stable averages per person. For example, a between-person analysis could investigate if the researchers that drink much coffee on average, also, on average, write more words per day. The exact term *between-person* then refers to studying individual differences in variables that do not vary within (potential) repeated measures of any particular person. It is important to note that the term does *not* refer to any interactions between people (such as social dynamics). Indeed, we usually assume people in the sample to be statistically independent of one another.[4]

A *within-person* analysis, on the other hand, solely focuses on studying the variance of variables in repeated measures of one particular person. A true within-person analysis always refers to the study of one particular person over time. For example, we can study one particular researcher, measure their coffee consumption and number of written words per day, and observe that on days this researcher drank a lot of coffee, they also wrote many words. This relationship is then established purely *within* the observations of one particular person, and this relationship may differ in other people.

Finally, the *fixed-effects* analysis collapses sampling over people rather than time. In such an analysis, within-person relationships of an average person are analyzed. For example, we can find that in general, whenever researchers drank more coffee than their average in a day, they also tended to write more words than their average on that day. This marks a distinct difference between true within-person analysis and fixed-effects analysis: A within-person analysis refers to relationships between deviations from the average of a particular person, whereas fixed-effects analysis refers to relationships between deviations from the average of a hypothetical average person. This person, however, does not exist. As such, the fixed-effects results are not established *within* the observations of one particular person. To this end, there is some debate on whether or not a fixed-effects analysis can really be interpreted as a within-person analysis.

---

4 Statistical independence between cases (subjects) means that the scores of one case cannot be predicted by the scores of another case. With this assumption, the likelihood function required to estimate parameters becomes a product of likelihoods per case, which is a fundamental property in many statistical routines.

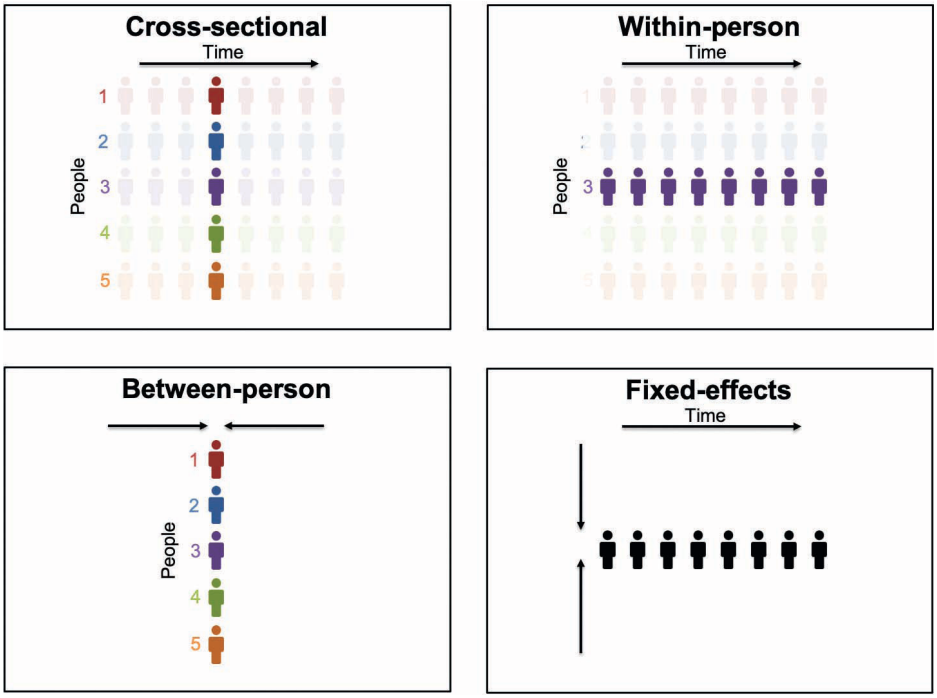


Figure 2.2. Different types of analysis that can be performed. In cross-sectional analysis, only one observation per person is analyzed, which represents a snapshot of those people over time. In between-person analysis, relationships between stable averages or traits are analyzed. In within-person analysis, only repeated measures from one person are considered. In fixed-effects analysis, the within-person effects of "the average person" are studied.

2.4 Differences between data and analysis

The previous sections introduced several different terms. For data, we can distinguish between four broad categories: (A) Single measurement data: Data of many subjects with only one observation per subject, (B) Panel data: Data of many subjects measured on a few occasions, (C) $N = 1$ time series data: Data of a single subject measured on many occasions, and (D) $N > 1$ time series: Several subjects measured on many occasions. In terms of analyses, we can also distinguish roughly between four categories: (1) Cross-sectional analysis: Analyzing a relationship between variables across people that can be expected to change on these variables over time, (2) Between-person analysis: Analyzing a relationship across people on variables/aspects that are not expected to change over time (or on the stable part of these variables), (3) Within-person analysis: Analyzing a relationship across time in a single individual, and (4) Fixed-effects analysis: The within-person relationship of an average person (aggregated over people).

With our vocabulary now established, we can continue discussing how these types of data and analysis relate to one another. In general, a cross-sectional analysis can be performed on single-measurement data and on temporal slices of panel and $N > 1$ time series data. A within-person analysis

can be obtained through $N = 1$ time series analysis or in $N > 1$ time series analyses (by analyzing each person separately). A between-person analysis can be performed on panel data and $N > 1$ time series data sets using methods that adequately separate (fixed-effect) within-person effects from between-person effects. A between-person analysis can also be performed on a single measurement data set if it can be assumed that these responses reflect stable averages over time. Finally, a fixed-effects analysis can be performed in panel data and $N > 1$ time series data, or on single measurement data if it can be assumed that the single observations correctly represent deviations from the person-wise averages.

Single-measurement data, therefore, could be used in principle to perform cross-sectional analyses, between-person analyses, and fixed-effect analyses, depending on the content of the variables in the data set. It should be noted, however, that evidence suggests that responses to trait-like questions on stable averages (e.g., 'are you a person that on average drinks a lot of coffee?') are usually still impacted by the state of that person in that particular moment in time (e.g., Brose et al., 2013). In principle, it is also possible to ask questions on deviations from the person-wise average (e.g., 'did you drink more coffee than average today?'), in which case a fixed-effects interpretation can be used in the analysis. Such designs, however, are not common yet. As such, the interpretation of analysis performed on single-measurement data discussed in the previous chapters relies on the question of validity: do the variables measure what they intend to measure? This question goes beyond the scope of this chapter, but in general shows that considering any analysis on single-measurement data a cross-sectional analysis is incorrect.

2.4.1 Ergodicity

Many authors consider cross-sectional analysis insufficient, as cross-sectional analysis can be considered a blend of within- and between-person effects. Furthermore, they consider the within-person effects to be the general target of inference in psychology – such as establishing that drinking coffee is a useful intervention to improve productivity for one particular researcher – and that cross-sectional analysis can fail to retrieve these within-person relationships adequately. Taking it one step further, this line of reasoning can also be used to argue against the study of fixed-effects, as these effects model an average person and not actually an effect of the people in the analyzed data set. In his seminal work on this topic, Molenaar (2004) showed that only when a concept known as *ergodicity* holds, cross-sectional and fixed-effects analyses align with within-person analyses for every person in the sample. In the relevant cases for this thesis (chapters 6 – 8), ergodicity mainly entails that it needs to be assumed that every person is a virtual replication of one another measured merely at different points in time. That means: every person has the same means and the same model structure that is also stable over time. Each person having the same means also indicates that there should be no between-person relationships (e.g., an empty network at the between-person level). Only in this case will cross-sectional analysis and fixed-effects analysis align with within-person analysis of every individual. These assumptions are so strong that ergodicity is never likely to hold in psychological data.

While this reasoning is undeniably true, it is important to note again here that *within-person* does not refer to things that happen within a person, but rather to relationships that can be established within the potential repeated measures of that person. The term *between-person*, on the other hand,

refers to relationships that can only be obtained by investigating individual differences. As such, *between-person* does not solely place the relationship outside of the person, and relationships at the between-person level can also be indicative of dynamics within an individual (potentially over time).

For example, Figure 2.3 (based on an example by Hamaker, 2012) shows a positive (fixed-effect) within-person relationship between typing speed and spelling errors, but also a between-person relationship between the same variables. While seemingly paradoxical (the term Simpson's paradox is often used for such a discrepancy), these relationships are not paradoxical and can readily be understood. Whenever a person writes faster than their average, that person also tends to make more spelling errors. However, people that on average write very fast are likely very experienced writers and, therefore, also likely to make fewer spelling errors. Thinking in terms of interventions here: suppose a researcher wants to make as few as possible spelling errors, then that researcher would likely consider writing slower than usual.

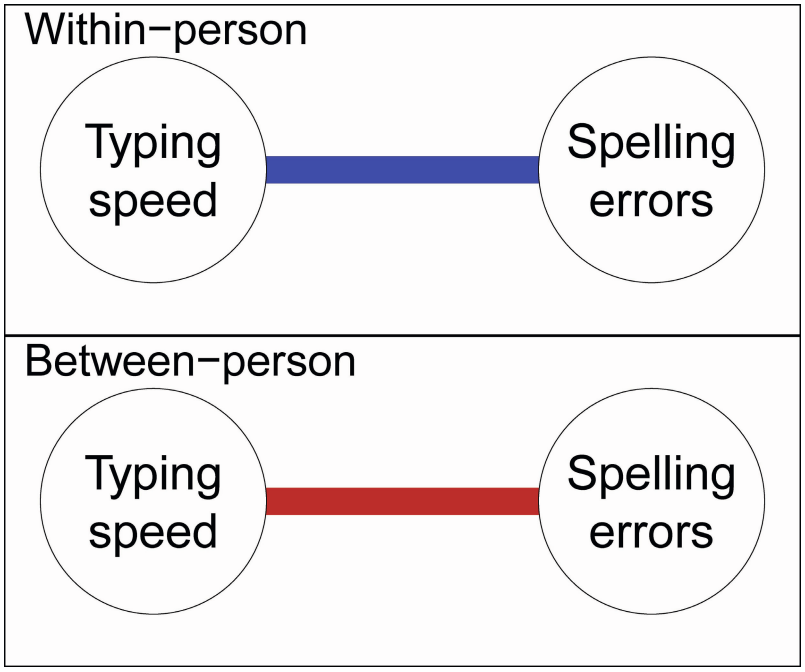


Figure 2.3. Hypothetical examples of within-person (or fixed-effect) and between-person relationships that could be found. Adapted from Epskamp, Waldorp, et al. (2018).

2.4.2 The importance of time scales

An important consideration to the interpretation of results is the time frame in which the study was conducted. For example, Figure 2.4 shows hypothetical data of a person measured over the period of a few weeks or months on 'anxiety' and 'stress.' Using this person's repeated measures, we may establish a within-person correlation between these two variables: whenever this person experienced stress this person was likely also experiencing anxiety. Figure 2.5, on the other hand, shows an entirely different scenario in which 'lung cancer' and 'smoking' are measured. In any time

series analysis of only a few weeks or months, we will not establish any within-person correlation for any of the people in the sample, as there is not enough variance for statistical analysis. However, if we perform a $N > 1$ time series analysis over a few weeks or months, or a cross-sectional analysis at any point in time, we may obtain a between-person correlation between these two variables. In a longer panel design, we may even obtain a within-person relationship between these variables, but likely only in fixed-effects, as none of the subjects show enough variation in their scores to warrant statistical analysis.

As such, relationships that are between-person in one study can be within-person in another study, depending on the time frame over which the study took place. This shows that the argument for within- and between-person effects is not black and white: both levels of analysis are essential for understanding relationships between variables of interest. To this end, it is vital to separate these levels of analysis whenever possible. Beyond separating between-person from within-person effects (at least the fixed effects), the temporal dependency of time series data also allows for further separation of within-person effects: separating effects that take place over time (temporal effects) from effects that take place within the same window of measurement (contemporaneous effects). This separation will be discussed in the next section.

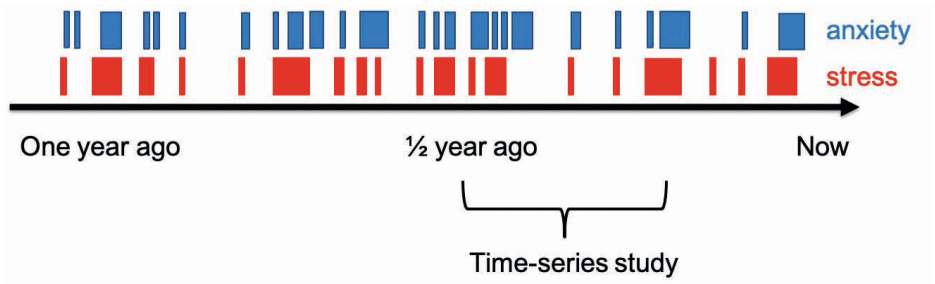


Figure 2.4. Example of $N = 1$ data collected in a relatively short time frame. Colored boxes indicate times that one of these variables would be endorsed. In this example, we could perhaps establish a within-person relationship between anxiety and stress, potentially even a temporal effect depending on the choice of lag-interval (time between measurements) in a vector auto-regressive analysis.

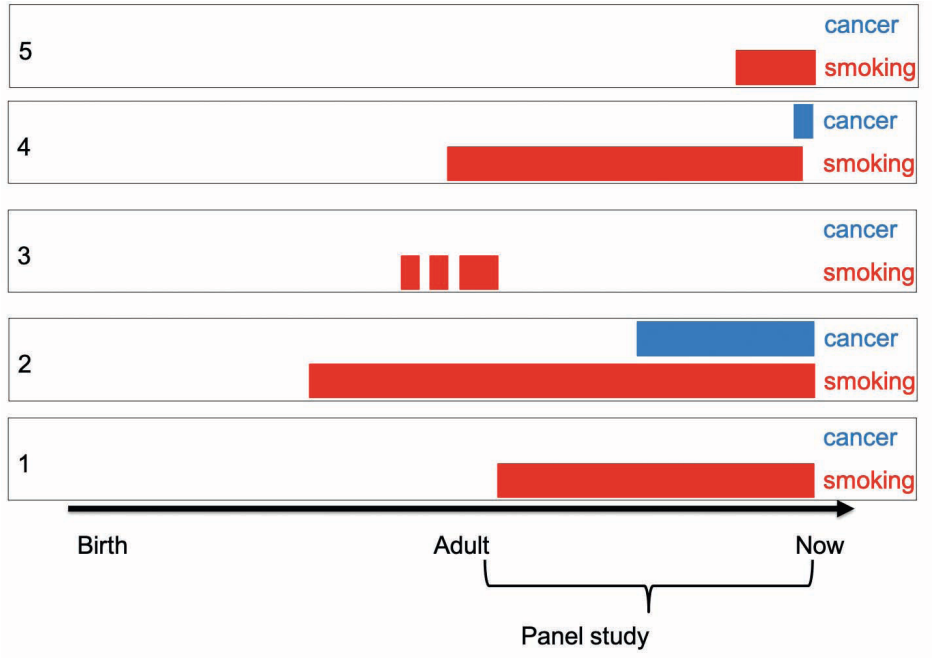


Figure 2.5. Example of $N > 1$ panel data. Colored boxes indicate times that one of these variables would be endorsed. In this case, we would likely not have enough variance to establish a within-person relationship in any person.

2.5 Separating contemporaneous and temporal effects

When analyzing longitudinal data, we need to consider that responses per person are temporally ordered and that there are likely substantial temporal dependencies between these responses. For example, suppose a person on average is very energetic but reports to feel very tired at 12 pm on a given day. Then, we can predict that this person is still tired at 3 pm, regardless of this person being energetic on average.

The most commonly used method for handling temporal ordering in data is through the use of the lag-1 vector auto-regressive model (VAR). In VAR, linear regression is used between consecutive time points to model temporal dependencies, allowing one to gain separate estimates for relationships across time (temporal effects), and relationships in the same window of measurement after controlling for temporal effects (contemporaneous effects). The VAR variant that is the core focus of this thesis is the *graphical VAR* (GVAR) model in which the contemporaneous effects are further modeled as a Gaussian graphical model (GGM; see Epskamp, Waldorp, et al., 2018g). As such, the GVAR model returns two network structures: a *temporal network* which is a directed network of

temporal relationships,[5] and a *contemporaneous network*, which is an undirected network of contemporaneous relationships. Figure 2.6 gives an example of such networks and how these can lead to different interpretations: these networks show that when a person is exercising they are more energetic than their average, but after a person is exercising they are less energetic than their average. The GVAR model separates two levels of within-person variance: Temporal and contemporaneous relationships. Important to note is that both of these have a within-person interpretation: These relationships only investigate relationships between deviations from the person-wise mean. It may also be interesting to investigate relationships of the means across people – between-person effects. When $N > 1$ data is available, this can be done by forming a separate GGM on the variance--covariance structure of the means per person.

As such, we can distinguish between three network structures: (1) *Temporal networks* which contain within-person (dynamic) relationships over time; (2) *Contemporaneous networks* which contain within-person relationships in the same window of measurement, and (3) *Between-person networks* which describe relationships of stable averages across people. Temporal and contemporaneous networks can be obtained per person if there are enough observations, or can be obtained through a fixed-effects analysis. The next chapter will discuss estimation of the GVAR model in more detail.

5 The temporal network used in GVAR modeling can also be termed a lag-1 network, with the word 'lag' indicating the number of time steps between measurements in an effect. In principle, such a model can be extended with lag-2 effects, lag-3 effects, and so forth, but in practice doing so will likely lead to an intolerable number of parameters that need to be estimated. As such, we only focus here on lag-1 models. A lag-0 model can be seen as a model in which no correction has been performed for temporal effects. Such a network should not be confused with a contemporaneous network, which controls for temporal effects as well.

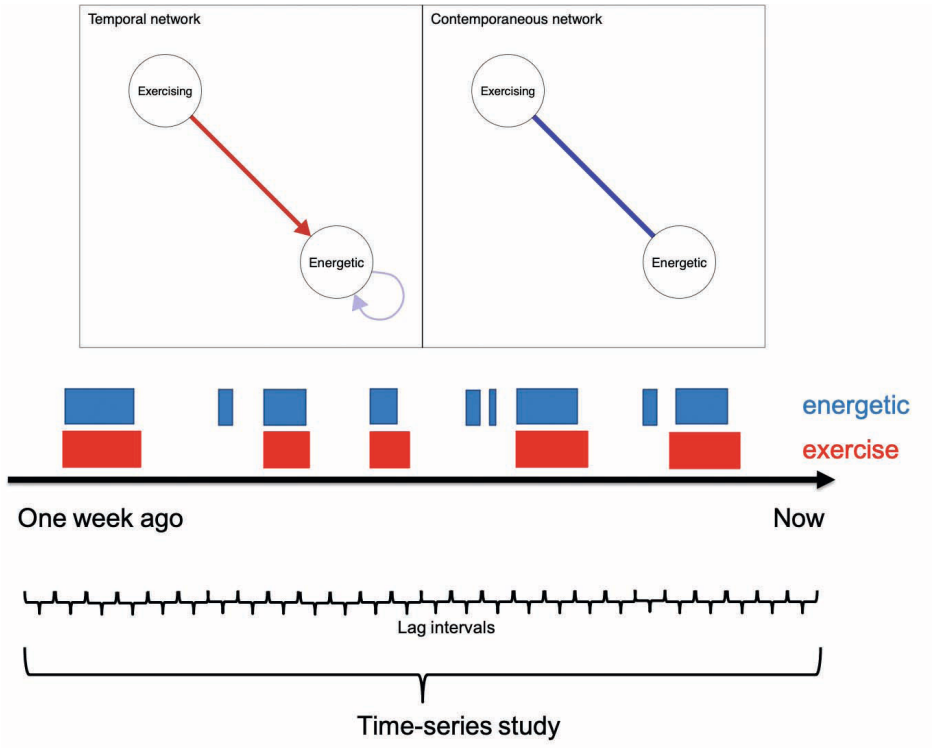


Figure 2.6. A graphical vector auto-regressive (GVAR) model with corresponding hypothetical data of two variables: 'exercise' and 'energetic.'

2.6 Conclusion

This chapter discussed challenges that emerge when repeated measures are introduced to the sampling plan. We distinguished four types of data: single measurement data (one measure per person), panel data (many people measured a few times), $N = 1$ time series data (one person measured many times), and $N > 1$ time series data (several people measured many times). We also discussed different analysis options: cross-sectional analysis, between-persons analysis, within-persons analysis, and fixed-effects analysis. The interpretation of results strongly relies on the type of data analyzed as well as the content of the variables included. When longitudinal data is analyzed, the graphical VAR (GVAR) model can be used to separate temporal within-person relationships from contemporaneous within-person relationships. As such, two within-person networks can be obtained: a temporal and a contemporaneous network. In the next chapter, we will discuss the addition of temporal dynamics and the estimation of network models from panel data and time series data in more detail.

CHAPTER 3

NETWORK ESTIMATION FROM TIME SERIES AND PANEL DATA

CHAPTER 3

Abstract

This chapter discusses how to estimate graphical vector auto-regressive (GVAR) network models from time series and panel data. The GVAR model can be used to estimate *temporal networks* (within-person relationships over time), *contemporaneous networks* (within-person relationships in the same window of measurement), and *between-person networks* (relationships between the means of persons in the data). The chapter explains how such network structures can be estimated using the R-packages *graphicalVAR*, *psychonetrics*, and *mlVAR*. We conclude with a discussion of current practical and methodological challenges, including the power of $N = 1$ networks, heterogeneity, missing data, model assumptions, and the importance of identifying appropriate time scales.

This chapter has been adapted from: Burger, J., Hoekstra, R. H. A., Mansueto, A. C., & Epskamp, S. (2022). Chapter 10. Network estimation from time series and panel data. In Isvoranu, A. M., Epskamp, S., Waldorp, L., J. & Borsboom, D. (Eds.). *Network psychometrics with R: A guide for behavioral and social scientists.* Routledge, Taylor & Francis Group.

3.1 Introduction

The previous chapter introduced *time* into the sampling design of studies. With the addition of time, longitudinal analysis of multiple measures per person becomes possible. As discussed in chapter 2, this step is vital in separating within- from between-person relationships. Time adds a new level of complexity to the modeling frameworks we have used before: temporal dependencies between observations of the same person over time. This complicates the models, as many more parameters need to be estimated, but comes with the substantial benefit of allowing researchers to study temporal effects over time – often termed 'dynamical relationships.' With the advent of modern data collection methodologies, such as electronic diaries or wearables, time series data have become a new data source to estimate networks, and with it, the use of dynamical network models has grown prominent. This chapter introduces the estimation of dynamical networks from longitudinal data. The chapter begins with a summary of the main modeling framework used, followed by explanations on how to estimate these models from $N = 1$ and $N > 1$ data sets. The chapter will then conclude with an overview of practical and methodological challenges in dynamical network analysis.

3.2 Graphical vector auto-regression

The main model we focus on in this chapter is the lag-1 graphical vector auto-regressive (GVAR; Epskamp, Waldorp, et al., 2018; Wild et al., 2010) model introduced in chapter 2, and further detailed in Box 3.1. In this model, a person's responses on a certain measurement are modeled as a Gaussian graphical model (GGM; see Epskamp, Waldorp, et al., 2018) after conditioning on their responses in the previous measurement. Alternatively, the GVAR model can also be interpreted as a multivariate regression on the previous responses, with the residuals (then termed *innovations*) being modeled through a GGM. The GVAR model includes two network structures: A *temporal network* that encodes how well deviations from the person-wise mean in one variable at a certain measurement occasion predict deviations from the person-wise mean in the *next* measurement occasion,[6] and a *contemporaneous network* that encodes relationships between variables within the same measurement occasion and after controlling for temporal effects. These networks allow for a within-person interpretation and can be estimated per person or for the average person (fixed effects). In $N > 1$ data, relationships between the means can further be investigated to construct a GGM termed the *between-person network*.

Figure 3.1 shows an example of these three network structures, estimated from time series data collected by Fried, Papanikolaou, et al. (2022).[7] The temporal network in the left panel of Figure

6 Temporal connections can also be said to encode *Granger causality* (Eichler, 2007; Granger, 1969) as they encode temporal predictions. However, that does not mean the edges necessarily encode *causal* relationships. Ultimately, temporal edges are just partial correlations between a lagged (encoding the previous time point) and a non-lagged variable after controlling for all other lagged variables. As such, similar reservations to causal interpretations in the temporal network should be taken as discussed in (Epskamp, Haslbeck, et al., 2022).

7 The data, including a detailed overview of the measures used, are available online at https://osf.io/mvdpe/.

3.1 contains self-loops (auto-regressions) and edges between nodes (cross-lagged regressions). For example, the self-loop on 'difficulty relax' indicates that when a person had more difficulty relaxing than their average in one measurement occasion, that person likely still had a higher than average difficulty to relax in the next measurement occasion.[8] Edges between nodes indicate similar predictions but then for different nodes. For example, we can see that 'angry' and 'irritable' predict one another well over time. The contemporaneous network in the middle panel of Figure 3.1 indicates, for example, that a person who is currently more 'worried' than their average is likely also more 'nervous' than their average at the same time. Lastly, the between-subjects network in the right panel of Figure 3.1 shows, for example, that individuals who 'worry' a lot on average also tend to be individuals that are 'irritable' on average. The between-person network also shows an interesting negative relationship between 'feelings of hopelessness'[9] and 'difficulty relaxing'. As explained in Epskamp, Haslbeck, et al. (2022), this could be due to a common effect structure. For example, perhaps both being a person who often feels hopeless and being a person who often has difficulty relaxing leads one to become a person who often worries.[10] Interestingly, this common effect structure can also be seen at the temporal level. Another explanation is Simpson's paradox, in which an effect becomes different when conditioning on a different level of the data (Hamaker, 2012; Kievit et al., 2013).

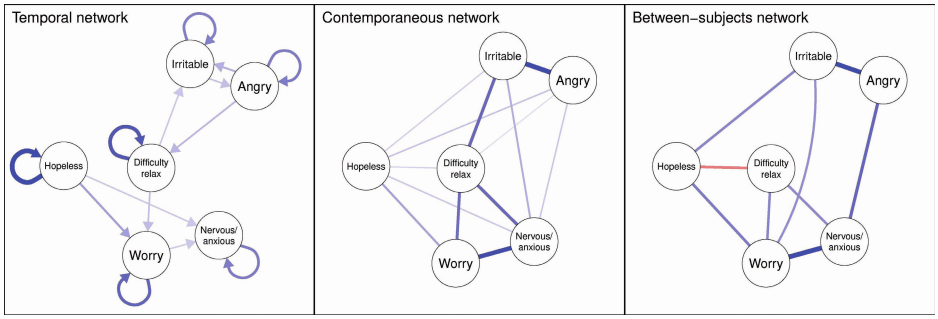


Figure 3.1. Example of a temporal (fixed effects), contemporaneous (fixed effects), and between-subjects network with six nodes based on data collected by Fried, Papanikolaou, et al., (2022). The data consist of four measurements per day for 14 subsequent days, filled in by 80 undergraduate students. Networks were estimated using two-step multi-level estimation with the *mlVAR* package using correlated random effects. The network structure was obtained by thresholding edges at $\alpha = 0.05$ (using an AND-rule for the contemporaneous and between-person networks).

8 The inverse interpretations are also true: a positive edge also indicates that whenever people experienced *less* difficulty relaxing than their average in one measurement occasion, they likely also experienced *less* difficulty relaxing in the following measurement occasion.

9 The actual measure used for this node was "I felt that I had nothing to look forward to.".

10 These data were collected during the first weeks of the 2020 COVID-19 pandemic outbreak. As such, any between-person effects could also be due to the general atmosphere of this time.

The vector auto-regressive (VAR) model is a model that assumes responses from longitudinal data of a single subject to be normally distributed after controlling for the previous measurement. Let $y_{p,t}$ represent a set of responses from a subject p measured at time point t (for simplicity we do not denote random vectors different from realizations here). The VAR model can then be written as:

$$y_{p,t} | \, y_{p,t-1} \sim N\left(\mu_p + B_p\left(y_{p,t-1} - \mu_p\right), \Sigma_p^{(C)}\right)$$

The B_p matrix encodes these temporal regression effects, with elements β_{ijp} (row i, column j) encoding the temporal effect from variable j to variable i for subject p. The transpose of this matrix therefore encodes a *temporal network*.[11] The matrix $\Sigma_p^{(C)}$ can be interpreted as the *contemporaneous* variance-covariance matrix: The variance-covariance structure after controlling for temporal dependencies. The subscript p indicates that these matrices can be modeled per person. In *graphical* vector auto-regression (GVAR; Epskamp, Waldorp, et al., 2018; Wild et al., 2010), we further model $\Sigma_p^{(C)}$ through the use of a Gaussian graphical model (GGM; see Epskamp, Waldorp, et al., 2018):

$$\Sigma_p^{(C)} = \Delta_p^{(C)}(\mathrm{I} - \Omega_p^{(C)})^{-1}\Delta_p^{(C)}$$

The matrix $\Omega_p^{(C)}$ encodes a *contemporaneous* network. Finally, in $N > 1$ data, the means can subsequently be treated as a random variable and also be modeled with a GGM:

$$\mu_p \sim N\left(0, \Sigma^{(B)}\right); \quad \Sigma^{(B)} = \Delta^{(B)}(I - \Omega^{(B)})^{-1}\Delta^{(B)}$$

in which the matrix $\Omega^{(B)}$ encodes the *between-persons network*.

Box 3.1. Technical description of the graphical vector auto-regressive model.

3.3 $N = 1$ estimation: Personalized network models

The GVAR model can be estimated in various ways from time series data of a single subject ($N = 1$). Doing so will return a temporal and a contemporaneous network.[12] Estimation of the GVAR model from $N = 1$ data mostly follows the same principles as estimating GGMs from single measurement data: the GVAR model can be estimated through multivariate or univariate (nodewise) estimation, using frequentist and Bayesian estimation, and model selection can be per-

11 Unlike the contemporaneous and between-person networks, the temporal network is not standardized in this notation. Typically, all variables are standardized before analysis to make the B matrices interpretable (Bulteel et al., 2016). Alternatively, these coefficients can be standardized to *partial directed correlations* (Wild et al., 2010). Elements from $\Omega_p^{(C)}$ are sometimes termed *partial contemporaneous correlations*, and elements from $\Omega^{(B)}$ are sometimes termed *partial between-person correlations* (PBC).

12 It is not uncommon that only the temporal network obtained through a (G)VAR analysis is of interest. Not reporting the contemporaneous network, however, is not recommended, as (1) the temporal network is only half the statistical model, (2) the method might lack sensitivity, especially in temporal connections (which are often weak), and (3) relationships in the same window of measurement can be interesting as well (Epskamp, Waldorp, et al., 2018).

formed through regularization or other model search strategies. This section will discuss estimation strategies for the GVAR, followed by model selection strategies.

3.3.1 Maximum likelihood estimation

Epskamp, Haslbeck, et al. (2022) discussed maximum likelihood estimation as an approach to establish multivariate structures: parameters are found under which the data were most likely to occur. This process requires a *joint likelihood* expression of the entire data. For single measurement data, computing the joint likelihood is relatively straightforward: this quantity can be computed by multiplying (summing) the (log) likelihoods of every individual case – the rows in the data set. While proper maximum likelihood estimation of the GVAR model parameters from time series data is possible in principle as well, it becomes computationally much more challenging in practice compared to analyzing single measurement data (Ciraki, 2007). This is because an inherent property of time series data is that cases in the data set are no longer *independent*. As a result, the likelihood can no longer be formed easily. To this end, the covariance between every case needs to be modeled, resulting in a covariance matrix that contains a row/column for every variable of every case. In the case of 100 measures on 10 variables, the variance–covariance matrix modeled would be a 1,000 times 1,000 matrix, which is too large for most software packages to handle.

3.3.2 Data augmentation

To overcome the computational challenges of maximum likelihood estimation, GVAR estimation typically relies on a trick that involves augmenting the data (Hamaker et al., 2002; Lane et al., 2019). While this trick no longer results in 'true' maximum likelihood estimation, the resulting parameters and standard errors are comparable to true maximum likelihood estimates, especially at larger sample sizes. Figure 3.2 demonstrates how the data can be augmented, by making a copy of the data, shifting that copy by one row, and appending the shifted data set to the original data set. This way, each row t contains both responses at time t as well as the previous time $t - 1$. If the mean structure is not explicitly modeled, the data can also be centered. The variance–covariance matrix of this augmented data takes the form of a block Toeplitz matrix and can be modeled in the same manner as the variance–covariance matrix of cross-sectional data (see Box 3.5 at the end of this chapter). Alternatively, regression models can be used on the augmented data: A multivariate regression model can be used with the set of responses at t as dependent variables and the set of responses at $t - 1$ as independent variables, resulting in the parameters of the temporal network. Subsequently, the residuals can be analyzed using the methods discussed in Epskamp, Haslbeck, et al. (2022) to obtain the contemporaneous network. Another alternative is to use univariate estimation to obtain all temporal effects by regressing each variable on all lagged variables first, and subsequently, to obtain all contemporaneous effects by using univariate estimation tools on the residuals as discussed in Epskamp, Haslbeck, et al. (2022).

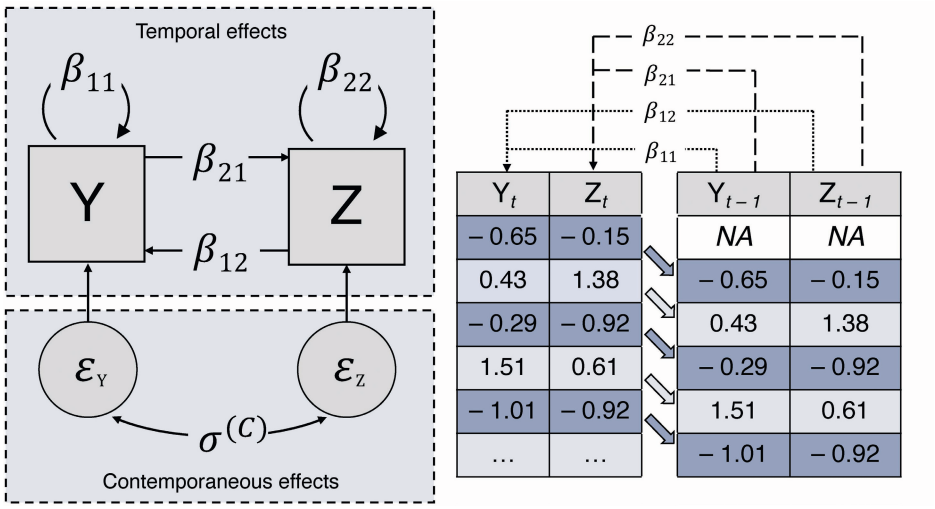


Figure 3.2. Example of a graphical vector auto-regressive (GVAR) model (left) and data augmentation usually used to estimate model parameters (right). In the data augmentation, all variables (here *Y* and *Z*) are copied and shifted by one row (also termed *lagged*). Rows in the augmented data that cross a night or non-equal measurements can be removed before analysis. The temporal effects (β parameters) encode within-person prediction over time, and the contemporaneous effects (here $\sigma^{(C)}$) can be used to form a GGM encoding relationships in the same window of measurement. For interpretable parameters the data should also be within-person centered or the mean structure should be explicitly modeled.

3.3.3 Model selection

Model selection can be performed in similar manners as discussed in Blanken, Isvoranu et al. (2022). For example, edges could be selected based on some threshold or through stepwise model selection search strategies, which has been implemented in the *psychonetrics* package, using mostly the same code as described in Epskamp, Haslbeck, et al. (2022) and Blanken, Isvoranu, et al. (2022) except that the gvar model function is used instead of the ggm model function (Epskamp, 2020b). A popular way in which the GVAR is estimated is through a regularization procedure closely related to the *EBICglasso* procedure used in GGM models. The multivariate regression with the covariance estimation (MRCE; Rothman et al., 2010) algorithm can be used to sequentially estimate a regularized temporal network (using LASSO regularization) and a contemporaneous network (using the GLASSO algorithm) until convergence. This algorithm utilizes two tuning parameters, one for the temporal coefficients and one for the contemporaneous coefficients, which can be selected using EBIC model selection (Abegaz & Wit, 2013). This algorithm has been implemented in the *graphicalVAR* package (Epskamp, 2020a), and the *SparseTSCGM* package (Abegaz & Wit, 2021). In Box 3.2, we illustrate an example for estimating temporal and contemporaneous networks using *graphicalVAR* and *psychonetrics*.

Suppose a data frame *data* in R has the following form:

subject	day	beep	worry	relax	angry
1	1	1	2	1	2
1	1	2	2	2	2
1	1	3	3	1	1
1	2	1	2	1	1
1	2	2	2	3	1
...	...	...	...	...	...

We can estimate GVAR model parameters from this $N = 1$ data set using the R packages *graphicalVAR* (Epskamp, 2020a) for regularized estimation (Abegaz & Wit, 2013; Rothman et al., 2010) and *psychonetrics* (Epskamp, 2022) for maximum likelihood estimation. The input to both packages is comparable, and requires information about the columns in the data to be stored first:

```
vars <- c ("worry", "relax", "angry") # Variables used in the model
dayvar <- "day" # The day variable, only use with >1 assessment/day
beepvar <- "beep" # The beep variable
```

These objects correspond to argument names in the R packages. The vars argument specifies the variables used in the analysis, the optional dayvar argument specifies the days and is used to cut out pairs of measurements that cross a night,[13] and the optional beepvar, corresponding to the measurement number within each day, can be used if the data contain missing measurements. Now, the *graphicalVAR* package can be used as follows:

```
library("graphicalVAR")
graphicalVAR(data, vars = vars, dayvar = dayvar, beepvar = beepvar)
```

The estimated network structures are stored as *partial directed correlations* (PDC) and *partial contemporaneous correlations* (PCC) for the temporal and contemporaneous networks respectively, and we can visualize the networks using the plot function. In *psychonetrics* the gvar model function can be used:

```
library("psychonetrics"); library("dplyr")
gvar (data, vars = vars, dayvar = dayvar, beepvar = beepvar, estimator = "FIML")
 %>% runmodel
```

Optionally, further model search functions can be applied such as prune and modelsearch. The weights matrices can be obtained using the getmatrix function, with omega_zeta indicating the contemporaneous network and PDC the standardized temporal network.

Box 3.2. Estimating $N = 1$ networks from time series data using *graphicalVAR* and *psychonetrics*.

3.3.4 Bayesian estimation

Another popular method for estimating VAR models (and, by extension, GVAR models) is through the use of multivariate Bayesian estimation by implementing the model in Bayesian sampling software such as JAGS (Plummer, 2003) or STAN (Carpenter et al., 2017). These software packages model a response vector as a function of the previous response vector by looping over the data when specifying the likelihood. To this end, the data need not be augmented as described above for other

13 Importantly: do not use the dayvar argument with only one observation per day, as then all data will be removed.

settings. Additional benefits of the Bayesian approach are that missing responses can easily be handled – even allowing for continuous time modeling (Ryan & Hamaker, 2022) – and that prior information could be used to improve estimation, as discussed in chapters 9 and 10. The GVAR model has also been implemented in the *BGGM* package (Williams & Mulder, 2020).

3.4 $N > 1$ estimation: Multi-level estimation

If intensive longitudinal data are available from multiple subjects, we might be interested in constructing a network of the average temporal and contemporaneous effects, the *fixed effects* network structures introduced in the previous chapter. A first, intuitive approach to estimating these fixed effects is to compute a network for each subject separately using the methods discussed above and subsequently calculate the averages for each parameter, as well as their variance–covariance structure. This approach of *pooling parameter estimates*, however, discards the nested structure of the data and relies on the – potentially underpowered – estimation of many (N) models separately. An alternative to estimating separate models per person is to estimate only one model for all observations. The most straightforward way to do this is to within-person center all variables,[14] combine all within-person centered data sets, and estimate a single GVAR model (making sure that responses from one person are not regressed on responses from another person). This process has been automated in the mlGraphicalVAR function in the *graphicalVAR* package[15] and the gvar function in *psychonetrics* (if the idvar argument is used). These methods provide good estimates of the fixed effect structures, but still do not properly take nesting of data points into account. These methods also do not provide insight into the variability of parameters across the sample, as individual networks have to be estimated separately per person.

3.4.1 Multi-level modeling

We can actively incorporate the nested structure of our data by estimating a *multi-level* (G)VAR model (Bringmann et al., 2013; Epskamp, Waldorp, et al., 2018). The term multi-level refers to the data being organized in two levels: Within-subject variance on level 1, and between-subject variance on level 2. In this approach, each parameter in the model (e.g., edge weights and means) is assumed to have a distribution over the population. Thus, in the estimation procedure, only these distributions (mean and variance of the parameter, and possibly the covariance between parameters) need to be estimated. The fixed effects can then be obtained from the centers of these distributions. Subsequently, the deviations from this center point – the *random effects* – can be sampled, which, together with the fixed effects, lead to estimates for the personal network models. Figure 3.3 shows how the parameter distributions of a multi-level VAR can inform fixed and random parameters in temporal networks.

14 For each variable removing for each person the mean of that person from the scores of the variable. This step is necessary to ensure that between-person effects are not included in the analysis.

15 Unlike the name suggests the mlGraphicalVAR function does *not* perform multi-level estimation. The function merely computes a pooled GVAR over all combined within-person centered data sets and runs the graphicalVAR function per person separately for individual networks (Epskamp, 2020a).

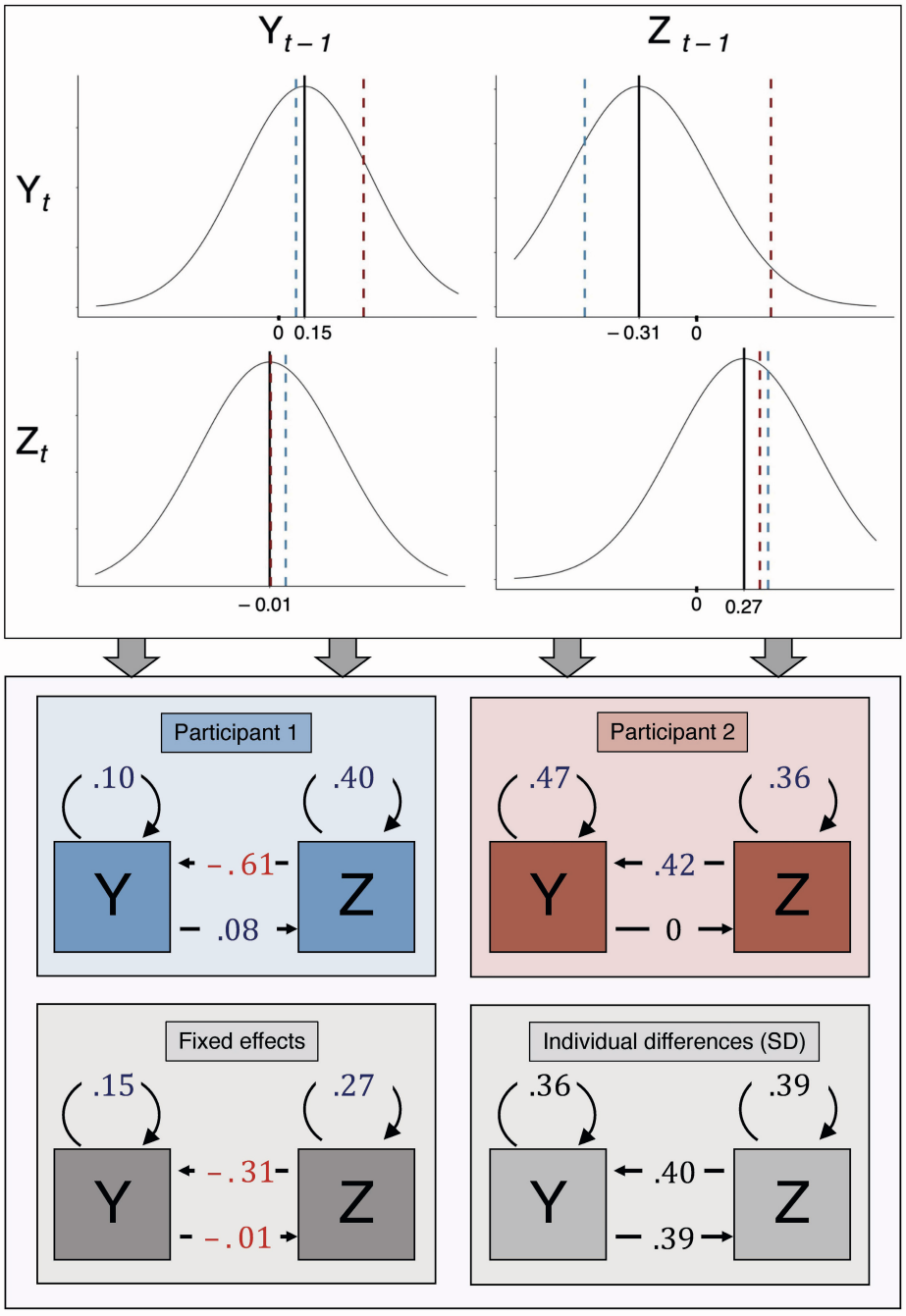


Figure 3.3. Multi-level model for the temporal effects of two variables Y and Z, with data simulated for N = 150. The distributions for the temporal effect parameters (top panel) are used to establish the random temporal effects for participant 1 and 2 (bottom panel, top row), as well as fixed temporal effects and the standard deviations of subjects on these effects (bottom panel, bottom row).

Estimating GVAR models through multi-level estimation has four main benefits. First, a single analysis can be performed on the entire data set, leading to a well-powered analysis based on a large sample size, especially for the estimation of fixed effects. Second, the multi-level analysis provides not only insight into the fixed effects structure, but also into the heterogeneity around these fixed effects through the standard errors of the random effects. Third, multi-level modeling can be used to separate within- and between-person variances, which also leads to estimates of the between-person structure. Finally, and perhaps most importantly, estimated individual network structures are typically closer to the fixed effects estimate compared to parameters that are estimated for that individual directly. This is termed *shrinkage*, as the estimates of different persons in the sample are "shrunken" towards each other. In other words, individuals' effects are also informed by other individuals and result in estimates that lie close to each other. This can lead to better estimates of the personal network structures, requiring fewer observations per person than performing many $N = 1$ analyses separately.

Multi-level modeling also has some downsides. Assuming a (typical normal) distribution across the population on parameters entails that these parameters do not differ in *structure*, only in weight. For example, suppose that a temporal edge A → B is modeled with a normal distribution across the population with mean (fixed-effect) 0.2 and standard deviation (of the random-effect) 0.1. This means that we would expect roughly 95% of the population to have individual edge weights for A → B between 0 and 0.4, with the remaining 5% lower than 0 (negative edge weights) or higher than 0.4. The model, therefore, does not assume that any of these persons have an edge-weight of *exactly* 0, which would lead to the edge not being included in the temporal network. To this end, multi-level modeling does not estimate individual network structures, only individual network parameters. These parameters are also shrunken towards the fixed effect, which makes it questionable if the individual network structures genuinely allow for a within-person interpretation (after all, the networks were not estimated within the data of every person separately). Another prominent downside of multi-level modeling is that the models quickly become very complicated and computationally too challenging to be estimated. This is because a joint distribution over all parameters needs to be estimated. To this end, it is generally not possible to include many nodes in multi-level analyses. Multivariate estimation methods typically only allow for a few nodes to be included, which is why the main estimation method we will discuss uses univariate estimation. In univariate estimation, about 6 (with correlated random effects) to 20 (with orthogonal random effects) can be included at most.

3.4.2 The two-step multi-level VAR algorithm

Univariate estimation – using sequential univariate multi-level models and combining the result – was first proposed for the multi-level VAR model in psychological literature by Bringmann et al. (2013). Epskamp et al. (2021) extended this approach for estimating GVAR models by separating within- and between-person effects (allowing for the estimation of between-person networks) and by estimating contemporaneous networks. This extension is termed *two-step multi-level GVAR* estimation (see Box 3.3), and is implemented in the *mlVAR* package (Epskamp, Deserno, et al., 2021), further described in Box 3.4. In step 1, the algorithm estimates the temporal and between-subjects networks by performing univariate multi-level modeling, predicting each variable from within-per-

son centered lagged variables and person-wise means. In step 2, the estimated residuals of the models run in step 1 are used in a new sequence of univariate multi-level models to estimate the contemporaneous effects. For separating within- and between-person variances, the algorithm makes use of within-person centering: using the sample means from every person separately to center variables. This requires decent estimates of the within-person means, meaning that several (at least about 20) measures have to be available per person.[16] To this end, two-step multi-level GVAR estimation can be used with $N > 1$ time series data, but not with panel data. The *mlVAR* package uses the *lme4* package (Bates et al., 2015) for multi-level estimation of all effects.

The *two-step multi-level* GVAR estimation algorithm, proposed by Epskamp, Waldorp, et al. (2018), is an algorithm for estimating multi-level GVAR models through a series of univariate multi-level regression analyses. First, the entire data set used is standardized to z-scores (subtracting the mean and dividing by the standard deviation). Let $z_{t,p}^{(worry)}$, $z_{t,p}^{(relax)}$, and $z_{t,p}^{(anger)}$ represent three variables answered by person p at measurement occasion t. To separate within- and between person variances, we within-person center lagged variables as predictors (Hamaker & Grasman, 2015): $\tilde{z}_{t-1,p}^{(\dots)} = z_{t-1,p}^{(\dots)} - \bar{z}_{p}^{(\dots)}$, in which $\tilde{z}$ represents a within-person centered variable and $\bar{z}$ the person-wise mean. In step 1, for each variable a univariate multi-level regression is performed using that variable as a dependent variable and all within-person centered lagged variables together with the person-wise means as independent variables:

$$z_{t,p}^{(worry)} = \beta_{0p} + \beta_{11p}^{(T)} \cdot \tilde{z}_{t-1,p}^{(worry)} + \beta_{12p}^{(T)} \cdot \tilde{z}_{t-1,p}^{(relax)} + \beta_{13p}^{(T)} \cdot \tilde{z}_{t-1,p}^{(anger)} + \beta_{12}^{(B)} \cdot \bar{z}_{p}^{(relax)} + \beta_{13}^{(B)} \cdot \bar{z}_{p}^{(anger)} + \varepsilon_{t,p}^{(worry)}$$

The $\beta^{(T)}$ parameters form the individual temporal networks, and the $\beta^{(B)}$ parameters can be used to form a GGM in the same way univariate regressions in univariate GGM estimation are averaged to partial correlation coefficients (Epskamp, Waldorp, et al., 2018). In the second step, the estimated residuals of the multi-level regression models in step 1 are used in a second round of univariate multi-level models:

$$\hat{\varepsilon}_{t,p}^{(worry)} = \beta_{12p}^{(C)} \cdot \hat{\varepsilon}_{t,p}^{(relax)} + \beta_{13p}^{(C)} \cdot \hat{\varepsilon}_{t,p}^{(anger)} + \zeta_{t,p}^{(worry)}$$

The $\beta^{(C)}$ parameters are subsequently used to form the contemporaneous networks.

Box 3.3. Two-step multi-level graphical vector auto-regression.

3.4.2.1 Parameter covariance

A challenging aspect of multi-level modeling is that often covariances between random effects need to be estimated as well. For example, it could be that people that have strong edges between some variables also tend to have strong edges between other variables (Pe et al., 2015). Estimating sequential univariate models, as is done in two-step multi-level GVAR estimation, provides a computationally efficient alternative to estimating the multi-level GVAR model because the univariate multi-level models do not include all parameters; the estimation routine does not have to include

16 With too few observations per person, the estimated network structures will likely be biased. This bias is termed *Nickel's bias* (Jordan et al., 2020), and most notably leads to erroneous negative auto-regressions (self-loops in the temporal network).

all potential covariances between random effects.[17] More specifically, univariate models only include the intercept and (incoming) edge-weights that are connected to the dependent variables. In *mlVAR*, the covariance between these included random effects can be estimated by using the arguments temporal and contemporaneous. The default (correlated) will include correlated random effects, which is feasible for up to about 8 to 10 nodes. For about 10 to 20 nodes, uncorrelated (orthogonal) random effects can be used, which introduces a limitation to the estimation procedure, as random effects can be assumed to be correlated. Figure 3.4 shows that estimating models with no correlated random effects will only assume some random effect covariances to be 0, not all (as not all random effects are included in each model).

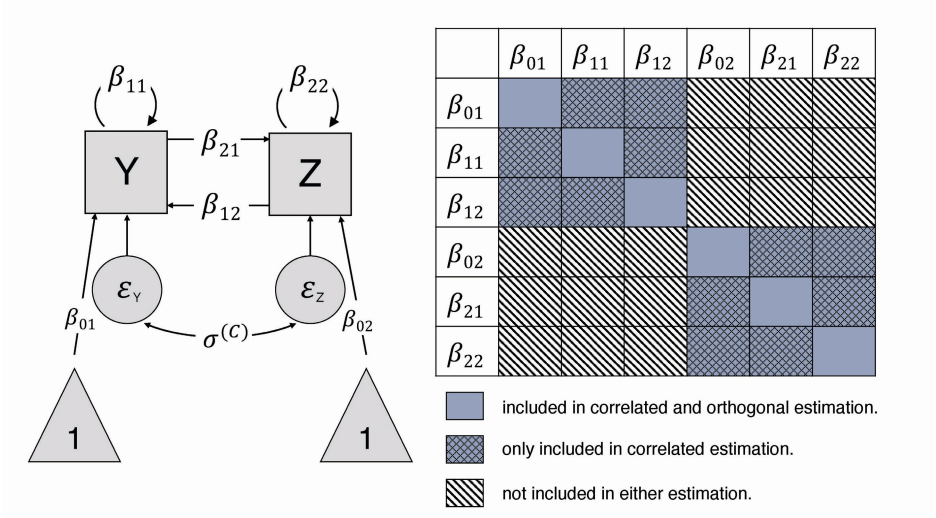


Figure 3.4. Parameter covariation included in *mlVAR* for the estimation of temporal effects. Choosing orthogonal estimation assumes that parameters are independent, whereas correlated estimation considers some (but not all) parameter correlations.

17 The upside of not having to estimate covariances between random effects also comes with the downside of not being able to investigate these covariances. This is why the between-person effects are estimated through level 2 predictors rather than by studying the random effects covariances between means.

We can estimate the multi-level VAR model from $N > 1$ time series data using the *mlVAR* package and from panel data using the *psychonetrics* package (fixed effect networks only). Both packages can handle data as structured in Box 3.4. The *mlVAR* package uses the mlVAR function, which can be used as follows (assuming a column in the data containing information on the subject id is called subject):

```
library("mlVAR")
mlVAR_results <- mlVAR(data, vars = vars, idvar = "subject", temporal = "correlated",
 contemporaneous = "correlated")
```

Optionally the dayvar and beepvar arguments can be used which work similarly as in Box 3.4. For non-correlated random effects, the temporal and contemporaneous arguments can be set to orthogonal. The plot method can be used to threshold and visualize the network. For example, the following command plots the temporal network with non-significant effects hidden:

```
plot(mlVAR_results, "temporal", nonsig = "hide")
```

Replacing plot for getNet will return the weights matrix instead. Networks showing the standard deviation of random effects can be obtained by setting SD = TRUE in plot(...) or getNet(...).

In *psychonetrics* the model can be estimated using the ml_gvar function:

```
library("psychonetrics"); library("dplyr")
ml_gvar(data, vars = vars, idvar = "subject", standardize = "z") %%
runmodel
```

Optionally, the beepvar argument can be used (the dayvar argument is not supported because this model is not designed for intensive time series), and further model search functions can be applied such as prune and modelsearch. Standardizing data is recommended to improve estimation. The contemporaneous network is stored as omega_zeta_within and the between persons network is stored as omega_zeta_between. If data are encoded in a wide format (variables encoded as different columns for each measurement), the panelgvar function can be used instead. Both ml_gvar and panelgvar are wrapper functions on the main dlvm1 function used for panel data modeling, which allows for some more options (e.g., modeling between-person effects as a Cholesky decomposition, which can be useful if between-person networks are seemingly not estimated well).

Box 3.4. Estimating multi-level GVAR models from $N > 1$ longitudinal data.

3.4.3 Multivariate estimation

3.4.3.1 Panel data

A considerable downside of the two-step multi-level GVAR estimation algorithm is that due to within-person centering with the person-wise sample means, a decent number of observations per person is required. To this end, it is not recommended to use this algorithm with less than about 20 observations, making it applicable to $N > 1$ time series data but not to panel data. One estimation method for estimating a GVAR model from panel data (also termed *panel GVAR*) has been proposed by (Epskamp, 2020b), and is implemented in the *psychonetrics* package, further

described in Box 3.4.[18] This model is a multi-level GVAR model with only random intercepts/means. This means that it assumes the same network structure for every person but allows people to differ on their averages. The variance–covariance structure of these random means is used to model the between-person network. The implementation in *psychonetrics* is a full-information implementation, meaning that all covariances between every possible measurement are included in the model. This makes the model computationally challenging to use with many nodes or many time points. It is recommended not to use this model with more than 10 measurements and more than around 10 to 20 nodes.

3.4.3.2 Bayesian estimation

A final powerful method for multivariate multi-level (G)VAR estimation is Bayesian estimation through sampling procedures. In these frameworks, all effects can be random and are included in the same model. As such, these methods return all possible random effect correlations (e.g., also allowing for between-person networks to contain edges). While it is possible to implement this model manually in software such as *JAGS* (Plummer, 2003) or *STAN* (Carpenter et al., 2017), doing so is quite challenging, requiring many prior distribution choices and likely leading to long computations if more than a few nodes are modeled. The Mplus software includes a module on dynamic structural equation models from version 8 onwards, which simplifies this process (Asparouhov et al., 2018; McNeish & Hamaker, 2020). This framework accommodates the multi-level VAR model, and while modeling GGMs is not included, partial correlation coefficients can manually be obtained from posterior samples (Epskamp, Waldorp, et al., 2018). The main downside is that even though the implementation in Mplus is very powerful, the number of nodes that can realistically be included in the analyses is still quite limited (about 6). Another downside is that Mplus is not open-source and, therefore, not free to use. A more detailed discussion on differences between Bayesian estimation and the two-step multi-level GVAR algorithm can be found in (Epskamp, Waldorp, et al., 2018).

3.5 Challenges to GVAR estimation

In this section, we discuss some of the most prominent practical and methodological challenges that researchers may face when estimating GVAR models from data.

3.5.1 Power and feasibility

The required number of observations to estimate reliable networks from time series data of a single subject ($N = 1$) is at least comparable to the number of participants needed to estimate networks from single measurement data. In fact, the required number may even be higher, as the GVAR

18 The panel GVAR model is implemented as a special case of a larger modeling framework that also includes latent variables. Epskamp (2020b) termed this model the *panel-LVGVAR* model, and the *psychonetrics* package terms this model the dlvm1 (dynamic latent variable model with lag-1) model. The panel GVAR can be obtained by representing each observed variable with a latent variable, setting all factor loadings to 1, and all residual variances to 0. This is done automatically in the *psychonetrics* package if no latent variable structure is assigned.

model includes a temporal network and is estimated from data with auto-correlated responses (reducing effective sample size). Collecting large time series for a single subject, however, is challenging in most fields of psychological research. Not only can it be burdensome for participants, but extending the measurement to long periods may also hinder the assumption of stationarity (see section 3.5.4 Stationarity assumption). Furthermore, the number of time points needed depends on the estimated network structure and on the number of nodes included. Sparse and well-defined network structures containing a few strong edges can be retrieved with smaller samples than dense networks with fewer strong edges that stand out.

In a series of simulation studies, Epskamp, Waldorp, et al. (2018) showed that reliable estimation of sparse synthetic networks is possible with 100 time points and eight nodes. However, a recent simulation study by Mansueto et al. (2022) used empirical networks as generating structures and instead found a relatively poor sensitivity (power to detect edges) with around 100 observations. A different generating network with 6-nodes led to better recovery of the global structure at 100 observations, but weaker edges were still not reliably retrieved. Such weak edges do not necessarily represent small and negligible effects; edges may be weak because of sampling bias or slight variations of the variables in time and may still be relevant for research or clinical purposes. Unfortunately, there is no way to know if the generating structure (assuming data were generated through a GVAR model) was dense or sparse and included strong or weak edges. As such, it is questionable if GVAR estimation is feasible from $N = 1$ data sets that may realistically be obtainable, and this will rely on certain assumptions (notably, that the generating model is sparse). It is advisable to consider that with about 50 to 100 time points, sensitivity likely is low, meaning that only a few edges may be discovered. The best solution to this problem, outside of aiming to collect more data, is to keep the model as small as possible. For $N = 1$ GVAR models, it is generally advisable to include as few nodes as possible (e.g., less than 10).

3.5.2 Heterogeneity

In addition to discovering individual network structures, researchers may also be interested in how much people differ in their network structure (heterogeneity). The detection of heterogeneity between GVAR models is directly related to the reliability of GVAR estimation. If it is not feasible to estimate reliable network structures, visually comparing network structures of individuals may lead to an illusionary sense of heterogeneity. Hoekstra et al. (2022) discuss that even if the generating structure is the same for two people, network structures estimated from their data may differ substantively. For example, suppose that the generating model contains 10 (true) edges, but sensitivity (power) is only 50%, meaning that we only expect to find 5 out of 10 edges in the network of one particular person. Suppose also that the chance of including an edge is the same for all 10 true edges. Then, there is only a 0.000016 probability that the *exact* same edges are detected in two people. As such, even though the generating structure is the same, we would expect to find different networks. This entails high sensitivity (and specificity) are needed to separate true from illusionary heterogeneity when estimating individual network models, and to this end, it is advisable to not interpret differences in personal network models as evidence for heterogeneity, especially when these networks are sparse. If a large number of people are included in the data set, multi-level modeling can be used to gain insight in the heterogeneity of parameter values. When estimating

a multi-level network using the *mlVAR* package, the standard deviations of random effects across the population on the temporal and contemporaneous network parameters are returned and can be visualized as networks (e.g., Figure 3.3). The width of the edges in this network shows the degree to which network parameters exhibit individual differences. Bringmann et al. (2013) recommend using a cut-off score of 0.10 for the edge weights. Alternatively, random effects can in principle be tested statistically by comparing a model with random effects to a fixed effects only model, although this may be hard in practice.[19]

3.5.3 Missing data

In intensive time series designs missing data are very common. Usually, time series data are characterized by wave missingness, where every item at a particular measurement point is missing (McLean et al., 2017; Schafer & Graham, 2002). Many factors can affect missing data, for example, measurement frequency and timing, length of the measurement period, physical activity, substance use, and age (A. Jones et al., 2019; McLean et al., 2017; Ono et al., 2019; Rintala et al., 2019; Wen et al., 2017). Techniques based on imputation or maximum likelihood can be used to handle data missing completely at random (MCAR) and at random (MAR), while with data missing not at random (MNAR), these may yield biased estimates. For example, Kalman filter imputation (Hamaker & Grasman, 2012; A. C. Harvey, 1990) can be applied prior to network estimation with the R package *imputeTS* (Moritz & Bartz-Beielstein, 2017). Alternatively, full information maximum likelihood estimation is implemented in the R package *psychonetrics*, which estimates a model using only observed responses (Epskamp, Isvoranu, et al., 2022). Finally, Bayesian estimation methods are well capable of handling missing data (McNeish & Hamaker, 2020). Mansueto et al. (2022) propose that such methods for handling missing data may lead to promising avenues for reducing participant burden through the use of planned missingness and adaptive testing – only asking a subset of questions in each measurement (Graham et al., 2006; Schafer & Graham, 2002).

3.5.4 Stationarity assumption

As any statistical model, (multi-level) GVAR estimation relies on a set of assumptions. A core assumption of the VAR model is *stationarity*. A stationary time series does not indicate changes over time in its defining characteristics, such as the means, variances, and network parameters. Violations of this assumption arise if there are trends in the time series, for example, seasonal or linear trends or changes in volatility. Deviations from the stationarity assumption are not implausible in psychological time series. For example, we might observe mean-shifts in symptoms following certain life events or obtain seasonal patterns for affect variables depending on the time of the year. To understand the dynamic process without such trends, a time series can be broken down into its constituent trend components through a method called *decomposition*, resulting in a trend component, a seasonal/cyclical component, and a residual/regular component. Another scenario introducing non-stationarity is the presence of a so-called *unit root*. A unit root is present if the auto-regressive parameter of a time series equals one, and can be detected using the (Augmented-)

19 A limited implementation for testing random effects of temporal coefficients is implemented in *mlVAR* in the mlVARcompare function.

Dickey-Fuller test (Dickey & Fuller, 1979). Figure 3.5 visualizes a stationary distribution, as well as different cases of non-stationarity.

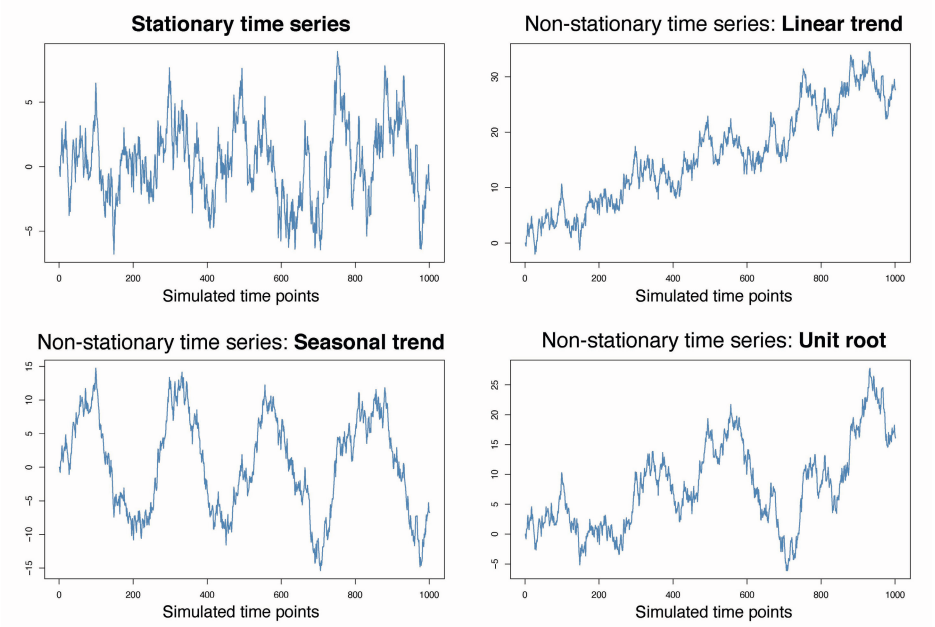


Figure 3.5. Simulated time series data under four conditions. Top left: A stationary time series with auto-regressive effect $\beta = 0.95$. Top right: Non-stationarity, due to a linear trend added to the time series used for the first plot. Bottom left: Non-stationarity due to a seasonal trend added to the time series used for the first plot. Bottom right: Non-stationarity due to the presence of a unit root, time series with auto-regressive effect $\beta = 1$.

There are several methods of handling non-stationarity. Generally, most of these methods aim to remove existing linear trends or seasonal components. For example, linear trends in the data can be accounted for by performing a regression on time and subsequently modeling the residuals as the time series adjusted for linear changes. Simulation studies showed that it is generally recommended to detrend present linear trends before estimating networks (Epskamp, Van Borkulo, et al., 2018). In this study, detrending all versus only significant trends performed comparably, while not detrending led to lower specificity (especially in temporal networks) and lower sensitivity (especially in contemporaneous networks). However, in many situations changing means, variances, or network parameters are of central interest, and therefore removing them from the data through detrending would be detrimental. Instead, these changes over time can be explicitly modeled in time-varying network models, which are further discussed in (Haslbeck, Ryan, et al., 2022).

3.5.5 Assumption of equidistant measures

The GVAR model establishes temporal dependencies in the form of lagged relationships. In doing so, it treats all lags of the same level as equally distant from one another. In other words, it is assumed that the time difference between any two subsequent assessment points is equal, an assumption referred to as *equidistant measures*. Two primary manners in which this assumption may be violated

are (1) when there are multiple measures per day, because there will likely be a larger time difference between the last assessment of a day and the subsequent day's first assessment compared to time differences between other consecutive measurements, and (2) when a participant failed to fill in all measurements, and the data is not properly encoded with missing values on the measurements that were not filled in. These problems can adequately be handled using the software packages discussed in this chapter through the use of the dayvar argument (removes pairs of observations that cross a night) and beepvar arguments (removes pairs of responses that are not consecutive). Another way this assumption may be violated is (3) when measurement occasions are at random time intervals. While it should not be a big problem if time intervals are roughly equal (e.g., sometimes 2 hours and sometimes 3 hours), it may be problematic if time intervals show large differences (e.g., sometimes 10 minutes and sometimes 4 hours). In this setting, an alternative is to use continuous time modeling, further discussed by (Ryan & Hamaker, 2022).

3.5.6 Time scales

In the GVAR model, we aim to predict dynamics as lagged relationships, typically including only one time lag. Consequently, to interpret temporal effects, we need to make sure that the time scale chosen for our analysis matches the type of dynamics we want to investigate. For example, if we want to model a temporal effect A → B, we want to ensure that we also capture this effect by appropriately timing our assessment intervals. This, however, is not always possible or feasible; in many cases, we either do not know the true time scale our processes are operating at, or it is not feasible to measure at the desired frequency. A mismatch between true time scale and modeled lags can lead to problematic inferences in two situations: first, the true dynamic process can unfold *faster* than specified in our assessment (e.g., panic symptoms occur within seconds, but we measure every two hours). In this case, the effects will not be captured in the temporal prediction. Such fast effects might be found in the contemporaneous rather than the temporal effects (Epskamp, Van Borkulo, et al., 2018). An alternative approach to modeling discrete time lags is to conceptualize dynamics on a continuous level, for example, using continuous structural equation modeling (Driver et al., 2017; Ryan & Hamaker, 2022) or differential equations. Second, the true dynamic process can unfold *slower* than specified in our assessment (e.g., investigating mood dynamics in relation to hormones, but we assess hormone levels every two hours). Such slower effects might be better understood using panel designs (Epskamp, 2020b) because they require more distance between assessments.

Epskamp (2020b) details multivariate estimation of the GVAR model from $N = 1$ time series and panel data. To estimate a GVAR model, the data can first be augmented:

$$Y^{(aug)} = \left[Y^{(lag)} \quad Y\right],$$

in which $Y^{(aug)}$ represents the augmented data set, Y represents the original data (with measurement t on row t), and $Y^{(lag)}$ the original data set shifted by one row (measurement $t - 1$ on row t). If needed, several rows of $Y^{(aug)}$ can be removed, especially when the pair of measurements $t - 1$ and t feature a large gap in time, such as across a night. The variance-covariance matrix of $Y^{(aug)}$ takes the following form:

$$\Sigma = \begin{matrix} \\ t-1 \\ t \end{matrix} \begin{matrix} \begin{matrix} t-1 & t \end{matrix} \\ \begin{bmatrix} \Sigma^* & \Sigma_1^{\mathrm{T}} \\ \Sigma_1 & \Sigma_{0.} \end{bmatrix} \end{matrix}$$

Also termed the Toeplitz variance-covariance matrix. The block Σ_0 can be modeled with the following expression:

$$Vec(\Sigma_0) = (I \otimes I - B \otimes B)^{-1} Vec\left(\Sigma^{(C)}\right)$$

In which $\Sigma^{(C)}$ is the innovation variance-covariance matrix that can further be modeled as a GGM (see Box 3.1). The lag-1 variance-covariance matrix can subsequently be modeled as:

$$\Sigma_1 = B\Sigma_0$$

Finally, with large samples we would expect $\Sigma^* = \Sigma_0$ In the *psychonetrics* package, however, Σ^* is modeled using Cholesky decomposition instead:

$$\Sigma^* = LL^T$$

Such that this block is always positive semi-definite and such that the stationary variance-covariance structure is not modeled twice. This box explains the main expressions used in maximum likelihood estimation of $N = 1$ GVAR models. The variant for panel data follows mostly the same steps, but creates a larger Toeplitz matrix with all waves of data and models between-person variance in addition to the within-person variances discussed here.

Box 3.5. Toeplitz variance–covariance structure for the GVAR model.

3.6 Conclusion

Time series analysis is a fruitful field for constructing dynamical networks. The graphical vector auto-regressive (GVAR) model separates longitudinal information in *contemporaneous*, *temporal*, and *between-persons* network structures. This chapter discussed how these networks could be estimated from time series analyses for single subjects and multiple subjects, using intensive longitudinal data collected via novel ambulatory assessment techniques or panel data. Current challenges in estimating time series networks span from a trade-off between power, stationarity, and feasibility to identifying appropriate time scales for lagged relationships.

While this chapter provides an introduction to network analysis from longitudinal data, the topic of longitudinal data analysis itself goes beyond the scope of this book. Indeed, entire textbooks could be written on this topic. Important to note is that the GVAR model, which was the focus of this chapter, is only one of several possible models. Another approach to constructing networks from time series data is the estimation of a structural VAR model (SVAR; Chen et al., 2011; Gates et al., 2010). In contrast to GVAR, the SVAR model uses *directed* effects for the contemporaneous

network. Structural VARs can be estimated by transforming (G)VAR results (Lütkepohl, 2005) or through unified structural equation modeling (Beltz & Molenaar, 2016; Gates et al., 2010; Kim et al., 2007), and structure estimation is usually done through step-wise model search. In $N > 1$ data, *group iterative multiple model estimation* (GIMME; Gates & Molenaar, 2012) is an often-used method for estimating structural VAR models for multiple persons. In short, GIMME searches for qualitative similarity across people – using step-wise model search strategies through structural equation models – to find network structures that contain group-level (edges that are included for every person in a group) as well as person-specific temporal and contemporaneous effects. Other variants of network models estimated from time series data are time-varying VAR networks and VAR networks that include non-Gaussian variables. These are discussed elsewhere (Haslbeck et al., 2022).

CHAPTER 4

USING NETWORK ANALYSIS TO EVALUATE MENTAL HEALTH INTERVENTIONS

CHAPTER 4

Abstract

The network approach to psychopathology, which assesses associations between individual symptoms, has recently been applied to evaluate treatments for mental disorders. While various options for conducting network analyses in intervention research exist, no guidelines have been established. To evaluate the potential of different analytic options, we conducted a review by searching the literature with combining terms on *network analysis*, *mental health problems*, and *intervention studies*. Studies were included if they constructed a symptom network, analyzed data that was collected before, during or after a treatment for mental disorders, and yielded information about the treatment effect. Across the 56 included studies, network analyses varied widely. About half of the studies estimated cross-sectional networks without a treatment node, about 20% of studies analyzed cross-sectional networks including a treatment node and a third of the studies estimated longitudinal networks. Studies differed on how networks were estimated, which network parameters were calculated, and which statistical tests were applied. This chapter highlights that current methodological practices limit the information that can be gained through network analyses and that methodological advances are needed to unleash the full potential of the network approach in intervention research. Analytic options need to be further investigated and structured guidelines need to be developed.

This chapter has been adapted from: Schumacher, L., Burger, J., Echterhoff, J., & Kriston, L. (2022). Methodological and statistical practices of using symptom networks to evaluate mental health interventions: A systematic review. *PsyArXiv Preprint, under review.*

4.1 Introduction

About 10 years ago, the network approach to psychopathology was proposed by Borsboom and Cramer (2013). This approach suggests that mental health problems develop and are sustained by symptoms mutually activating each other and defines mental disorders as networks of interacting symptoms (Borsboom, 2017). Next to describing the symptomatology of a specific client group with symptom networks, it was suggested to apply the network approach to plan and evaluate treatments for mental disorders (Blanchard & Heeren, 2022; McNally, 2016). In this framework, treatment effects are discussed in regards to the treatment's impact on symptom networks. More specifically, interventions may change the severity of specific symptoms, the interactions between symptoms or impact symptom-triggering variables in the external field (Borsboom, 2017). So far, interventions for mental disorders have mostly been evaluated by analyzing the presence/absence of a diagnosis or a composite score indicating the aggregated severity of several symptoms. Additional information could be gained with symptom networks since these allow the analysis of the treatment effect on specific symptoms and symptom associations. Studying treatment effects on the symptom level seems promising as, first, the effects of treatments for mental disorders might be symptom specific, i.e. some symptoms are positively affected while others are not (e.g. Bekhuis et al., 2018), and solely focusing on composite-scores or the presence of a diagnosis cannot reveal such symptom-specific effects (Kaiser, Herzog, et al., 2021); second, large variations in symptom-expressions have been observed for individuals with the same diagnosis, therefore, a diagnosis might not be a good description of the experienced problems of the target population (Fried & Nesse, 2015); and third, when investigating symptom networks over time throughout the treatment, possibly on the level of an individual, the changes in symptom associations might give some insights into the working mechanisms of the treatment (Hofmann et al., 2020). Thus, using symptom networks to evaluate mental health interventions could potentially broaden the knowledge on treatment effects by focusing on individual symptoms and their relations.

Statistical methods were developed to estimate symptom networks from empirical data (Bringmann et al., 2013; Epskamp & Fried, 2018). Such estimated networks consist of nodes indicating observed symptoms, and edges which show statistical relationships between the symptoms (Epskamp, Waldorp, et al., 2018). In the last years, network analysis has been frequently applied to investigate the symptomatology of a specific client group (Robinaugh et al., 2020). Here, network analysis often assesses the strength of edges, i.e. how strongly a symptom relates to another symptom, the centrality of nodes, i.e. how strongly a symptom is associated with all other symptoms, and the overall connectivity, i.e. how strongly all symptoms are, at average, associated with each other. Additionally, researchers have started to use network analysis to evaluate treatments for mental disorders. For example, symptom networks were compared before and after treatment (e.g. Kaiser, Boschloo, et al., 2021), or between treatment groups (e.g., Blanco et al., 2020). It was also suggested to add a treatment node in the network, which indicates the allocation to a treatment group, a control group or an alternative treatment (Blanken et al., 2019).

However, so far, there is no standard procedure or structured guidelines available on how network analysis should be applied to assess intervention effects. At the same time, there are many analyses options for network analysis in intervention research. Firstly, cross-sectional or longitudinal

networks with different underlying models can be constructed. Cross-sectional networks that display between-person symptom associations can be estimated using Graphical Gaussian Models (GGMs, for continuous variables; Epskamp, Waldorp, et al., 2018), Ising models (for binary variables; Finnemann et al., 2021) or Mixed Graphical Models (MGMs, for mixed variables; Haslbeck & Waldorp, 2020). Alternatively, longitudinal networks that investigate temporal within-person symptom associations can be estimated using multilevel Vector-Autoregressive Models (mlVAR; Bringmann et al., 2013). Other models and adaptions of these models are possible. Further, results highly depend on what (kind of) variables are included as nodes in the network (Bringmann et al., 2022). Researchers need to decide which symptoms to include and if non-symptoms such as moderating variables and/or a treatment node as proposed by Blanken and colleagues (2019) should be included. The kind of variable, e.g. if nodes constitute an absolute or a change score for a symptom, also strongly influences the interpretation. Finally, there are various ways to analyze the estimated networks. Different network parameters can be calculated, different comparisons can be made (e.g. between treatment groups, at different time points or between treatment responders and non-responders), and different statistical analyses can be conducted.

The potential of network analysis in intervention research is likely to largely depend on such analytic choices. We realize that different analytic procedures are probably valuable for different research questions and contexts. Still, to our knowledge, there is no consensus on which methodologies are best to choose for the evaluation of treatment effects for different research questions and not even an overview of which analyses choices have been previously made is available so far. Therefore, we systematically reviewed intervention studies that used network analysis to evaluate treatments for mental disorders. Through this review, we aimed to gain an overview of the employed methods and assess benefits and drawbacks of these. This can inform future studies using the network approach, or even structured guidelines for evaluating interventions for mental health problems and hopefully increase the value of the network approach in intervention research.

4.2 Methods

4.2.1 Study search

We searched three bibliographic databases (PsycINFO, MEDLINE, and Web of Science) for intervention studies that utilized symptom network analyses. The title, abstract, keywords, and subject headings were explored with combining search terms from three categories: (1) network analysis as the method of data analysis, (2) intervention study as the study design, and (3) individuals with mental health problems as the target population. The specific search terms can be found in the supplementary materials,[20] Text S1. Additionally, we performed forward and backward reference search for the included studies and searched Google Scholar with the term "network intervention analysis". Finally, we checked the references of reviews on using the network approach in mental health and psychopathology.

20 All supplemental material can be found here: https://osf.io/n4xp5/files/osfstorage .

4.2.2 Inclusion criteria

We aimed to include studies with individuals, who received an intervention directed at mental health problems (population) and studies which assessed symptom-relations with network analysis (concept). We wanted to include of all kinds of interventions/treatments directed at mental health problems, all kinds of mental disorder, and of all kinds of control groups (i.e. open context). Thus, the following inclusion criteria were applied: (1) the study conducts a network analysis which investigates the relation among psychological symptoms (and possibly other variables), (2) the study analyses data which was collected before, after or during a treatment which was directed at psychological problems or mental disorders, (3) the analysis yields information in regards to the effect of the treatment, (4) the study is published in a peer-reviewed journal. Network meta-analyses and network analyses in which nodes represented people or neural connections were excluded.

4.2.3 Study selection and data extraction

Titles and abstracts were screened with the above outlined inclusion criteria using the software Rayyan; 20% of the titles and abstracts were double-screened by two independent raters. The full-texts of all studies that were found eligible in the first step were examined by two independent raters regarding the decision to include the study. Disagreement between raters was resolved by discussion. Information on the sample characteristics, the intervention, the research design, the estimated networks, the statistical analysis, and the use of open science practices were extracted and coded for all included studies. All variables are displayed in the supplementary materials, Table S1. The main variables describing the network estimation and further statistical analyses were extracted by two independent raters and disagreement was resolved by discussion.

This review was preregistered on the open science framework (https://osf.io/8txcy/?view_only=-833fa715381e470eaa24d93251457191). Our rationale for slightly adjusting the inclusion criteria can be found in the online supplemental materials, Table S2. All needed materials (exact search terms, inclusion criteria, excluded studies, extracted data) are publicly available in the online supplemental materials.[21] This study is reported according to the PRISMA-ScR guidelines.

4.3 Results

4.3.1 Search results

The search in the bibliographic databases in December 2021 yielded 4,519 records, of which 4,298 remained after deduplication. After title and abstract screening, the full-texts of 39 studies were screened and 34 studies met all inclusion criteria. The interrater-reliability was $\kappa = 0.89$ and $\kappa = 0.86$ for the abstract screening and the full-text screening, respectively. The additional search (forward/backward reference search, Google scholar search, and search in network reviews) in April 2022 yielded the inclusion of additional 22 studies, leading to a total of 56 included studies. Of note, such high turn-out of the additional search could be expected due to the diverse network terminology and the recency of these studies. A detailed overview of the study selection procedure is displayed

21 All supplemental material can be found here: https://osf.io/n4xp5/files/osfstorage .

in Figure 4.1. The references of all included studies and an overview of all excluded studies can be found in the online supplemental materials Text S2, and Table S3, respectively.

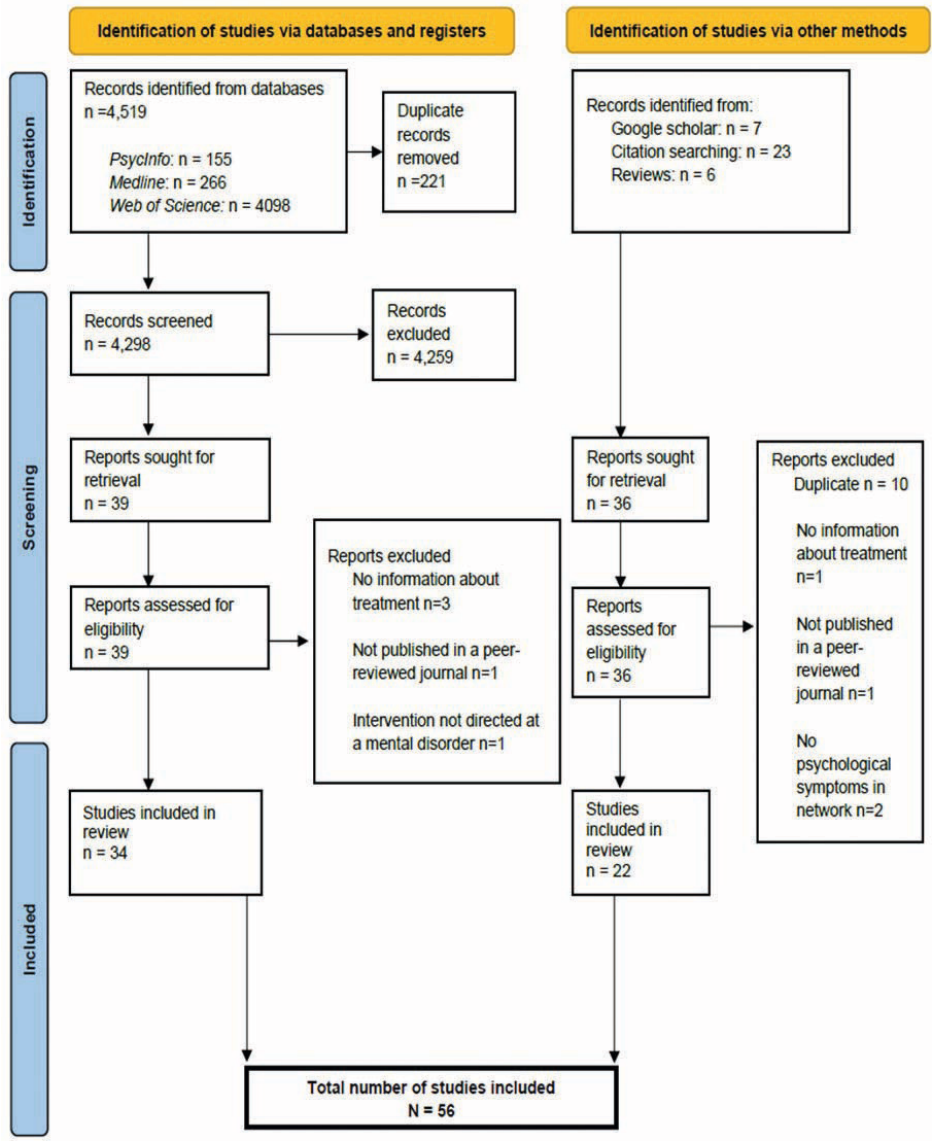


Figure 4.1. Flow chart for study selection.

4.3.2 Study characteristics

Most studies were conducted in Europe (n = 24, 42.9%) or Northern America (n = 16, 28.6%) between 2015 and 2022, with the majority of studies being published between 2020 and 2022 (n = 33, 58.9%). Across all studies, the average age of the samples had a median of 41.4 years and

the proportion of females had a median of 65.7%. The most often investigated client group were persons with depression symptoms or a diagnosis of a depressive disorder (n = 29, 51.8%). About a third of the included studies evaluated some form of cognitive behavioral therapy (CBT) and 21.4% evaluated antidepressants. The interventions had a mean length of 13.1 weeks, ranging between 2 and 52 weeks. About half of the studies classified as randomized controlled trials (RCTs, n = 29, 51.8%), the majority of the other half was observational (n = 26, 46.4%)[22]. One study conducted an individual client-data meta-analysis of RCTs. The total sample size ranged from 1 to 5,614, with a median of 229. The network analyses were based on data which included a median of three assessments per person, ranging between 1 and 120 assessments for each person. Half of all studies (n = 28, 50%) did not report how missing data was handled. Twenty studies (35.7%) indicated that there was no missing data or excluded participants with missing data. In four studies (7.1%) imputation methods were used to handle missing data. Study characteristics of all studies are displayed in Table S1 in the online supplemental material.

4.3.3 Network analysis

Most studies estimated two (n = 9, 16.1%), three (n = 10, 17.9%) or four (n = 13, 23.2%) separate networks. The sample sizes for the individual networks ranged from 1 to 5,614, with a median sample size of 151. The number of nodes for each network varied between 5 and 47, median = 12. The majority of studies did not report how they decided on which nodes to include in the network (n = 41, 73.2%). The nodes constituted absolute item scores in 46 studies (82.1%), absolute composite scores in 15 studies (26.8%), change item scores in eight studies (14.3%) and change composite scores in two studies (3.6%)[23]. Ten out of the 56 studies included non-symptoms other than "treatment" as nodes in the network. All studies estimated networks with weighted edges and edges were directed (as opposed to undirected) in the networks of 14 of the 56 studies. Forty-eight studies (73.8%) calculated network parameters to describe the networks. The three most common network parameters were node strength/ degree (n = 34, 60.7%), node closeness (n = 21, 37.5%) and node betweenness (n = 19, 33.9%). A list of all network parameters can be found in the supplemental material, Table S4. Of the studies reporting network parameters, 17 (51.5%) performed case-dropping bootstrap to evaluate the stability of the network parameters.

The majority of studies (n = 34, 60.7%) estimated cross-sectional networks without a treatment node. Cross-sectional networks including a treatment node were estimated by 21.4% of the studies (n = 12), and longitudinal networks were estimated by 32.1% of the studies (n = 18).[24] Estimation methods and statistical analyses will be reviewed for each of the three types of networks separately. The described network characteristics for every study can be found in Table S1 in the supplemental materials.

---

22 Data from five of these observational studies were originally from an RCT. However, data from just one treatment group was used in the reported analyses. Thus, we classified such studies as observational.

23 Some studies used several types of nodes.

24 Eight studies estimated different types of networks. Therefore, these studies are mentioned in both corresponding categories.

4.3.3.1 Cross-sectional networks without a treatment node

Out of the 34 studies on cross-sectional networks without a treatment node, 29 studies estimated a Graphical Gaussian Model (GGM). Three studies constructed a cross-sectional network from a multilevel VAR model, and two studies estimated an ISING model.[25] In most networks, edges were defined as partial correlations regularized by the LASSO (n = 23, 67.6%). In all networks, edges indicated pairwise associations. Fifteen studies (44.1%) examined the stability of edge weights using non-parametric bootstrap. One study compared their dataset characteristics with a previous simulation study to gauge the stability of the network. In seven studies (20.6%), the nodes were defined as changes scores, mostly describing the (estimated) change from pre- to post-treatment. The following R packages were used for estimating the networks: *qgraph*[26] (Epskamp et al., 2012), *mlVAR* (Bringmann et al., 2013a), *psychonetrics* (Epskamp, 2022), *NetworkX* (Python; Hagberg et al., 2008).

Nearly all studies on cross-sectional networks without a treatment node estimated several networks and compared them with each other (n = 30, 88.2%). Twelve studies (35.3%) compared networks across time, often before and after treatment. In nine of these studies an observational design was employed, i.e. all participants received the same intervention. In three studies, the data came from an RCT and data from both (treatment) groups were combined for the network analysis. Six studies (17.6%) compared networks between different response groups, e.g. from clients who did versus did not remit. Five studies (14.7 %) investigated network differences between treatment groups (four RCTs, one individual client data meta-analysis). Seven studies (20.6%) performed multiple comparisons. Of all 30 studies comparing networks, six studies compared networks only visually and 17 studies used the Network Comparison Test (van Borkulo et al., 2022). All details on networks and statistical analyses for cross-sectional networks without a treatment node can be found in Table 4.1.

Table 4.1. Network analysis characteristics of studies estimating cross-sectional networks without a treatment node.

Study	Study design	N of networks	N sample for networks	Model	Comparison	Statistical tests
Beard et al. (2016)	observational	2	1029/742	GGM	timepoints of the same group	NCT for global network strength, correlation between edge weights and centrality indices

25 Note that seven studies did not report the underlying model explicitly, but based on their description we assumed a GGM for six studies and an ISING model for one study.

26 One study refers to "graphics" but cites qgraph.

Table 4.1. Continued

Study	Study design	N of networks	N sample for networks	Model	Comparison	Statistical tests
Berlim et al. (2021)	RCT	6	151/ 151/ 74 /74 / 77/ 77	GGM	timepoints and groups receiving different treatment	permutation test for global strength invariance, network structure invariance and edge strength invariance
Blanco et al. (2020)	RCT	4	48/ 48/ 45/ 45	GGM	timepoints and groups receiving different treatment	NCT
Bos et al. (2018)	observational (original study was an RCT)	2	178/ 178	GGM	timepoints of the same group	NCT
Boschloo, Cuijpers, et al., (2019)*	RCT	2	399/ 395	n.r. (assumed GGM)	groups receiving different treatment	NCT
Boschloo, Bekhuis, et al., (2019)*	individual patient data meta-analysis of RCTs	2	500/ 570	n.r. (assumed GGM)	groups receiving different treatment	NCT for global connectivity and individual connections
Briganti et al. (2021)*	observational	3	100	GGM	timepoints of the same group	NCT for global strength
Calugi et al. (2021)	observational	2	214	GGM	timepoints of the same group	NCT
Carney et al. (2018)	observational	2	77/ 48	n.r. (assumed GGM)	groups defined by treatment response	none
Curtiss et al. (2021)	RCT	4	94/ 64/ 64/ 64	GGM	timepoints of the same group	none

Table 4.1. Continued

Study	Study design	N of networks	N sample for networks	Model	Comparison	Statistical tests
Elliott et al. (2020a)	RCT	5	142/ 119/ 113/ 105/ 142	GGM	timepoints of the same group	linear regression to test if symptoms with higher expected influence have a higher prognostic value
Esfahlani et al., (2017)	RCT	4	733/ 316/ 733/ 316	n.r. (assumed GGM)	timepoints and groups defined by treatment response	Kolmogorov-Smirnov test for differences in closeness and degree centrality
Esfahlani et al., (2018)	RCT	2	316/ 733	GGM	groups defined by treatment response	supervised machine learning to predict treatment responders with network parameters
Goldberg et al. (2020)	RCT	1	208	GGM	n.a.	various analyses to test if symptoms identified in the network analyses can predict treatment effects
Hilbert et al. (2020)	observational (original study was an RCT)	3	178/ 178/ 178	GGM	timepoints of the same group	NCT of global network strength, exploratory graph analysis for communities, predicting treatment response by network connectivity

Table 4.1. Continued

Study	Study design	N of networks	N sample for networks	Model	Comparison	Statistical tests
Hoffart et al. (2019)*	observational	1	65	mlVAR model	n.a.	none
Hoffart & Johnson (2020)*	RCT	1	60	mlVAR model	n.a.	none
Johnson & Hoffart (2018)	RCT	6	38/ 38/ 38/ 36/ 36/ 36	mlVAR model	groups receiving different treatment	none
Kaiser, Herzog, et al., (2021)	observational	2	5614/ 5614	GGM	timepoints of the same group	NCT
Komulainen et al. (2021)*	RCT	2	1033/ 2526	n.r. (assumed ISING model)	groups receiving different treatment	none
Kraft et al. (2019)*	RCT	4	149 / 153/ 149/ 153	GGM	groups receiving different treatment	NCT for global strength and density
Levine & Leucht (2016)	RCT	3	437/ 437/ 437	GGM	timepoints of the same group	t-tests to compare number of positive and negative edges, permutation-based tests (not specified)
Lorimer et al. (2020)	observational	4	91/ 769/ 93/ 774	GGM	timepoints and groups defined by treatment response	none
Lydon-Staley et al. (2020)	observational (original study was an RCT)	4	523/ 496/ 457/ 426	GGM	timepoints of the same group	NCT
Madhoo & Levine (2016)	observational	3	2862/ 2585/ 2578	GGM	timepoints of the same group	NCT for global connectivity

Table 4.1. Continued

Study	Study design	N of networks	N sample for networks	Model	Comparison	Statistical tests
McElroy et al. (2019)	observational	9	566 / 2277 / 174	GGM	groups defined by treatment response	NCT for global-strength and structural invariance at baseline
O'Driscoll et al. (2021)	observational (original study was an RCT)	3	2858/ 956/ 1466	GGM	groups defined by treatment response	NCT for global connectivity and centrality parameters
Olatunji et al. (2018)	observational	4	5193/ 5193/ 2876/ 2301	GGM	timepoints and groups defined by treatment response	multiple regression to predict treatment outcomes from central symptoms
Papini et al. (2020)	RCT	1	306	n.r. (assumed GGM)	n.a.	correlation between pre-treatment network parameters and the impact that change of a specific symptom had on change in the whole network
Park et al. (2021)	observational	5	1152/ 801/ 522/ 409/ 281	GGM	timepoints of the same group	none
Schweren et al. (2018)	observational	2	233/ 232	n.r. (assumed GGM)	groups defined by treatment response	permutation testing of differences in density and node strength
Scott et al. (2020)	observational	4	900/ 900/ 122/ 466	ISING	groups defined by treatment response	NCT

Table 4.1. Continued

Study	Study design	N of networks	N sample for networks	Model	Comparison	Statistical tests
Smith et al. (2019)	observational	4	446/ 446/ 223/ 223	GGM	timepoints and groups defined by treatment response	NCT of global network strength
Zhou et al. (2022)	observational	4	474/ 474/ 254/ 220	GGM	timepoints and groups defined by treatment response	NCT

Note. RCT = randomized controlled trial, GGM = Graphical Gaussian Model, NCT = Network comparison test, mlVAR = multilevel Vector Autoregression, n.r. = not reported, n.a. = not applicable. *These studies report two different kinds of symptom networks. Information here just relates to the cross-sectional networks without a treatment node.

4.3.3.2 Cross-sectional networks including a treatment node

Twelve studies estimated cross-sectional network including a node which indicates membership to a treatment group as suggested by Blanken and colleagues (2019) (often termed Network Intervention Analysis). All but one study estimated a mixed graphical model (MGM) to construct the networks[27]. In 10 out of the 12 studies, edges were defined as regularized nodewise regression coefficients and in two studies as partial correlations. Testing the stability of edge weights with nonparametric bootstrap was reported by eight of the 12 studies. The package mgm (Haslbeck & Waldorp, 2020) was mostly used (n = 10, 83.3%). Five studies (41.7%) estimated cross-sectional networks including a treatment node with nodes indicating pre- to post-treatment change scores. A network was estimated at various time points before, during and after the treatment by the remaining seven studies (58.3%). Here, no study conducted a formal comparison between the networks. All data came from RCTs or individual client data meta-analysis of RCTs. All analysis details of studies on cross-sectional networks including a treatment node are displayed in Table 4.2.

4.3.3.3 Longitudinal networks

The 18 studies on longitudinal networks used most commonly an mlVAR model (n = 11, 61.1%, of which three also included a time trend). Two studies used dynamic time warp distance matrices and one study cosine similarity measures to construct networks. One VAR model, one graphical vector auto-regressive (GVAR) model and one unified structural equation model were estimated. In 12 out of 18 studies (66.7%), the edges described lag-1 temporal associations. In the other studies, edges describe contemporaneous associations (n = 4, 22.2%), temporal associations between latent

27 Note that one study did not report the underlying model, but based on their description we assumed it was a MGM.

variables (n = 1, 5.6%), the similarity of time series between variables (n = 3, 16.7) or temporal associations between improvements (n=1, 5.6%). Several different (R) packages were used: *mlVAR* (Bringmann et al., 2013), *dtw* (Giorgino, 2009), *pheatmap* (Kolde, 2019), *parallelDist* (Eckert, 2022), *qgraph* (Epskamp et al., 2012), *nmle* (Pinheiro, 2022), *pompom* (Yang, 2021), NetworkX (Python; Hagberg et al., 2008), and Stata mixed commands. Stability was not assessed, except for one study (5.6%), where the characteristics of the dataset were compared to a previous simulation study. Five observational studies and one RCT (27.8%) investigated the temporal symptom associations with one longitudinal network over treatment (data from both treatment groups were combined for the RCT) and did not compare longitudinal networks with each other. The networks were compared between treatment groups in six studies (33.3%) and four observational studies compared longitudinal networks in different response groups (22.2%). Two RCTs (11.1%) compared different treatment groups at different timepoints. Seven out of the 12 studies (58.3%) comparing networks did not use formal tests to compare the networks, and the other studies used various statistical analyses such as permutation tests for edge differences for the comparison. All analysis variables are displayed in Table 4.3 for studies on longitudinal networks.

Table 4.2. Network analysis characteristic of studies estimating cross-sectional networks with a treatment node.

Study	Study design	N of networks	N sample for networks	Model	Comparison	Statistical tests
Bekhuis et al. (2018)	RCT	1	186	n.r. (assume MGM)	n.a.	none
Bernstein et al. (2019)	RCT	9	120/107/97/89/ 88/87/92/84/74	MGM	timepoints before, during and after treatment	none
Blanken et al. (2021)	RCT	6	143/ 133/ 126/ 118/ 113/ 105	MGM	timepoints before, during and after treatment	none
Blanken et al. (2019)	RCT	10	104	MGM	timepoints before, during and after treatment	none
Boschloo, Cuijpers, et al., (2019)	RCT	1	794	MGM	n.a.	none
Boschloo, Bekhuis, et al., (2019)	individual patient data meta-analysis of RCTs	1	1070	MGM	n.a.	none

Table 4.2. Continued

Study	Study design	N of networks	N sample for networks	Model	Comparison	Statistical tests
Cervin et al. (2020)	RCT	9	215/215/215/209/209/209/488/488/488	MGM	timepoints during treatment	none
Kaiser, Boschloo, et al., (2021)	RCT	10	794/ 572/ 530/ 754/ 511/ 508/ 498/ 475/ 482/ 692	MGM	timepoints during and after treatment	none
Kraft et al. (2019)*	RCT	1	322	GGM	n.a.	none
Lancee et al. (2022)	RCT	10	125/113/ 99/97/94	MGM	timepoints before, during and after treatment	none
Monteleone et al. (2021)	RCT	4	187/ 146/ 121/ 115	MGM	timepoints before and after treatment	none
Mullarkey et al. (2020)	RCT	1	295	MGM	n.a.	none

Note. RCT = randomized controlled trial, MGM = Mixed Graphical Model, n.r. = not reported, n.a. = not applicable. *These studies report two different kinds of symptom networks. Information here just relates to the cross-sectional networks with a treatment node.

Table 4.3. Network analysis characteristics of studies estimating longitudinal networks.

Study	Study design	N of networks	N sample for networks	N datapoints for network	Model	Comparison	Statistical tests
Booij et al. (2021)	observational	134 (133 individual, 1 group-level)	133 x 1/ 133	mean = 6.2/ 819	dynamic time warp distance matrices	groups defined by treatment response	t-test of difference in average dynamic time warp distance
Briganti et al. (2021)*	observational	1	100	300	GVAR	n.a.	none
Bringmann et al. (2015)	RCT	3	182/ 99/ 83	2595	mlVAR model including a time trend	groups receiving different treatment	community structure analysis, BIC differences

Table 4.3. Continued

Study	Study design	N of networks	N sample for networks	N datapoints for network	Model	Comparison	Statistical tests
Groen et al. (2019)	observational (original study was an RCT)	2	30/ 30	348/ 330	mlVAR model	groups defined by treatment response	permutation tests for differences of edges and connectivity
Hebbrecht et al. (2020)	observational	256 (255 individual, 1 group-level)	255 x 1 / 255	mean = 5.8/ 1480	dynamic time warp distance matrices	groups defined by treatment response	Wilcoxon signed-rank test comparing the average density
Hoffart (2018)	observational	1	35	645	n.r. (assume mlVAR model)	n.a.	none
Hoffart et al. (2019)*	observational	2	65/ 65	727	mlVAR model	n.a.	none
Hoffart & Johnson (2020)*	RCT	2	60/ 60	1176	mlVAR model	n.a.	none
Holzhauer et al. (2020)	RCT	6	55/ 55/ 55/ 44/ 44/ 44	220/ 220/ 220/ 176/ 176/ 176	mlVAR model	groups receiving different treatment	none
Johnson & Hoffart (2018)	RCT	6	38/ 38/ 38/ 36/ 36/ 36	n.r.	mlVAR model	groups receiving different treatment	none
Komulainen et al. (2021)*	RCT	2	1013 /2417	n.r.	n.r. (assume ISING model)	groups receiving different treatment	none
Kreiter et al. (2021)	RCT	6	9/ 9/ 9/ 5/ 5/ 5	480/ 358/ 397/ 254/ 268/ 247	mlVAR model including a time trend	timepoints and groups receiving different treatment	permutation test for edge differences
Lutz et al. (2018)	observational	3	58/ 35/ 23	3248/ 1960/ 1288	mlVAR model	groups defined by treatment response	prediction of dropout with network parameters

Table 4.3. Continued

Study	Study design	N of networks	N sample for networks	N datapoints for network	Model	Comparison	Statistical tests
Lydon-Staley et al. (2021)	RCT	1212 (1210 individual, 2 group-level)	1x1210/ 1210 / 1210	mean = 27.91/ 33771	mlVAR model/ unified structural equation model	groups receiving different treatment	estimating system recovery time with simulations and comparing these with t-tests between treatment groups
Mariotti et al. (2021)	observational	1	1	n.r.	VAR model	n.a.	none
Snippe et al. (2017)	RCT	8	23/ 23/ 27/ 27/ 57/ 57/ 62/ 62	n.r.	mlVAR model including a time trend	timepoints and groups receiving different treatment	permutation test to assess significance of lagged relations (edges), change in lagged relations and difference in change in lagged associations between treatment groups
Strauss et al. (2020)	RCT	2	83/ 161	n.r.	cosine similarity measure networks	groups receiving different treatment	repeated measure ANOVA and independent samples t-test for centrality indices

Table 4.3. Continued

Study	Study design	N of networks	N sample for networks	N datapoints for network	Model	Comparison	Statistical tests
Vittengl et al. (2022)	observational	4	290/ 185/ 84/ 62	3422/ 2183/ 1024/ 756	mlVAR model	n.a.	predicting treatment outcome variables including response and no response from the symptom linkage density

Note. RCT = randomized controlled trial, GVAR = Global Vector Autoregressive, mlVAR = multilevel Vector Autoregressive, n.r. = not reported, n.a. = not applicable. *These studies report two different kinds of symptom networks. Information here just relates to the longitudinal networks.

4.3.4 Reporting and open science practices

Several important variables were missing for a considerate number of studies, see online supplemental material Table S5. 41 out of the 56 studies (73.2%) did not report how nodes were selected, 28 studies (50%) did not report handling of missing data, 10 studies (17.9%) did not specify which software package was used and seven out of 56 studies (12.5%) did not report the model that was estimated. Software code to reproduce the analysis was available for 11 studies (19.6%). Only one study shared their data and three studies (5.4%) published correlation matrices to enable reproduction of the networks. None of the network analyses were preregistered; the original trial that the data was taken from was registered in nine studies (16%).

4.4 Discussion

We conducted this systematic review to gain an overview on how network analysis has been applied to evaluate treatments for mental disorders and evaluate the strength and weaknesses of the different analytic options. The application of network analysis in intervention research was more frequent than expected, with the majority of studies being published since 2020. It became apparent that how network analysis was used for the evaluation of treatments differed to a large extend. Studies varied on which kind of networks were chosen. The majority analyzed cross-sectional networks that indicated pairwise associations between symptoms. The remaining studies applied cross-sectional networks including a treatment node or longitudinal networks. Consequently, many different underlying models were adopted, most commonly GGMs, MGMs, mlVAR models and ISING models, and the definition of edges varied widely. There was also variation between studies in regards to which parameters were calculated to describe the networks and if networks of different treatment groups, response groups or at different time points were compared. All these different analytic decisions had an impact on the information that could be gained from the network anal-

ysis. Finally, reporting differed between studies and open science practices were rarely applied. In the following, we aim to provide information that can help researchers to make informed analytic decisions based on their specific research question when evaluating treatments for mental disorders with network analyses by discussing benefits and challenges of different analytic options.

4.4.1 General topics in network analysis

Compared to other multivariate techniques, network analysis is still relatively new. Although much progress has been made in the past decade, there are several issues concerning all kinds of network analyses (also outside of intervention research) that need to be addressed in future research. First, we found that only a small number of studies described their selection procedure for the variables that were included in the networks. However, as most networks show conditional dependencies and the resulting network structure completely depends on which variables are included, node selection needs to be carefully considered (Bringmann et al., 2022). Not including influential (possibly confounding) variables could lead to spurious symptom associations and misleading interpretation. This emphasizes the necessity to find clear criteria for the inclusion of variables for individual studies and the need to develop general standards for symptom inclusion for the field. These could, for example, be guided by symptoms included in the current classification systems like DSM 5 and ICD 11. Similar to Robinaugh and colleagues (2020), we found that more than 80% of the studies in this review included solely symptoms (or a treatment node) in the networks. To better understand mechanisms of change in treatments, it might be beneficial to also include hypothesized treatment processes as variables in the network (Hofmann et al., 2020). Johnson and Hoffart (2018) investigated, for example, the associations between symptoms and (meta) cognitions for clients receiving metacognitive therapy or CBT. Including potentially interesting variables in the network next to symptoms takes advantages of the fact that networks can display associations between multiple different variables and might improve the knowledge that can be gained about treatment effects from network analysis.

Further, this review showed that symptoms were most commonly measured with a single item, which contrasts to established measurement practices (Allen et al., 2022). A reliable measurement of included variables is a pre-requisite to a reliable network analysis (Bringmann et al., 2022), therefore, future studies need to closer assess the validity of this measurement practice (Allen et al., 2022). Third, optimal sample size, power and robustness of psychopathology networks is a debated topic and (initial) recommendations depend on the model type, number of nodes and expected effect sizes (Bringmann et al., 2022; Lafit et al., 2021). For cross-sectional networks, the stability of edges and network parameter can be assessed with bootstrapping methods to gauge robustness of the network (Epskamp, Borsboom, et al., 2018). For longitudinal networks, robustness has been only evaluated with cross-validation . Accordingly, no longitudinal study and only half of the cross-sectional studies included in this review formally assessed edge stability. Given that the median sample size of all studies was rather low (n = 151) and the stability of the estimations were not regularly investigated, the replicability of current symptom networks in intervention research is largely unknown. For network analysis to provide robust information in regards to intervention effects, an assessment of its replicability is indispensable.

4.4.2 Different types of psychopathology networks in intervention research

Different types of network analyses pose different challenges and potential for the evaluation of treatments. Cross-sectional networks can show how variables relate between persons and their robustness can be estimated with bootstrapping methods (Epskamp, Borsboom, et al., 2018). Ideally with data taken from an RCT, cross-sectional networks can be estimated for each treatment group and different time points before, during and after the treatment. Thus, change in symptom associations through treatment and differences between client groups can be investigated. Here, GGM, MGM or ISING models should be chosen depending on the included variables and networks should be statistically compared, e.g. with the network comparison test (van Borkulo et al., 2022). When including a treatment node in the network, indicating randomized treatment allocation, interpretation of the treatment effects is quite straight forward. Associations between the treatment node and a symptom reflect the causal effect of the treatment on the specific symptom (Cervin et al., 2020). Here, MGMs or ISING models should be used and accuracy analysis using nonparametric bootstrapping should be conducted (Haslbeck & Waldorp, 2020). In many studies, cross-sectional networks with a treatment node were estimated at different time points before, during and after the treatment, aiming to show the evolvement of symptom-specific treatment effects over time. Unfortunately, none of the studies statistically tested difference between these networks, thus, differences can be interpreted only tentatively. Another promising approach shown in the review is the estimation of cross-sectional network including a treatment node and nodes that indicate change in symptom severity (e.g. Boschloo, Cuijpers, et al., 2019). With such models, the direct and indirect effect of treatments on change in symptom severity can be assessed. The biggest drawbacks of these kind of analyses is their cross-sectional nature. As cross-sectional analyses do not disentangle within-person from between-person effects (Schuurman, 2023), it seems questionable how cross-sectional network analyses relate to within-person treatment effects (Bos et al., 2018). It has been argued that for relating treatment effects to the individual person, within-person and not between-person treatment effects need to be assessed (Epskamp, Waldorp, et al., 2018; Molenaar, 2004; Schuurman, 2023). Therefore, it needs to be carefully considered to what extend results from between-person networks can generalize to treatment effects for the individual client.

In contrast, using longitudinal networks to evaluate treatment effects has the benefit that within-person processes can be investigated (Epskamp, Waldorp, et al., 2018). Here, one major obstacle is that the VAR model, including GVAR and mlVAR, assumes that symptom relations stay the same over time, i.e. are stationary (Bringmann et al., 2013). When analyzing data which was collected during treatment, this assumption is likely to be violated. As seen in this review and a previous review of longitudinal networks (Blanchard & Heeren, 2022), most studies handle the violation of stationarity by detrending the data. However, as symptoms are hypothesized to change through treatments, their relations are also likely to change. Similarly, network theory proposes that treatment changes the symptom associations (Borsboom, 2017). Thus, by ignoring the change in symptom associations over time, we run the risk of missing a large part of the treatment effect. Additional obstacles in using longitudinal networks for treatment evaluation are the lack of possibilities in assessing edge stability and the lack of standard procedure on how to formally compare longitudinal networks. New methods to estimate time-varying VAR-models (Haslbeck, Bringmann, et al., 2021) or dynamic time warp analysis (Hebbrecht et al., 2020; Booij et al., 2021) do

not assume stationarity in symptom relations, and have thus a larger potential for using longitudinal networks in intervention research. Similarly, estimating longitudinal networks for a time period before and a time period after treatment as done by Kreiter and colleagues (2021) and Snippe and colleagues (2017) seems promising, as here the assumption of stationary symptom associations is more likely to be true.

4.4.3 Reporting and open science practices

As network analysis (in intervention research) is rather new, no reporting standards have been established yet. When it is unclear what to report, being transparent of your methods, i.e. the use of open science practices, also becomes more difficult. In this review, information on the investigated interventions and the analysis was missing for a considerable number of studies. Chapter 5 of this thesis proposes reporting standards, which hopefully will lead to more consistent reporting in network analyses. Further, only a few studies preregistered their analysis and shared their code and data, making it more difficult to reproduce the analyses. While many studies were exploratory, it still seems recommendable to preregister a priori analytic decisions to limit the impact of (post-hoc) analytic decisions. We realize that within clinical contexts, data is often more sensitive and difficult to share while protecting clients' anonymity. Still, especially when code is openly available, the analyses are more easily reproducible and researchers can learn from each other. Precise reporting and good reproducibility are especially important under the light of the various analytic options for network analysis in intervention research.

4.4.4 Limitations of the current review

Not all studies that were included in the review originally aimed to evaluate a treatment with the network approach, possibly inflating the variability in methods found. These were still included because they analyzed data that were collected within the realm of treatment and conducted some analyses that allowed some inference in regards to the treatment effect. Additionally, this review included all kinds of interventions that were directed at mental health problems. Different network methodologies might be more suitable for different kind of interventions and this could not be assessed in this study. We took this broad approach as we wanted to gain an overview of all kinds of network analyses that were facilitated in previous intervention research.

4.4.5 Overall evaluation of the current methodological state and outlook

This review showed that the previously used methodological and statistical practices have several limitations. In previous studies, the importance of variable selection, variable measurement, and the assessment of the robustness of the analysis has not yet been comprehensively addressed. Thus, it is not yet clear if the observed results can be expected to replicate. Since these aspects are indispensable to gain valuable knowledge from network analysis, the field needs to move forward addressing these issues. Even more importantly, it became apparent that currently used statistical network models can only indirectly and to a limited extended provide the information that was hoped to be gained through the network approach. This is, it was suggested that the network analyses can inform on symptom-specific treatment effects (Blanken et al., 2019), provide information on how (causal) symptom associations change through treatment and, thereby providing insights into working

mechanisms of investigated treatments (Hofmann et al., 2020). Symptom-specific treatment effects could be shown by network analyses including a treatment node. However, so far, these analyses were only applied cross-sectional, therefore, the relation between these between-person results and within-person treatment effects is unknown. Similarly, when investigating change in symptom associations through cross-sectional networks before and after treatment or of different responder groups, only change in between-person associations and not within-person processes are displayed. Finally, as cross-sectional networks mostly display (partial) correlations, no causal interactions between symptoms can be inferred and therefore, inference about possible working mechanisms seems difficult. When investigating longitudinal network models, inference about within-person effects and possible causal mechanisms is more appropriate. However, as most longitudinal models assume stationarity, i.e. no change in symptom associations, no change due to treatment could be directly investigated. Finally, it should be noted that network theory suggests that individuals differ in their symptom networks (Borsboom, 2017; Borsboom & Cramer, 2013). As the majority of studies investigated group-level symptom networks, individual differences for treatment effects on symptom networks are largely unknown. In sum, current methodological and statistical practices provide limited information on the causal, dynamic (i.e., changing) and possibly person-specific interactions among symptoms and how these are impacted by treatment.

Importantly, we think that suboptimal methodological practices are to be expected in such a new, emerging field. There are ongoing methodological developments, e.g. Bayesian approaches (Huth et al., 2023), time-varying approaches (Haslbeck, Bringmann, et al., 2021) and new estimation approaches like Group Iterative Multiple Model Estimation (Sanford et al., 2022), which are likely to further advance the field. In our opinion, further methodological and statistical developments would be highly valuable as the network approach to psychopathology offers a new perspective on the treatment of mental disorders and has the potential provide new insights about treatments and their effects. Network analyses allow the exploration of associations among various specific symptoms, the display of a complex picture of a multitude of variables and an assessment of symptom specific treatment effects. These kind of analysis are likely to provide more detailed information than analyses focusing on the composite score of various different symptoms or the presence of a diagnosis. With methodological advances, it can be a tool for the investigation of treatment processes and personalization of treatments, as will be discussed in chapters 9 and 10, which are also both increasingly important topics in intervention research. To assess changes in symptom-associations, a time-variant longitudinal network model, taking into account individual differences, could be estimated for each treatment group from data of a randomized controlled trial. Then, within-person symptom associations and their change in response to treatment could be investigated and compared between treatment groups. Similarly, longitudinal symptom networks including a treatment node based data from individuals receiving different treatments could show within-person symptom-specific treatment effects. With methodological and statistical advances, network analysis provides a promising tool for the exploration of symptom interaction and could be valuable besides other methods such as structural equation modeling and computational modeling which require stronger theoretical knowledge.

We are aware that much progress is made currently, and several newer approaches already address some of the mentioned weaknesses. We hope that the current chapter can give directions

for applied researchers about which methods are available and which issues need to be considered. Future studies need to address the scope of applicability of different analytic options. With a better understanding of which network analyses are suitable for which kind of data and which questions, network analyses in intervention research can hopefully help us to learn more about symptom-specific treatment effects and with this increase our understanding of treatments for mental health problems.

CHAPTER 5

REPORTING STANDARDS FOR PSYCHOLOGICAL NETWORK ANALYSIS IN CROSS-SECTIONAL DATA

CHAPTER 5

Abstract

Statistical network models describing multivariate dependency structures in psychological data have gained increasing popularity. Such comparably novel statistical techniques require specific guidelines to make them accessible to the research community. So far, researchers have provided tutorials guiding the *estimation* of networks and their accuracy. However, there is currently little guidance in determining what parts of the analyses and results should be *documented* in a scientific report. A lack of such reporting standards may foster researcher degrees of freedom and could provide fertile ground for questionable reporting practices. Here, we introduce reporting standards for network analyses in cross-sectional data, along with a tutorial and two examples. The presented guidelines are aimed at researchers as well as the broader scientific community, such as reviewers and journal editors evaluating scientific work. We conclude by discussing how the network literature specifically can benefit from such guidelines for reporting and transparency.

This chapter has been adapted from: Burger, J., Isvoranu, A.-M., Lunansky, G., Haslbeck, J. M. B., Epskamp, S., Hoekstra, R. H. A., Fried, E. I., Borsboom, D., & Blanken, T. F. (2022). Reporting standards for psychological network analyses in cross-sectional data. *Psychological Methods*. Advance online publication.

5.1 Introduction

Over the past decade, there has been a rapid increase in empirical contributions applying network analytic methods across many psychological disciplines. The increasing interest in networks (Barabási, 2012; Watts & Strogatz, 1998) led to empirical applications in various fields of psychology (Robinaugh et al., 2020) and resulted in a large number of special issues in journals such as *Psychometrika*, *The European Journal of Personality, The European Journal of Psychological Assessment*, *BMC Medicine,* and *The Journal of Traumatic Stress.* However, there is a lack of clear guidelines on how to report psychological network analyses. The present chapter introduces such guidelines, aiming to enable researchers to identify all elements of their analyses that should be included in a scientific report. We argue that reporting guidelines can facilitate the evaluation of network contributions by the broader scientific community, including reviewers, editors, journalists, and science writers.

5.1.1 Questionable reporting practices and the benefit of reporting standards

While there are several tutorials on *estimating* networks from psychological data (Costantini et al., 2015; Epskamp, Borsboom, et al., 2018; Epskamp & Fried, 2018; Haslbeck, Bringmann, et al., 2020; Jones et al., 2018; Williams & Mulder, 2020), as of yet, there is no guidance for how researchers should *report* the results of network analyses in a scientific paper. There are general reporting standards for statistical analyses, such as the *Journal Article Reporting Standards for Quantitative Research in Psychology* published by the APA Publications and Communications Board Task Force (Appelbaum et al., 2018). However, specific types of multivariate analyses contain explicit elements that go beyond the scope of generic reporting standards (Hoyle & Isherwood, 2013). For this reason, more tailored reporting standards do exist for other types of multivariate analyses, such as structural equation modeling (Schreiber et al., 2006). At present, however, there are no explicated standards on how to report the results of network analyses.

A lack of clear reporting standards, in turn, may hinder rigorous scientific communication: (Wigboldus & Dotsch, 2016) highlight that a large part of the degrees of freedom in empirical research resulting in questionable research practices are in fact gray areas that pertain to questionable *reporting* practices. To this end, objective reporting standards for network analysis are an important contribution towards making empirical network studies more rigorous. Since such norms are not yet established in the network literature, the goal of the present chapter is to explicate what we refer to as "minimal shared norms" in reporting psychological network analyses. By making these shared norms explicit, they can be extended and debated, and they will increase the replicability and reproducibility of network analysis, both of which will move the field of network psychometrics forward.

5.1.2 A brief introduction to psychological network analysis

While a detailed introduction to psychological network analysis is beyond the scope of this chapter, in this section we briefly introduce this methodology as to keep the chapter self-contained. A more extensive primer on network analyses in psychological science has recently been published (Borsboom, Deserno, et al., 2021), and a textbook dedicated to the emerging field of network

psychometrics is currently in press (Isvoranu et al., 2022). In addition, we include a glossary that provides an overview over the most important network-specific concepts discussed in this chapter.

A *network* is any system which can be represented with *nodes* (circles), which are connected by *edges* (lines) denoting a strength of connection between the nodes. In psychological networks, nodes represent observed variables, and edges are used to represent the strength of associations between two variables, typically after controlling for all other variables in the dataset. This type of model is termed a *Markov Random Field*, which includes commonly used network models depending on the data used: Gaussian graphical models (GGM) – also termed partial correlation networks – for continuous data (Epskamp, Waldorp, et al., 2018; Lauritzen, 1996), Ising models for binary data (Epskamp, Maris, et al., 2018; Ising, 1925; Marsman et al., 2018; Van Borkulo et al., 2014), and mixed graphical models (MGM) for mixed data (Haslbeck & Waldorp, 2020). Psychological networks can be estimated with (penalized) maximum likelihood estimation (Epskamp & Fried, 2018), Bayesian estimation (Williams & Mulder, 2020), or pseudo-likelihood estimation (i.e., nodewise regression) where each variable is regressed on all other variables, after which results are combined to form a network (Epskamp, Maris, et al., 2018; Haslbeck & Waldorp, 2020; Van Borkulo et al., 2014).

As is the case for statistical models in general, a crucial aspect of psychological network analysis is that estimated models are subject to sampling variation. As a result, edges may falsely be included while not being present in the true model, and differences in edge weights may be strong merely due to chance. To address such chance fluctuations, psychological network analyses should always include both model selection methods and checks for stability and accuracy. *Model selection* algorithms are diverse but generally fall under one of three categories (Blanken, Isvoranu, et al., 2022): (1) *Pruning/thresholding* methods, which merely remove or hide edges that do not meet some criterion as defined by a classical statistical significance level or a lower Bayes factor; (2) *Model search* strategies, which use extensive model search methods to iteratively arrive at an optimal network structure, typically informed by an information criterion; and (3) *Regularization* methods, which use penalized maximum likelihood estimation to shrink parameters to zero, potentially removing them from the network. Each of these strategies has its pros and cons (Isvoranu & Epskamp, 2021). For example, regularization techniques (Meinshausen et al., 2006; Ravikumar et al., 2010; Tibshirani, 1996) may work well in retrieving an interpretable structure at low sample sizes, but may also feature a lower specificity rate than desired (Williams et al., 2019). Furthermore, in such circumstances one must be careful to interpret the sparsity of the network, as this is, at least in part, a consequence of the estimation method used (Epskamp, Kruis, et al., 2017). Checks for stability and accuracy usually involve the use of data-driven resampling methods such as bootstrapping (Epskamp, Borsboom, et al., 2018) or Bayesian sampling methods (Williams & Mulder, 2020) to assess and visualize uncertainty around parameter estimates.

5.1.3 Scope of models and software

In this chapter, whenever we refer to "network models," we intend to designate statistical models that are designed to capture pairwise statistical interactions between variables and that are estimated on cross-sectional data. Our focus lies on cross-sectional networks, because network analyses for this type of data account for the largest part of empirical network contributions over the past ten

years (83% of the identified empirical papers between 2008 and 2018 as reported by Robinaugh et al., 2020). Of note, there are many other types of psychological network analyses than the ones we discuss here, including models estimated in panel data and time series data (Epskamp, 2020b; Gates & Molenaar, 2012; Haslbeck, Bringmann, et al., 2020) or moderated network models (Haslbeck, Borsboom, et al., 2019). These are beyond the scope of the present chapter as they require different reporting standards due to differences in data structure, estimation methods, and model assumptions.

Within the domain of cross-sectional network analysis, there is a wealth of software options. Depending on the choice of software, different reporting elements, such as specific test statistics, might be required to ensure interpretability of the results. Here, we focus on software implemented in the open-source environment R (R Core Team, 2015), specifically on packages that have been most frequently used in the past decade in empirical, psychological network contributions (Robinaugh et al., 2020). An overview of the software packages that we cover in this chapter can be found in Table 5.1. While we focus on a specific set of R-packages, most of the discussed reporting standards represent core elements of cross-sectional network analysis in psychological data. We therefore expect that the introduced reporting standards will also be applicable to other software, albeit not in regard to the specific test statistics included in this chapter. For instance, reporting *parameter uncertainty* is not a unique standard of the packages discussed in this chapter but should be included for any contribution that estimates partial correlation networks. Consequently, the listed packages should be seen as examples of how the core reporting standards introduced here can be applied to software that is frequently used in the literature, rather than restricting the domain of reporting standards to this type of software alone.

Lastly, the presented guidelines may also be applicable to some aspects of reporting simulation studies on network analyses. For example, simulation studies should include information on how networks were derived from the simulated data. However, simulation studies may require specific additional reporting elements, such as information on data-generating mechanisms and performance measures (e.g., bias or mean squared error). We therefore recommend considering additional guidelines for simulation studies, such as the guidelines provided by Morris, White, & Crowther (2019).

5.1.4 Organization of the proposed reporting standards

This chapter adopts the typical structure of a psychological report according to APA standards (American Psychological Association, 2020) and can therefore be used as a reference for authors who prepare their work for submission to an APA journal. Of note, some of the recommendations discussed below, such as *reporting on the variable selection procedure*, are not unique reporting elements for network analyses. We included those elements for two reasons: First, to ensure that these guidelines are standalone readable, and second, because some more general elements deserve specific attention when using network analyses (e.g., variable selection is related to the problem of *topological overlap*, see Box 5.1).

We provide a reporting routine for both the "Methods" and the "Results" sections of an empirical APA report (sections 5.2 and 5.3, respectively), using the following structure:

5.1.4.1 General analysis routine

These sections contain reporting standards that are applicable to all analyses as defined above, independent of specific research questions. These routines include the reporting of general features of the data, the statistical approach, details about the sample and variables, as well as accuracy and stability checks. We recommend to always report these elements.

5.1.4.2 Analysis-specific routine

These sections contain reporting standards that apply only to specific research questions and analyses within the network analytic framework, such as reporting on group comparisons, centrality analyses, edge differences and visualization. Not all of these will be of interest for every empirical network contribution and are therefore only applicable if they align with their specific research question.

5.1.4.3 What to watch out for

The main focus of this chapter lies on providing reporting standards and not interpretation guidelines. However, some reporting standards are closely related to interpretation. Therefore, in the "What to watch out for" boxes, we discuss some of considerations that are important when applying network analyses to psychological data.

5.1.4.4 Illustrative examples

To illustrate these norms and reporting standards, we include two examples of network analyses on openly available data with two distinct research goals. Further, we include an overview of most network estimation packages and functions referred to in this chapter, along with information on important arguments, current estimation defaults, applicable input data, and parameter interpretation (Table 5.1).

5.2 Reporting standards for the 'Methods' section

5.2.1 General analysis routine

5.2.1.1 Sample collection

We recommend to specifically consider and report how and from which population the participants were recruited and whether a sub-population was included in the analyses (e.g., depressed clients; see Box 5.1, *Biases due to subsample selection*). Subsample selection can occur because of recruitment strategies (e.g., collecting data in clinical practice) or by selection after data collection (e.g., only include participants that scored higher than a certain cut-off). Make sure to report on subsamples in either case. Report the number of participants for whom data was collected and the number of participants that were included in the network analyses.

5.2.1.2 Variable selection procedure

As with any other study, it is important to precisely report what instruments were used to collect the data, as well as the versions of these instruments, if applicable (Flake & Fried, 2019). We rec-

ommend specifically considering the instrument, as some questionnaires might include multiple items that have the same relations to other nodes (i.e., topological overlap), which can lead to problematic inferences in networks (see Box 5.1, *instrument design*). With regard to network analyses, we recommend to additionally report on the number of variables on which data were collected. When the data are preprocessed before being included in the analyses (e.g., variable selection or transformation), report on these processing steps and indicate the number of variables included in the network analysis. Pre-processing choices concern, but are not limited to, collapsing variables (e.g., aggregating variables such as 'loss of appetite' or 'increase of appetite'), collapsing categories (e.g., binarization of Likert-scale data), data transformations (e.g., in case of violating assumptions; see Box 5.1, *variable distribution*), and imputation or removal of missing data (e.g., listwise deletion of cases). An exhaustive list of choices that warrant justification is listed elsewhere (Flake & Fried, 2019). For the variables that are included in the network, we recommend comparing the distribution of the variables with the assumptions of the estimation method and checking any violations (e.g., skewness of the data; see Box 5.2, *variable distribution*). If variables are removed/included following network stability analyses (see *Accuracy and stability of edge-estimates*), this should be reported as well.

5.2.1.3 Deterministic relations between variables and skip-structures

The manuscript should specifically report if the scale used to construct the network contains a so-called *skip-structure*, i.e., some questions in the questionnaire are skipped based on responses to previous questions. This can occur when participants are instructed to only answer one question or the other (e.g., report either on weight loss or weight gain) or when certain follow-up items are only administered to a subset of participants (e.g., only assessing nuanced depressive symptomatology if one of the core depression symptoms is present). This creates a missingness problem for the data that should be addressed, and the report should indicate precisely how this problem has been handled. This is important because some methods, such as imputing zeroes for skipped items, will induce dependency relationships in the data that bias the network structure and can lead to faulty inference (see Borsboom et al., 2017). The latter problem will hold for any deterministic relationship included in the network (e.g., including a sum-score variable together with the components that make up the sum-score) and should be avoided. To our knowledge, no validated methods for handling such structures exist to date and therefore it is recommended not to analyze skip-structure questionnaires using network analysis. In the case of large diagnostic questionnaires (e.g., SCID, CIDI), one alternative could be to focus on the diagnostic category questions that all subjects have answered rather than on follow-up skip items.

5.2.1.4 Estimation method

We recommend to specifically mention in the manuscript how the data was modeled (i.e., continuous, ordinal, binary, etc.). The measurement level is linked to the estimation method used when performing a network analysis, which should always be reported as well (e.g., *EBICglasso*, *IsingFit*, *MGM*, etc.; see Table 5.1 for a description of commonly used estimation techniques). In addition to the estimation method, mention any additional specifications. For example, when the networks are thresholded, report the chosen thresholds; when regularization is used, report the parameter

specifying the search for appropriate regularization. Of note, even if researchers stick to default arguments (i.e., the standard settings that are used in the estimation procedure), we recommend reporting them, since defaults in software packages can change which in turn would make reproducing analyses difficult.[28] Finally, we advise considering the assumptions of each estimation method (see Box 5.1, *variable distribution*), as well as how each estimation method handles missing data (see Box 5.2, *missing data*).

5.2.1.5 Accuracy and stability of edge-estimates

As with any procedure that involves parameter estimation, it is important to assess how accurate our estimates are (Fried, Epskamp, et al., 2022). In the context of the currently most common estimation techniques in network analysis, accuracy can be assessed via a bootstrap procedure implemented in the R-package *bootnet* (Epskamp, Borsboom, et al., 2018) using the function *bootnet* and specifying the argument *type* as "nonparametric"). In this procedure, the model is estimated repeatedly under resampled or simulated data and statistics of interest (e.g., edge weights) are computed (Efron, 1979). As such, bootstrapping allows to approximate the sampling distribution of the parameters in the population. The sampling distribution can then be inspected visually (for details see e.g., Epskamp, Waldorp, et al., 2018). Specifically, in the methodology section of the manuscript, we advise reporting the number of bootstrap samples, as well as the type of bootstrap method employed (in the above case "nonparametric"). For methods that make use of Bayesian inference, such as *BGGM* (Williams & Mulder, 2020), there are equivalent measures to assess accuracy and stability, such as credibility intervals for estimates and convergence diagnostics.

5.2.1.6 Statistical packages

Finally, we recommend reporting the statistical software and packages that are used, including their versions. Full reproducibility is guaranteed only if this information is shared along with code and data, because statistical packages can change estimation defaults when they are updated (Epskamp, 2019). With this information, the reader can mimic the analyses under identical estimation settings and reproduce all results, for example using the *checkpoint* package in R (Ooi et al., 2020). We further recommend including any seed-settings in the code that have been used in conducting analyses (e.g., if estimation techniques based on cross-validation or the *Network Comparison Test* were used; (Haslbeck & Waldorp, 2020; van Borkulo, Boschloo, et al., 2017). Note, however, that setting a seed does not fix results if parallel computing is used, as is often the case when drawing many bootstrap samples.

5.2.2 Analysis-specific routine

5.2.2.1 Group comparisons

If groups are compared, we recommend reporting which methods have been employed to compare groups (usually correlating weighted adjacency matrices; comparing networks using the *Network Comparison Test* (van Borkulo, Boschloo, et al., 2017); comparisons based on the posterior

28 Within the *R* statistical software (R Core Team, 2015), the defaults of each package can be checked using the "?" + name of the function within a statistical package (e.g., ?*estimateNetwork*).

predictive distribution or model selection in Bayesian GGMs (Williams et al., 2020); estimating moderated network models in *mgm* (Haslbeck, 2020; Haslbeck & Waldorp, 2020); or through using multi-group network modeling (Epskamp, Isvoranu, et al., 2021)). If groups are compared using multiple methods, we recommend reporting all comparisons that were made and in addition reflect on the consistency of the results. Of note, these methods are dependent on the sample size and identifying no differences may sometimes reflect power issues.

5.2.2.2 Centrality indices

One particular application of network analysis is to identify nodes that could be particularly influential, for example because they are well connected to other nodes. In graph theory and network analysis, the quantification of this relative influence based on the network flow is referred to as *centrality analysis*. Centrality metrics can be computed that quantify the role of each node in a network (Costantini et al., 2015; Jones et al., 2019; Opsahl et al., 2010), for example via the *qgraph* package in R (Epskamp et al., 2012; using the functions *centrality*, *centralityPlot*, or *centralityTable*), or via the *networktools* package in R (Jones, 2017; using the function *bridge*). If such inferences are of interest, we recommend carefully selecting centrality metrics that relate to the specific research question. For example, if the research question involves identifying the most strongly connected nodes (as is the case in for example Elliott et al., 2020), "strength centrality" may be most suited, whereas if the research question involves identifying nodes that bridge different clusters (as is the case in for example Levinson et al., 2018) "bridge centrality" measures may be most informative. There may also be research scenarios in which a combination of these metrics is of interest (as is the case in for example Isvoranu et al., 2021). We recommend reporting all centrality metrics that were computed, alongside the accuracy of their estimates (e.g., case-drop bootstrap in the *bootnet* package, using the function *bootnet* and argument *type* set to "case", for more information see Epskamp et al., 2018; see also Box 5.1, *centrality*). Suppose the differences between node centralities are not robust. In that case, it cannot reliably be determined which node is "most central" (note that this does not imply the network was estimated with low accuracy; it is also possible that there simply are no differences in centrality between nodes; see Box 5.2). In this case, we recommend only reporting that the centrality metric was computed, but that the centrality differences between nodes will not be further interpreted because these differences are not stable.

5.2.2.3 Differences between edges within one network

If edges within a network are compared with one another, we recommend reporting the method of comparison (e.g., the bootstrapped difference-test in the *R* package *bootnet*, using the *differenceTest* function; Epskamp et al., 2018). Further, if hypotheses are tested in a Bayesian context (Williams & Mulder, 2020), these should be stated explicitly (e.g., A – B > C – D).

5.2.2.4 Clustering

Clustering refers to the tendency of a network to exhibit groups of nodes that arise from their specific interconnections. If clustering of nodes is of interest, we recommend reporting which clustering method was employed when running the analyses (e.g., *Exploratory Graph Analysis*; Golino & Epskamp, 2017), why this particular method has been chosen (Hennig, 2015), as well as if and how the stability of the identified clusters was checked.

Dataset	**Instrument design**. It is important to consider how the instrument used to gather the data was constructed. For instance, variables included in a network may come from a single questionnaire that was constructed to measure a latent variable, and is therefore intended to measure a single underlying construct. If a set of items does in fact depend on the same latent variable, but the items are interpreted as measuring distinct factors, possible distortions in e.g., centrality estimates should be taken into account (Fried & Cramer, 2017).
Variable distribution	**Assumptions of estimation methods.** For each estimation method, model assumptions should be considered and violations of these assumptions should be addressed. Main assumptions include (1) independent cases; (2) the presence of (log) linear relationships and pairwise interactions only; (3) missing data are Missing (Completely) at Random (Rubin, 1976); (4) relevant distributional assumptions of the variables included in the network. **Variance.** Certain restrictions to variance, such as floor/ceiling effects or restrictions in range, can affect statistical relationships. This should be considered when interpreting edges and the importance of variables (e.g., suicidal ideation is typically restricted in variance but clinically relevant; see also *centrality* below and (Fried et al., 2018). Note that these artifacts not only pertain to networks estimated from continuous data but also to those estimated from binary data; for example, if symptoms are coded as present versus absent and most participants in the sample are healthy individuals without symptoms, floor effects may occur.
Subsample selection	**Biases due to subsample selection (e.g., Berkson's bias).** Sample selection is important because it can lead to unexpected patterns in the data. For example, if a sub-population (e.g., depressed clients) is recruited based on a cut-off on the total score of symptoms included in the network structure, one may find that, in that sub-population, many edges between symptoms are negative. The reason for this result is that the total score is composed of the individual item scores. As a simple example that illustrates the effect, suppose one throws coins A and B repeatedly and only selects cases in which only one of them falls heads (i.e., total score = 1). Within this set of throws (i.e., conditioning on the total score), the correlation between the outcomes of the tosses for the two coins will be negative because if coin A falls heads then, given a total score of 1, coin B must have fallen tails. This effect has been referred to as Berkson's bias (de Ron et al., 2020). However, it has also been noted that Berkson's bias is but one of various effects of conditioning, and that these need not constitute bias in the statistical sense (Haslbeck, Ryan, et al., 2020). Nevertheless, it is important for researchers to realize that creating subsamples based on functions of the variables in the network will often have strong effects on the network structures found in these subsamples.

Variable inclusion	**Variable selection**. The structure of network estimation results depends on which variables were included in the analysis. This is due to the fact that conditional dependencies are used in network estimation: conditioning on different sets of variables can therefore lead to different network structures. This implies that the network structure may change if variables are included in or excluded from the model. **Item-scores versus sum-scores.** Depending on the research question, item-scores may sometimes be preferred, whereas sum-scores may be the best option at other times. For example, the general comorbidity of different psychopathologies can be shown at the sum-score level, but the specific symptoms that connect these clusters can only be identified at the more detailed item level. This is illustrated in the paper by Deserno et al. (2017), where the authors show how the relation between autism and well-being yields different information at different levels (item scores, subscale scores, sum-scores) and can be used to answer different research questions. Another option is to use latent network modeling, in which the indicators are modeled through the use of a latent node and independent measurement error (Epskamp, Rhemtulla, et al., 2017). Ultimately, what level to include in the network depends on the research question. The guiding principle should be to match the level of the included variables with the resolution at which inferences are ought to be made.
Centrality	**Local network properties**. Centrality is *not* a characteristic of a variable, but it is determined within the estimated network (see also *variable distribution* and *variable inclusion*; (Bringmann et al., 2019; Fried et al., 2018). Thus, a variable that is peripheral in one network may be central in another. For instance, the symptom of insomnia may be on the periphery of a depression network and of a generalized anxiety network. At the same time, it may connect the depression network to the generalized anxiety network and thus may be highly central in the combined network.

Box 5.1. What to watch out for, 'Methods' section.

5.3 Reporting standards for the 'Results' section

5.3.1 General analysis routine

5.3.1.1 Final sample size

As with general statistical guidelines (Appelbaum et al., 2018), all information regarding sample size should be reported. This includes all operations that are relevant to the sample size, such as removal of outliers and missing data, data imputation, data transformations, split-half approaches, etc. For further details please refer to Table 5.1 and Box 5.2.

5.3.1.2 Results of the accuracy and stability checks

Results on how accurate parameters are estimated (Epskamp, Borsboom, et al., 2018) should be reported. Usually, reports include plots giving information on bootstrapped confidence intervals (CIs), inclusion probabilities, or case-drop bootstraps, but which specific method to use is based on the choice of software. It is important to note that the bootstrapped confidence intervals discussed here cannot always be interpreted in the same manner as traditional confidence intervals (for detailed information, see Box 5.2 as well as Epskamp et al., 2018; Fried et al., 2022). Of note, which stability analysis to use is conditional on the research questions to be addressed (e.g., if centrality

is not analyzed, reporting stability results for centrality may not be relevant). For most existing analyses and research questions, stability analyses are available.

5.3.2 Analysis-specific routine

5.3.2.1 Network visualization

When a network plot is included in the manuscript, we recommend using a colorblind-friendly theme, as well as reporting: (1) What the edges represent (for example, partial correlations in the GGM or averaged logistic regression coefficients in the Ising model. In networks estimated using *mgm* (Haslbeck & Waldorp, 2020), edges between Gaussian variables can be interpreted as partial correlations, whereas relations that involve categorical variables can be interpreted in terms of (averaged) regression coefficients; for details on which type of coefficient is relevant, see Table 5.1); (2) Information about the plot, such as the size of the smallest and largest edges in the network and whether any specific visualization tools were used (e.g., in *qgraph*; Epskamp et al., 2012; whether a *minimum*, *maximum* or *cut* value were used when plotting the network); and (3) How the layout of the network was set (e.g., manually or using a pre-defined algorithm).

5.3.2.2 Network density and average absolute edge weights

The network density refers to the number of estimated edges relative to the total number of possible edges and is used to give an indication of the sparsity of the network. If the overall network structure is of interest, we recommend reporting the network density and average absolute edge weights. When visualized with *qgraph* (Epskamp et al., 2012), parameters adjust the color saturation and width of an edge to the absolute weight and scale relative to the strongest weight of the graph. One cannot get a clear notion of the average edge weight from visualization alone (Epskamp et al., 2012b), and thus reporting this is essential.

5.3.2.3 Centrality indices

If centrality is of interest (Costantini et al., 2015; Jones et al., 2019; Opsahl et al., 2010), we recommend including a supplementary table or appendix reporting the raw centrality scores in addition to visualizing raw centrality scores in the centrality plot itself,[29] as exact parameter values can often not be inferred from centrality plots with high precision. To assess the degree to which centrality estimates are subject to sampling error, we recommend reporting results of *centrality stability* (i.e., a case-drop bootstrap plot for the reported centrality indices), as well as the *correlation stability coefficient* (CS coefficient; Epskamp, Borsboom, et al., 2018). In addition, the bootstrapped difference test allows to test for differences in centrality between two nodes, which should be reported in case a centrality comparison between two particular nodes is of interest. The bootstrapped difference tests can also be used to compare specific edge pairs in a network, see *Specific nodes and edges*.

29 The default behavior in *qgraph* up to version 1.6.9 provides z-scores instead of raw-scores. This, however, may inflate dissimilarity between centrality indices, and we therefore recommend to use raw scores instead.

5.3.2.4 Predictability

The predictability of a node quantifies how well that particular node can be predicted by all remaining nodes (Haslbeck & Fried, 2017; Haslbeck & Waldorp, 2018, 2020). If predictability of nodes is of interest, we recommend specifying which predictability measure was chosen for which type of variable (e.g., R^2), and including the predictability measures in the network plot. In addition, we recommend including a supplementary table or appendix reporting the raw predictability scores, as exact predictability values typically cannot be inferred from the visualization.

5.3.2.5 Specific nodes and edges

If more specific features of the network are of interest, such as a particular edge A – B, we recommend reporting the stability of that particular edge. Likewise, if specific nodes are of interest, say node A, it is important to report the stability of the edges between node A and its connecting nodes, as well as the stability of the centrality for that particular node (see also *Centrality indices*). When comparing the strength of two edges, we recommend reporting the results of the bootstrapped difference test. These may also be informative in other settings, e.g., if one is interested in the overall stability of the network structure. Finally, if clustering of nodes is of interest, we recommend reporting the number of resulting clusters, as well as the stability of the clusters.

5.3.2.6. Group comparisons

When interested in comparing the network structure between different groups, we recommend reporting: (1) The sample size per group after data preprocessing choices (e.g., removal of outliers, removal of missing data, data imputation, data transformations); (2) whether a particular statistical test was used to compare the groups: the resulting *p*-values or Bayes Factors, and whether these were adjusted for multiple testing; and (3) whether the chosen comparison method allows, the stability of each network structure should be reported alongside the network comparisons.

When comparing networks visually, arguments used for visualization become crucial (e.g., *minimum*, *maximum*, and *cut* values; whether the same layout was used, etc.), as well as the correlation between the weighted adjacency matrices of the two (or more) network structures. We thus recommend: (1) Using the same layout when comparing network structures. Note that merely comparing networks visually may be misleading and is not recommended in isolation (e.g., without also carrying out a statistical test), even if the layout is fixed across networks (e.g., equal layouts might suggest that network structures are more similar than they actually are); and (2) setting the same value as the strongest edge in both networks (e.g., in *qgraph* by setting the same *maximum* value) in both network structures.

Features of the network structure	**Sparsity**. A central assumption of most of the models highlighted in the current manuscript is the *assumption of sparsity*, i.e., the true network structure can be expressed as a simplified, "sparse" network. If this assumption is violated, the performance of regularized estimation algorithms may be suboptimal (Epskamp, Kruis, et al., 2017), because many edges that are small but nonzero will be incorrectly set to 0. In this case, a nonregularized method (without model selection) can be used as an alternative (Williams et al., 2019), or the low-rank estimation approach proposed by Marsman and colleagues (2015). **Collider structures**. Collider structures occur when a variable is a common effect of two or more variables. If a true causal collider structure (A -> B <- C) underlies the data and the variables A and C are marginally uncorrelated or weakly positively correlated, then the undirected network could feature an edge between the causes (A – C), which is negative if both causal effects are positive. As such, collider structures can produce strong and unexpected negative edges in the network structure, which may hamper the interpretation of results. While there is no principled way to detect collider structures, one way to detect at least potential collider structures is by comparing the partial correlations to marginal correlations. If a partial correlation is of a different sign (e.g., negative) than a marginal correlation (e.g., positive), then this can signal conditioning on a collider (in this case, also check whether the two variables are both strongly connected to a third, which may be a common effect). **Network architecture.** When interpreting a network structure, it is important to keep an eye open for global features of the network. For instance, are there hubs in the network? How do these hubs influence the network structure? Is the network structure dense? Are there subnetworks? Global network aspects can inform and drive the interpretation of the network. Network architecture refers to the structure of the network as a whole; for instance, well-known architectures include small world, scale-free, and random graphs (Newman, 2018). Network architecture has been suggested to influence the recovery of the network structure (van Borkulo et al., 2014). For example, if a network features locally dense structures in the form of strong hubs (as in a scale free network), regularized estimation may have trouble recovering this (as it promotes sparsity). In contrast, in a ring graph (as e.g. used by Epskamp & Fried, 2018) each node has only two neighbors, which a regularized estimation technique can easily recover.
Network visualization	**Plotting algorithms.** Network plots are always dependent on the chosen plotting settings, i.e., settings that determine the spatial position of nodes in the network. Some plotting algorithms, such as the Fruchterman-Reingold algorithm (Fruchterman & Reingold, 1991a), can be sensitive to small changes (e.g., small differences in edge weights). Although network plots are informative visual representations, the exact placement of nodes should not be interpreted as standing in a one-to-one relation with features of the data. In order to arrive at representations that optimally represent patterns in the data, one may utilize MDS-based algorithms (Jones et al., 2018).

Unstable network structures	**Accuracy and stability.** Network stability is typically assessed by investigating whether the same ordering of edge strengths or centrality estimates arises across random subsamples of the data. Importantly, an unstable network structure does not necessarily imply that the analysis failed and the network should be discarded. This is because there are two reasons why orderings of edges may be unstable under bootstrapping: (1) there are estimation problems (e.g. *N* is too small), and (2) all edges are equally strong so that there is no ordering in the first place (e.g., the network is a Curie-Weiss model; Marsman et al., 2018). However, unstable network structures do limit the interpretation of the network (e.g., if the centrality ordering is unstable for whatever reason, centrality differences should not be interpreted). In general, instability should be acknowledged, and findings from unstable network models should be presented with caution. **Using bootstrapped confidence intervals.** Unless saturated (no model selection or regularization) maximum likelihood estimation is used, we argue against checking if bootstrapped CIs do (not) include 0, because the model selection methods themselves are already designed to put edges to zero. Therefore, doing additional checks on the CIs may lead to double thresholding. To this end, bootstrapped CIs of, for example, regularized network edges should never be used to assess for "significance" of edges (Fried, Epskamp, et al., 2022), and seeing bootstrapped CIs that include zero is in no way evidence for instability or inaccuracy of parameter estimates. Rather, the width of CIs reflects the accuracy of parameter estimates, irrespective of whether they include 0 or not (Epskamp, Borsboom, et al., 2018). Wide confidence intervals imply caution in interpretation, especially when interpreting the strength of edges, or the presence of weaker edges. While a clear definition of what *wide* represents is not established, this resolution can be driven by the specificity of a research question. For example, if the research question focuses on a specific edge (as for example done in Blanken et al., 2020, then it is particularly important to investigate the stability and accuracy of that edge: the wider the bootstrapped CI is for that edge, the less confidence we can attach to the estimate, and the more careful our inferences should be. **Case-drop bootstrapping.** To assess the stability of centrality indices, an alternative method must be used, the *case-dropping bootstrap.* This is because centrality indices rely on absolute edge weights, and consequently, an edge weight of 0 is at the boundary of the parameter space. Bootstrapping parameters near the boundary of the parameter space is highly problematic and leads to false inferences. Since edge weights of 0 are to be expected in PMRFs, Epskamp et al. (2018) propose an alternative method to circumvent this problem by correlating the centrality indices from the whole sample with centrality indices obtained through estimating networks on subsets of the sample (i.e., the case-dropping bootstrap). Epskamp, Borsboom, et al. (2018) term this *stability* (of the centrality rank order), as such correlations cannot say how *accurate* centrality estimates are. For example, suppose that all nodes in a network feature the exact same centrality. Then, any differences in centrality are due to chance, and we should expect these correlations then to be low even if the centrality measures are closely estimated to their true values (Borsboom et al., 2017).

Missing data	**Missing values**. It should be noted that not all estimators can handle missing data (see *Table 5.1*). Besides the use of (multiple) imputation strategies, which have not yet been studied in detail for network models, there are currently two ways for handling missing data when estimating GGMs. First, some estimators, such as *EBICglasso* and *ggmModSelect*, only require a correlation matrix as input, which can be estimated using pairwise observations. The *bootnet* package (Epskamp, Borsboom, et al., 2018) does this by default for these estimators and will use the average of pairwise sample sizes as a proxy for the sample size (e.g., for BIC computation; Epskamp, 2020c). Specifically, the sample sizes used when estimating each pairwise correlation separately are computed, and the average of these is taken as the final sample size in the analyses. Second, the *psychonetrics* package includes full information maximum likelihood estimation (Epskamp et al., 2020), which will only use observed data to estimate the network structure. We recommend to include the portion of missing data, as well as to consider and report any potential source of systematic missingness. If such systematic influences are present, using any statistical strategy can lead to problematic inferences because accurate inferences will depend on strong assumptions regarding the missingness mechanism (e.g., that data are missing at random or missing completely at random; Rubin, 1976). An example of such a systematic influence would be that missingness primarily occurs in participants with specific clinical features, such as high symptom levels.
Error rate	**Error rate**. The error rate, as well as the circumstances under which the error rate changes, should be considered. It is thus essential for researchers to consider whether they are favoring the sensitivity (true positive rate) or the specificity (true negative rate) of a model. Some estimation techniques (e.g., the *EBICglasso* algorithm; Epskamp & Fried, 2018) have high sensitivity but lower specificity. This means that weaker edges in the estimated network may be more prone to be false positives (i.e., Type I errors). Other estimation routines may be more conservative, retaining high specificity but featuring lower sensitivity (i.e., some edges may be missing from the network). As is typically the case in diagnostic situations, researchers face a trade-off between sensitivity and specificity: if one is more lenient to include edges in the estimated network, sensitivity will increase at the cost of specificity. Researchers can choose to err on the side of *discovery* (favor sensitivity over specificity) or to err on the side of *caution* (favor specificity over sensitivity). This choice is also driven by the research question. For example, in the study by Isvoranu and colleagues (2020), the aim was to identify edges between a polygenetic risk score and symptoms, which are generally weaker than edges between symptoms themselves. While good sensitivity is required to identify such small edges (and this was achieved in the paper as a result of a large sample size), high specificity is essential to justify interpreting the smaller edges in substantive terms. The authors therefore chose *ggmModSelect* as an estimator, which has been shown to have good specificity in large sample sizes (Isvoranu & Epskamp, 2021).

Box 5.2. What to watch out for, 'Results' section.

5.4 Illustrative examples

To illustrate the highlighted norms and reporting standards, we provide two examples of network analyses on openly available data, with two distinct research goals, in the supplementary materials. Both examples contain the elements described under the general analysis routine, as well as analysis specific elements matched with the indicated research goal. For an overview of elements covered in both examples, see Table S5.1. This table may also be used as a summary checklist of the chapter. First, using data from the empirical analysis in chapter 8, we aim to highlight the analysis specific routine on group comparisons, network visualization, and global network properties. Second, using open data (https://openpsychometrics.org/tests/TMAS/) collected on the *Taylor Manifest Anxiety Scale* (Taylor, 1953), we aim to highlight the analysis specific routine elements on centrality, differences between edges, network visualization, and local network properties.

Table 5.1. Overview and detailed information of commonly applied estimation routines in R.

Model (data)	Parameter interpretation	<package name>::<main function>	Description	Input type*
Ising Model (binary)	Logistic regression coefficients / loglinear interactions	IsingFit::IsingFit	Regularized estimation nodewise logistic regressions and EBIC model selection	Raw data (0/1 encoded)
		IsingSampler:: EstimateIsing	Unregularized estimation using psuedolikelihood, loglinear modeling or univariate logistic regressions	Raw data (any binary encoding)
		psychonetrics::Ising	Maximum likelihood estimation	Raw data (any binary encoding) or summary statistics (means + covariance matrix)
Gaussian Graphical Model (normal or ordinal data)	The parameters (i.e., edges) represent the unique association among two variables, after conditioning on all other variables in the network****	qgraph::EBICglasso	Regularized estimation using glasso and EBIC Model selection.	Variance-covariance / correlation matrix
		qgraph::qgraph(..., graph = "pcor")	Unregularized network (saturated).	Variance-covariance / correlation matrix
		qgraph::ggmModSelect	Unregularized estimation using extensive model search	Variance-covariance / correlation matrix
		psychonetrics::ggm	(Full information) maximum likelihood estimation	Raw data or summary statistics (means + covariance matrix)

Main defaults	Bootnet default set	Bootnet default differences	Missing data handling	Notes
Gamma hyperparameter is set to 0.25 and an AND-rule is used (both regression estimates required to be nonzero).	"IsingFit"	Automatic missing data removal (listwise) and automatic median split if input data is not binary	None (rows with missing data need be removed before analysis)	Regularization via *glmnet* package.
Pseudolikelihood estimation (method = "pl").	"IsingSampler"	Loglinear model used for up to 20 nodes and univariate logistic regressions for >20 nodes	None (rows with missing data need be removed before analysis)	Saturated or pre-defined model only (bootstrap threshold with bootnet::bootThreshold)
psychonetrics model**	N/A	N/A	Listwise deletion (pairwise covariance matrix can potentially be used as input)	Not possible with many (over 20) nodes.
Gamma hyperparameter is set to 0.5	"EBICglasso"	Bootnet correlation defaults used when raw data is used as input***	Pairwise deletion (sample size can be set to average of sample sizes for each pair of variables)	Poor performance in large sample sizes with dense network structures
Saturated model (all edges included).	"pcor"	Bootnet correlation defaults used when raw data is used as input***	Pairwise deletion (sample size can be set to average of sample sizes for each pair of variables)	Edges that are not significant (based on *p*-values or bootstraps) can be hidden, but no model selection is performed.
Gamma hyperparameter is set to 0 (BIC)	"ggmModSelect"	Bootnet correlation defaults used when raw data is used as input***	Pairwise deletion (sample size can be set to average of sample sizes for each pair of variables)	Slow with many nodes (>30) unless stepwise = FALSE is used. Note that this setting, however, also has its drawbacks (Isvoranu & Epskamp, 2021), and the decision should not only be based on the complexity of the network but also on the implications for the algorithm.
psychonetrics model**	N/A	N/A	Missing data handling through full information maximum likelihood is supported with estimator = "FIML"	Ordinal data supported with ordered = TRUE (uses weighted least squares estimation).

Table 5.1. Continued

Model (data)	Parameter interpretation	<package name>::<main function>	Description	Input type*
Mixed Graphical Model (normal / categorical / count)	(logistic / linear / multinomial) regression weights based on standardized data	mgm::mgm(..., lambdaSel = "EBIC")	Regularized nodewise regressions with EBIC model selection, potentially with interaction effects (3-way, 4-way, etcetera)	Raw data
		mgm::mgm(..., lambdaSel = "CV")	Regularized nodewise regressions with *k*-fold cross-validation model selection, potentially with interaction effects (3-way, 4-way, etcetera)	Raw data
Correlation Network (any)	Bivariate marginal correlations	qgraph::qgraph(..., graph = "cor")	Bivariate estimation	Variance-covariance / correlation matrix
		psychonetrics::corr	Maximum likelihood estimation	Raw data or summary statistics (means + covariance matrix)
Relative Importance Network (continuous)	Normalized *lmg* metric relative importance masures	relaimpo::calc.relimp	Relative importance	Raw data

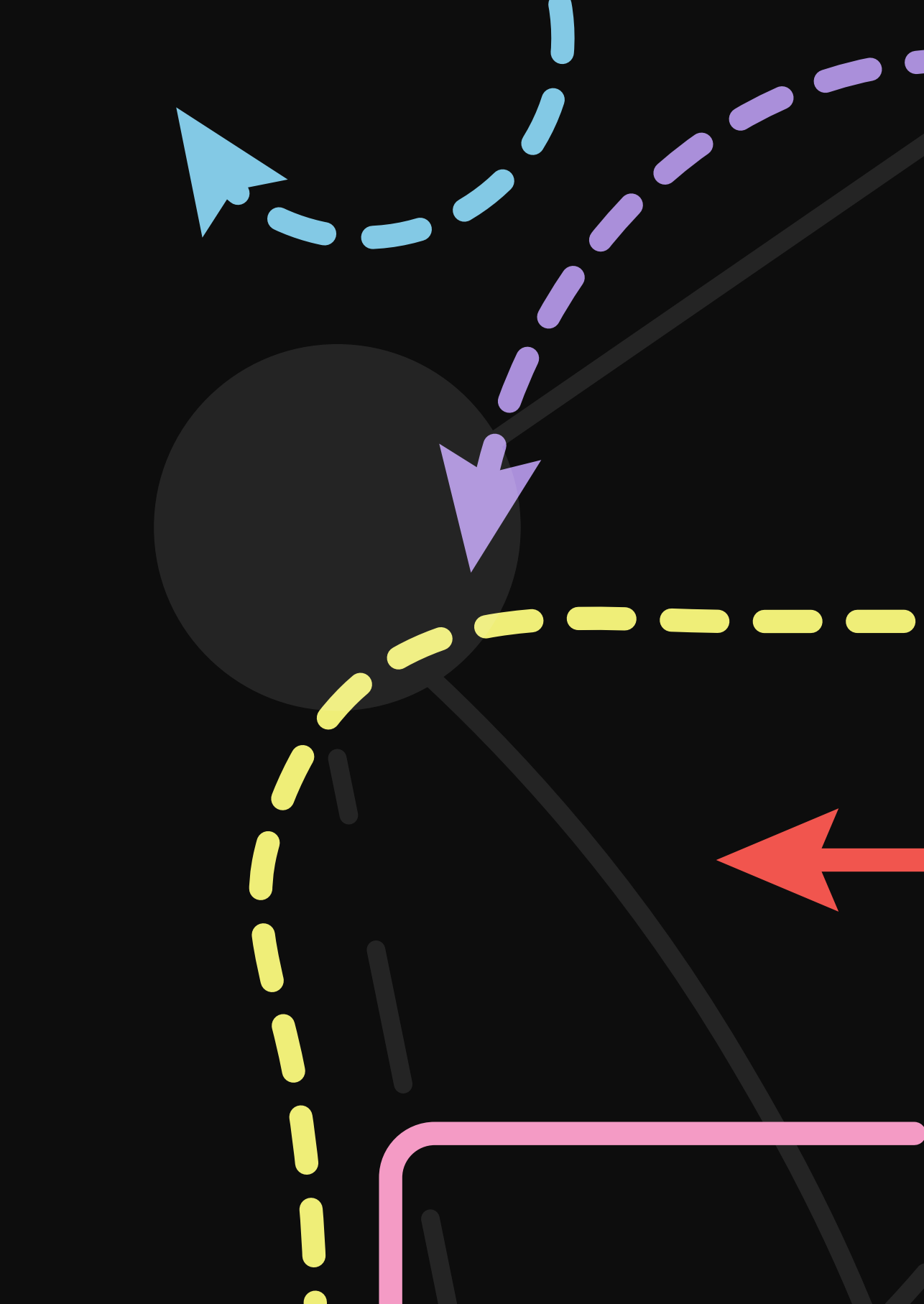
Main defaults	Bootnet default set	Bootnet default differences	Missing data handling	Notes
Gamma hyperparameter is set to 0.25	"mgm" (criterion = "EBIC")	Gamma hyperparameter is set to 0.5, type and level arguments are automatically set, edges are automatically signed if possible, listwise deletion automatically applied to data.	None (rows with missing data need be removed before analysis)	Default when using *bootnet* but not when using *mgm*. Reduces to Ising model with only binary variables. Edge weights between continuous variables are **not** partial correlation coefficients (but have the same interpretation)
Number of folds is set to 10	"mgm" (criterion = "CV")	type and level arguments are automatically set, edges are automatically signed if possible, listwise deletion automatically applied to data.	None (rows with missing data need be removed before analysis)	Default when using mgm but not when using *bootnet*. Reduces to Ising model with only binary variables. Edge weights between continuous variables are **not** partial correlation coefficients (but have the same interpretation)
Saturated model (all edges included).	"cor"	Bootnet correlation defaults used when raw data is used as input***	Pairwise deletion (sample size can be set to average of sample sizes for each pair of variables)	Edges that are not significant (based on *p*-values or bootstraps) can be hidden, but no model selection is performed.
psychonetrics model**	N/A	N/A	Missing data handling through full information maximum likelihood is supported with estimator = "FIML"	Ordinal data supported with ordered = TRUE (uses weighted least squares estimation).
Saturated model (all edges included).	"relimp"	No automated function outside of bootnet wrapper	None (rows with missing data need be removed before analysis)	Returns directed (not causal) network

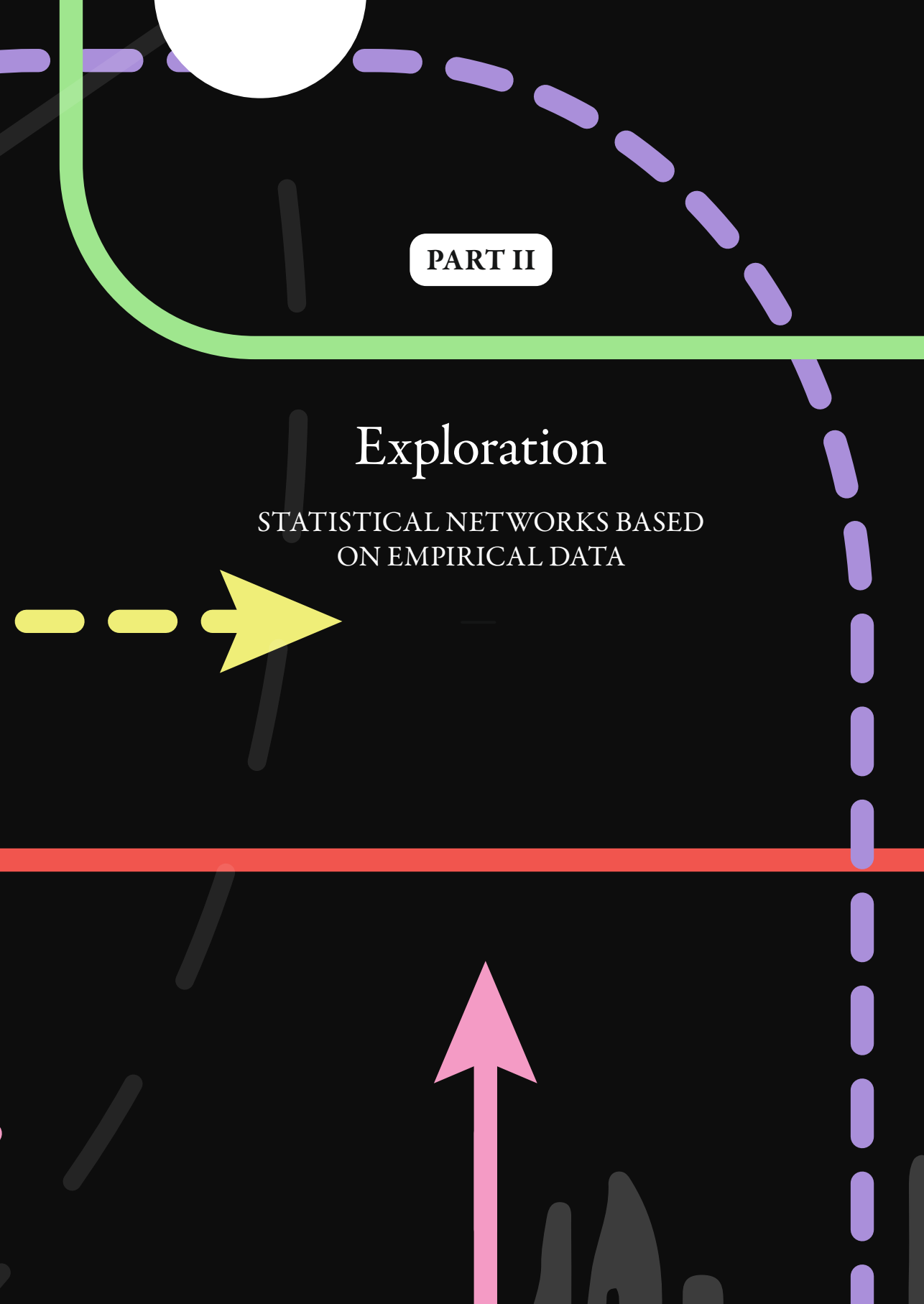
5.5 Conclusion

As clear norms have not yet been established in the network literature, the current chapter explicates minimal shared norms in reporting psychological network analyses. While network psychometrics is a relatively young field of research, we recognize that many norms discussed here have important implications for commonly used inferences. We therefore included two "what to watch out for" boxes, where we discussed important considerations for network analysis, as well as potential sources of misinterpretation of network structures.

It should be noted, however, that our description of validity threats is not exhaustive and subject to ongoing research. For example, although robustness analyses allow one to assess the uncertainty of claims based on the model (relative to sampling error), methods for assessing the goodness-of-fit of the model as a whole remain underinvestigated (although model fit assessment techniques are available for confirmatory network analyses; Epskamp, 2020b). Currently, operational network analysis techniques are better viewed as exploratory analysis and visualization tools in the tradition of Tukey (Tukey, 1977), or as phenomena-detection tools that can generate a starting point for theory formation (Borsboom, van der Maas, et al., 2021; Haig, 2005, 2014), than as confirmatory theory-testing approaches in the tradition of SEM (Hoyle, 2012). Hence, we currently advise against strong inferences based on network analyses alone, while noting that considerable methodological research opportunities are open to extending network analysis in this direction (Epskamp, 2020b).

Clear reporting standards for network psychometrics improve transparency, which is necessary for reproducibility. Only if the scientific community can follow exactly what analyses were conducted can we vet inferences drawn by respective authors. This is especially relevant in a field that is still fairly novel such as network psychometrics, where we encounter new challenges regularly. Overall, we trust the highlighted directions to aid researchers in identifying elements of their analyses that are important to include in a scientific report, as well as to make empirical network studies more rigorous.




PART II
Exploration
STATISTICAL NETWORKS BASED
ON EMPIRICAL DATA

CHAPTER 6

NETWORK ANALYSIS OF ANXIETY SYMPTOMS DURING THE COVID-19 PANDEMIC

CHAPTER 6

Abstract

To understand the interplay between anxiety symptoms and their maintaining psychological processes in the population, an analysis of longitudinal within-person relationships is required. A sample of 1,706 individuals completed daily measures during a 40-day period with strict mitigation protocols. Data of 1,368 individuals who completed at least 30 assessments were analyzed with the multilevel vector autoregressive (mlVAR) model. This model estimates a temporal, a contemporaneous, and a between-person network. Uncontrollability of worry, generalized worry, fear of being infected, fear of significant others being infected, and threat monitoring had the highest outstrength within the temporal network, indicating that daily fluctuations in these components were the most predictive of next-day fluctuations in other components. Of specific connections, both fear of self and fear of close others being infected predicted generalized worry and threat monitoring. In turn, generalized worry and threat monitoring engaged in several positive feedback loops with other anxiety symptoms and processes. Also, intolerance of uncertainty was predictive of other components. The findings align with the mechanisms both in the metacognitive therapy (MCT) model and in the intolerance of uncertainty model of generalized anxiety disorder (GAD).

This chapter has been adapted from: Hoffart, A., Burger, J., Johnson, S. U., & Ebrahimi, O. V. (2023). Daily dynamics and mechanisms of anxious symptomatology in the general population: A network study during the COVID-19 pandemic. *Journal of Anxiety Disorders*, *93*, 102658.

6.1 Introduction

The COVID-19 pandemic and the social distancing protocols used to impede the spread of the virus have been associated with an increase in adverse mental health symptoms (Salari et al., 2020). In particular, and not least due to the life-threatening nature of the virus, anxiety symptoms have increased (Salari et al., 2020). Most of the pandemic literature has focused its efforts on the identification of prevalence estimates and the risk factors associated with the alterations in anxiety symptom levels. There is a lack of studies investigating the cognitive and behavioral processes that contribute to the maintenance of anxiety symptoms and disorders, and therefore may be involved in their mechanisms of change. That is, changes in these processes could lead to changes in symptoms and disorders. Thus, a study of such processes could identify targets of intervention.

For a study of anxiety in general, the symptoms of generalized anxiety disorder (GAD) are a reasonable focus. GAD is one of the most common disorders in the population (Bandelow & Michaelis, 2015). Moreover, many of its symptoms – anxiety, generalized worry, inability to relax, restlessness, irritability, and fear of awful events – are also present in other anxiety disorders (American Psychiatric Association, 2013). Studies should also include panic attacks, which is a prominent transdiagnostic anxiety symptom (American Psychiatric Association, 2013), and symptoms of health anxiety, which are especially relevant during the present pandemic due to the threat of being infected.

Several transdiagnostic processes have been proposed to maintain anxiety disorders and symptoms (Harvey et al., 2004). Clinical processes based on research evidence include threat monitoring (i.e., selective attention to threat stimuli), recurrent thinking in the form of worry, avoidance behavior, and intolerance of uncertainty (Morris & Mansell, 2018). Notably, some of these processes (e.g., worry and avoidance) are also considered symptoms due to their inherent distressing or impairing features. The way in which many of these processes relate to each other and to anxiety symptoms is explained in the model underlying metacognitive therapy (Wells, 2009), a therapy that has shown promising outcomes (Normann & Morina, 2018). Metacognitive therapy (MCT) focuses on thought processes (e.g., worry, selective attention) and underlying metacognitive beliefs that are supposed to maintain emotional disorders. Worrying is directed to potential danger in the future (e.g., thoughts about being subject to rejection) and leads to exaggerated appraisals of danger and increased anxiety. More generalized worry (i.e., worry about many things) leads to wider threat monitoring which biases information processing, and thus inflates the sense of danger and anxiety. Threat monitoring leads to the detection of more dangers, thus extending the themes of generalized worry. Anxiety may reinforce worry through emotional reasoning: "I am anxious, therefore there is danger." Avoidance of perceived threat situations, both in order to prevent threat and anxious thoughts and feelings, precludes corrective experiences of their dangerousness and maintains anxiety. The repeated practice of worry contributes to a sense of loss of control over worry and to the development of negative metacognitive beliefs about the uncontrollability of worry (also a symptom of DSM-5 GAD; American Psychiatric Association, 2013) and about the dangerousness of thoughts, for instance, that some thoughts could lead to a loss of mind. These beliefs reinforce anxiety, and conversely, aspects of the anxiety experience, such as racing thoughts, may be taken as evidence of loss of control and thus strengthen the beliefs. Moreover, there is a reciprocal rela-

tionship between these two beliefs, as the belief that thoughts are dangerous strengthens the sense of their uncontrollability. Finally, the uncontrollability of worry influences generalized worry by representing an additional worry theme.

As an alternative to the MCT model, the intolerance of uncertainty model of GAD (Carleton et al., 2012) posits that uncertainty is experienced as a threat by itself and is the cause of key GAD symptoms. Worry is viewed as a strategy to remove uncertainty by mentally preparing for any eventuality; however, as complete certainty often is not achievable, worry persists and is easily experienced as out of control.

Previous research has also connected a multitude of situational and psychobiological state variables to anxiety, including interpersonal conflict (Nolte et al., 2011), social media use (Vannucci et al., 2017), physical activity (Ebrahimi, Hoffart, et al., 2021), sleep quality (Alvaro et al., 2013), and alcohol abuse (Kushner et al., 2000). These variables are relevant regarding their interactions with anxiety symptoms, both in pandemic and non-pandemic settings. Additionally, perception of sufficient information access about the pandemic and protective measures (Ebrahimi, Hoffart, et al., 2021) is specifically relevant in the current pandemic situation.

As change in maintaining processes leads to change in symptoms, these processes represent mechanisms of change. Most studies of mechanisms of change in anxiety so far have addressed the global level of disorder and not individual anxiety symptoms. That is, symptoms are viewed as effects and indicators of a latent disorder and aggregated in global (e.g., sum) scores. However, in the MCT model of GAD, mechanistic processes and individual symptoms are supposed causes of each other (e.g., more generalized worry elicits more uncontrollability of worry, which, in turn, leads to more anxious mood), and these causal interactions explain the co-occurrence of GAD symptoms. As outlined above, the intolerance of uncertainty model of GAD also focuses particular symptoms. The symptom level focus both of the MCT and the intolerance of uncertainty model makes them amenable to network analytic techniques which focus on the causal interaction of components of phenomena rather than of latent entities (Borsboom, 2017). Components (e.g., psychological processes, symptoms) are considered to have different causal roles in the network they constitute. For instance, the MCT model proposes that generalized worry is a particularly influential component in the anxiety network; that is, it is strongly connected to other components. In network analysis, centrality indices are used to estimate such interconnectedness (Opsahl et al., 2010). Given that mechanistic processes proposed by the MCT model are supposed to causally interact with symptoms as well as mediate the relationships between symptoms, they should be central components of a combined network of such processes and symptoms.

To date, most mechanistic studies have addressed between-person differences; that is, individual differences in mechanistic variables (e.g., intolerance of uncertainty) have been related to individual differences in outcome (e.g., anger) variables (Laposa & Fracalanza, 2019). However, mechanistic relationships concern covariance between changes in mechanistic variables and symptom changes *within* persons. Of note, between-person findings can only be generalized to within-person relationships under very strong and potentially unreasonable assumptions (Molenaar, 2004), which has been thoroughly demonstrated empirically (Fisher et al., 2018). A study of dynamic within-person processes necessitates a collection of longitudinal data, in which potential mechanisms and symptoms are repeatedly measured.

A few studies have used an intensive longitudinal design to estimate within-person networks of GAD symptoms. Anger and irritability were most and being worried was least influential in a study of individuals with GAD or major depressive disorder (Fisher et al., 2017) and anxiety predicted subsequent worry in a study of clients with mood and anxiety disorders who did not complete subsequent treatment (Lutz et al., 2018). During the beginning of the COVID-19 outbreak in the Netherlands, worry about health related to the pandemic predicted the GAD symptom generalized worry at the next measurement point among students (Fried, Papanikolaou, et al., 2022).

In a within-person network study, a time scale for the measurements should be used that reflects the speed of effects. An anxious mood has been shown to vary considerably from day to day (Maser & Cloninger, 1990), and daily anxiousness has been shown to predict daily depressed mood and other variables in clients with GAD and a history of depressive symptoms (Starr & Davila, 2012). Accordingly, a daily time scale appears to be appropriate for the longitudinal assessment of anxiety symptoms. Such daily measures can then be used to produce two types of within-person networks (Epskamp, Waldorp, et al., 2018). First, a temporal network illustrating predictive relationships between variables from one day to the next and, second, a contemporaneous network illustrating relationships between variables within the same day. In addition, a between-person representing the relationships between the person-means on the variables may be estimated.

The main purpose of the present study was to examine the within-person networks of anxiety symptoms, theory-derived cognitive and behavioral processes, and situational and psychobiological state variables in the general Norwegian population in a 40-day period during the COVID-19 pandemic. As temporal and contemporaneous networks show the within-person relationships between these variables, they could be used in an exploratory fashion to point to important targets for intervention. As detailed above and in the pre-registered protocol, we expected generalized worry, uncontrollability of worry, threat monitoring, and avoidance to be influential nodes (i.e., having high strength centrality) in the anxiety networks.

6.2 Methods

All elements of this study adhere to its preregistered protocol, which can be found at the online repository of the Center for Open Science (https://osf.io/zhakg/). All participants gave written informed consent to participate. The ethical approval for the study was granted by the Regional Committee for Medical and Health Research Ethics (reference: 125510).

6.2.1 Study type and design

This study is part of The Norwegian COVID-19, Mental Health and Adherence Project (MAP-19), encompassing a large ongoing longitudinal investigation of psychiatric symptomatology in the general adult Norwegian population during the COVID-19 pandemic. The present sub-study was an observational longitudinal study, conducting daily measures for 40 days. The 40-day period (i.e., February 17 to March 28, 2021) was characterized by a predominantly stable set of strict pandemic mitigation protocols, such as quarantine upon contact with infected individuals, isolation upon infection, closure of schools and universities, prohibitions of social gatherings, and travel and visitation restrictions. Some of these protocols (e.g., social gatherings, domestic travel, and

visitation restrictions) were slightly lightened during winter and start of Easter holiday intervals. These holiday intervals encompassed five weekends and two weekday periods. On February 17, 2021, the number of new infected cases was 358 and that of deaths from the virus was 4. Toward the end of March 2021, the number of new cases per day had gradually increased to about 1,000, whereas the number of deaths per day was still typically 4 (Worldometers, 2021).

6.2.2 Participants and procedure

Eligible participants included all adults (i.e., minimum age 18 years) residing in Norway. The participants were initially recruited in March 2020 through an online survey disseminated to a random selection of Norwegian adults through a Facebook Business algorithm and through national and local television, radio stations, and newspapers (for further information, see also chapter 7). Prior to the daily measurements conducted for the present study, the participants provided responses at four measurement waves since the onset of the pandemic. At the fourth wave of data collection (i.e., January 2021), the participants were questioned about participation in the present study, and 2,383 of them expressed interest, of which 1,706 individuals were formally enrolled in the study. During the 40-day period, the participants received a set of questions (items) every evening at 6:30 PM. Access to the questionnaire was closed at 10:30 PM.

6.2.3 Measurement

The participants reported their age, sex, civil status, education, presence of psychiatric diagnosis, region of residence, and whether they received treatment for anxiety. The 17 network items are listed in Table 6.1. Following Fried et al. (2022), the items were measured on a five-point response scale with scale information detailed in the table note. The items were selected with the aim of representing the theory-derived variables presented in the introduction. Of GAD symptoms, the core symptoms in the DSM-5 (American Psychiatric Association, 2013) were selected, using the three first items of the Generalized Anxiety Disorder-7 (GAD-7; Spitzer et al., 2006) scale. In addition, the GAD symptom irritability was included (GAD-7, Item 6) to be congruent with the items used in Fried et al. (2022). Fear of being infected was measured by Item 1 on the Obsessions with COVID-19 Scale (Lee, 2020). Threat monitoring, avoidance, and thoughts of loss of mind were measured by Item 2, 4, and 15 on the Cognitive-Attentional Syndrome-1 (CAS-1; Wells, 2009), intolerance of uncertainty was measured by Item 1 on the Intolerance of Uncertainty Short Form (Carleton et al., 2007), sleep satisfaction by Item 6 on the Bergen Insomnia Scale (Pallesen et al., 2008) and social media use by Item 10 used by Fried et al. (2022). Items for fear of close others being infected, physical activity, alcohol use, interpersonal conflict, and sufficient information were developed by the present authors. More details about the selection of items and their source are contained in the pre-registered protocol (https://osf.io/zhakg/).

Table 6.1. Items measured daily over the course of 40 consecutive days.

No.	Abbreviation	Item name and content
1	Anxiety	Anxiety: "Today I have been feeling nervous, anxious or on edge".
2	UnconWor	Uncontrollable worry: "Today I was not able to stop or control worrying".
3	GenerWor	Generalized worry: "Today I worried about a lot of different things".
4	Irritabi	Irritability: "Today I was easily annoyed or irritable".
5	SelfIInf	Fear of being infected: "Today I had disturbing thoughts that I may have caught the coronavirus".
6	PanicAtt	Panic attacks: "Today I feared experiencing sudden attacks or episodes where I feel anxious, scared, or panicky".
7	OtherInf	Fear of others being infected: "Today I thought a lot about that someone I know may become infected or die from the coronavirus".
8	ThreatMo	Threat monitoring: "Today I focused my attention on things I find threatening (such as bodily sensations, negative thoughts, possible infection, or danger)".
9	Avoidance	Avoidance: "Today I avoided at least one situation to protect myself from negative feelings or thoughts".
10	ThouMind	Thoughts lead to losing mind: "Today I thought that some thoughts could make me lose my mind".
11	IntolUnc	Intolerance of uncertainty: "Today I felt stressed by unforeseen events".
12	PhyActiv	Physical activity: "Today I spent ..... minutes/hours physically exercising to the extent that it lead to increased pulse or at least light sweating".
13	AlohUse	Alcohol use: "Today I drank ... units of alcohol".
14	InterCon	Interpersonal conflict: "Today I argued or had negative discussions with someone".
15	SleepSat	Sleep satisfaction: "Today I was satisfied with my sleep".
16	SuffInfo	Sufficient information: "Today I find that I have received enough information about how to deal with the pandemic and its associated protocols".
17	SocMedia	Social media use: "Today I spent ..... minutes/hours scrolling social media just to make the time pass".

Note. All Items were measured on a five-point scale (1–5). *Items 1–13*: 1 (Not at all), 2 (Slightly), 3 (Moderately), 4 (Very), 5 (Extremely). *Items 14–16*: 1 (0 min), 2 (1–15 min), 3 (15–60 min), 4 (1–2 h), 5 (Over 2 h). Item 15 below is measured on a 5-point scale: 1: No fear of such an attack; 2: Mild fear of such an attack; 3: Moderate fear of such an attack; 4: Strong fear of such an attack; 5: Actually experienced at least one such attack today. *Item 16*: 1 (0 min), 2 (10–15 min), 3 (15–30 min), 4 (30–60 min), 5 (Over 1 h). Item 17 is measured on a 5-point scale: 1 (0 units); 2 (1 unit); 3 (2 units); 4 (3 units); 5 (4 or more units).

The theory-based selection was followed by a data-driven procedure, affirming that the correlation matrix was positive definite and that the included items were not linear combinations of one another. Then, the goldbricker algorithm (Jones et al., 2021) was used to search for pairs of highly correlated items, in addition to items displaying similar behavioral patterns to the other items in the network. Dependent correlations were examined using the Hittner method (Hittner et al., 2003). This data-driven procedure supported the theoretical selection, identifying no redundant items.

6.2.4 Statistical analyses

As discussed in chapter 3, because the network analyses are based on within-person centering using sample means per person, it is not recommended to include individuals with less than 20 measurements (Epskamp, Waldorp, et al., 2018). The optimal balance between including participants with minimal missingness and retaining as many participants as possible was found to be at 30 completed assessments. This resulted in the inclusion of 1,368 out of the 1,706 participants.

The applied analyses of the time series assume that they are stationarity. Therefore, linear mixed models were used to examine whether there were systematic changes in the network variables over the course of the 40-day observation period. The SPSS 27.0 program was used. Any linear time trend and weekday versus weekend (Saturday and Sunday) effect were removed from each variable by subtracting these effects from each observation.

Then, the multi-level vector auto-regressive (mlVAR) method, implemented in the *mlVAR* package for R (version 4.1.0; Epskamp et al., 2023) was used to analyze the multivariate time series in the 1368 subjects. The R-code and the edge weight matrices are found here (https://osf.io/zhakg/). Three networks are estimated in *mlVAR*. The temporal network represents the average lagged within-person associations between the variables from one day to the next, controlled for each other. The contemporaneous network represents the within-person associations between the variables within the same day, controlled for each other and the temporal effects. The between-person network represents the associations between the person-means on the variables, given the person-means on the other variables. In networks, the links are called edges and the variable nodes. The networks were visualized using the *qgraph* package in R (Epskamp et al., 2012a). The arrangement of the nodes was based on the average layout of the networks that have been established via the Fruchterman-Reingold algorithm (Fruchterman & Reingold, 1991b). A significance level of $p < .001$, two-tailed, for visualization of edges in the networks was used throughout.

Network characteristics were also assessed in the form of centrality indices (Opsahl et al., 2010). These parameters indicate how central the position of a node is in the network (see Figure 6.3 and 6.4 for explanation of the indices). A novel visualization approach was used to visualize the different centrality metrics by the use of radar plots, portraying the contrasts in centrality across the different networks more clearly (Ebrahimi, Burger, et al., 2021; Fruchterman & Reingold, 1991).

Within the process measures of the 1,368 included participants, there were occasional missing completion of items, amounting to 3,986 (7.3%) of the expected number of 54,720.

6.3 Results

6.3.1 Sample characteristics and representativeness

The age of the 1,706 participants enrolled in the study ranged from 18 to 86 years (M = 37.3), with 1336 (78.5%) of the participants being female (population proportion: 49.8%), 962 (56.9%) having a university degree (population proportion: 31%), and 830 (49.4%) being married or in a civil partnership, and 67 (4.9%) received treatment for anxiety. Of the participants, 1,368 (80.2%) provided sufficient data to be included in the study, with no differences found between those with and those without sufficient data. The percentage of individuals with a pre-existing psychiatric diagnosis in this sample was 16.6%, representative of the known rate of psychological disorders in

the adult population of Norway, which is between 16.7% and 25.0% (Tesli et al., 2016). The sample was further geographically representative of Norway, with the number of participants sampled from each region being proportional to region size.

6.3.2 Sensitivity analyses

Demographic characteristics not fully representative of the Norwegian adult population were adjusted in sensitivity analyses encompassing a random selection of 598 participants fully matching the population characteristics (i.e., including sex and education). These sensitivity analyses replicated the results from the main sample across all analyses below, with the correlation between the matrices containing the results of the representative sample and the main sample ranging from .971 to .999.

6.3.3 Changes over the observation period

Each of the 17 variables was used as dependent variables in mixed models using day number and weekday versus weekend as predictors (see Supplement B, Table S6.1). The symptoms generalized worry, fear of being infected, and fear of close others being infected increased over the observation period, as did the processes of threat monitoring and intolerance of uncertainty. Sleep satisfaction also increased with time. Physical activity, alcohol use, sleep satisfaction, and social media use were higher during weekends. During weekends, the participants scored lower on the other variables, except interpersonal conflict.

6.3.4 Networks

Figure 6.1 visualizes the temporal network of significant connections between the variables from one day to the next, and the radar charts in Figure 6.3 depict the variables' outstrength and instrength centrality. The variability (SDs) of the temporal edges across the participants is presented in Figure S6.1. 'Uncontrollability of worry', 'generalized worry', 'fear of self being infected', 'fear of others being infected', and 'threat monitoring' had relatively high outstrength centrality, indicating that they predicted other variables to a large extent. 'Anxiety', 'uncontrollability of worry', 'generalized worry', and 'threat monitoring' exhibited high instrength centrality, indicating that they were the most predicted by other variables. The specific connections are addressed in the discussion part.

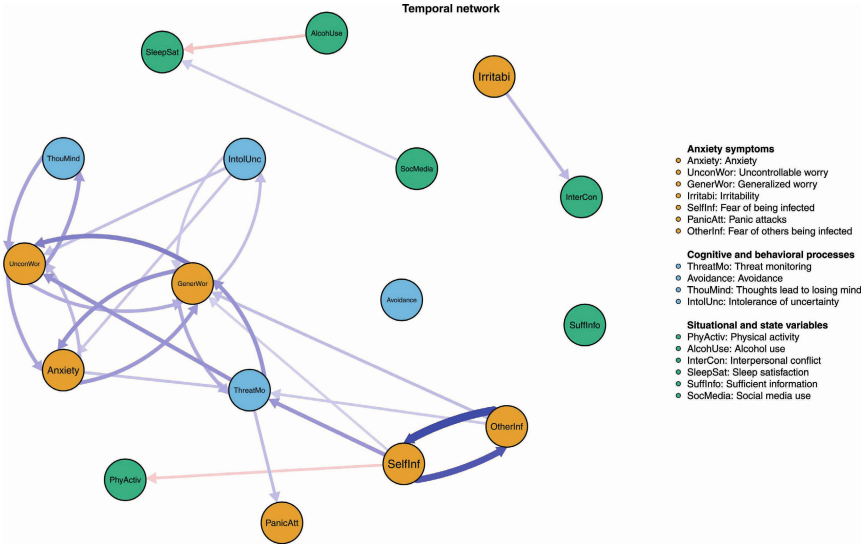


Figure 6.1. Temporal network derived from the multi-level vector auto-regressive (mlVAR) model. The temporal network shows the significant ($p < .001$) connections between the nodes, while controlling for all other nodes in the network. The thickness of an edge represents the strength of connection, relative to the strongest edge coefficient, which was the auto-regressive coefficient of .27 for the node 'SuffInfo' (i.e., 'sufficient information'). Blue edges represent positive connections and red edges represent negative connections. All the auto-regressive edge weights were significant but were omitted from the figure to enhance clarity of presentation.

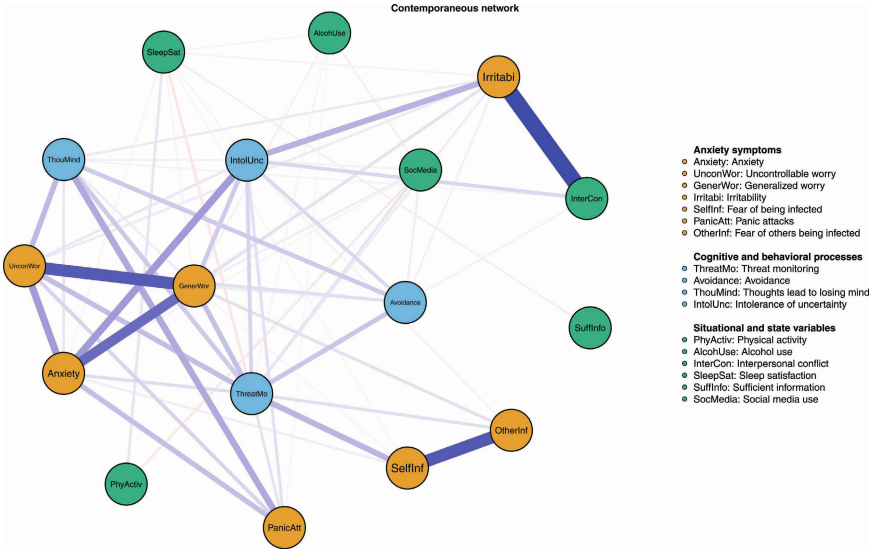


Figure 6.2. Contemporaneous network derived from the multi-level vector auto-regressive (mlVAR) model. The contemporaneous network shows the significant ($p < .001$) connections between nodes, while controlling for all other nodes in the network in addition to controlling for the temporal effects. The thickness of an edge represents the strength of connection, relative to the strongest edge coefficient, which was .44 between 'Irritabi' (i.e., 'irritability') and 'InterCon' (i.e., 'interpersonal conflict'). Blue edges represent positive connections and red edges represent negative connections.

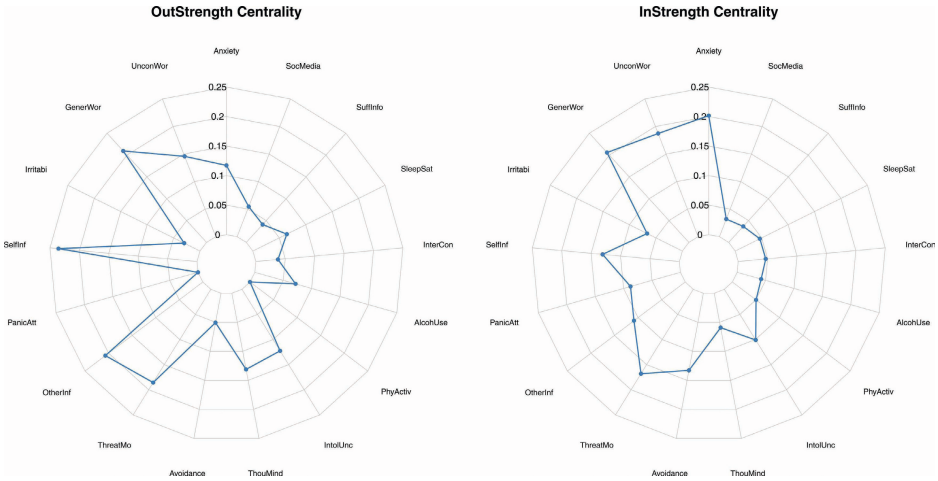


Figure 6.3. Radar chart showing the outstrength and instrength centrality in the temporal network.[30]

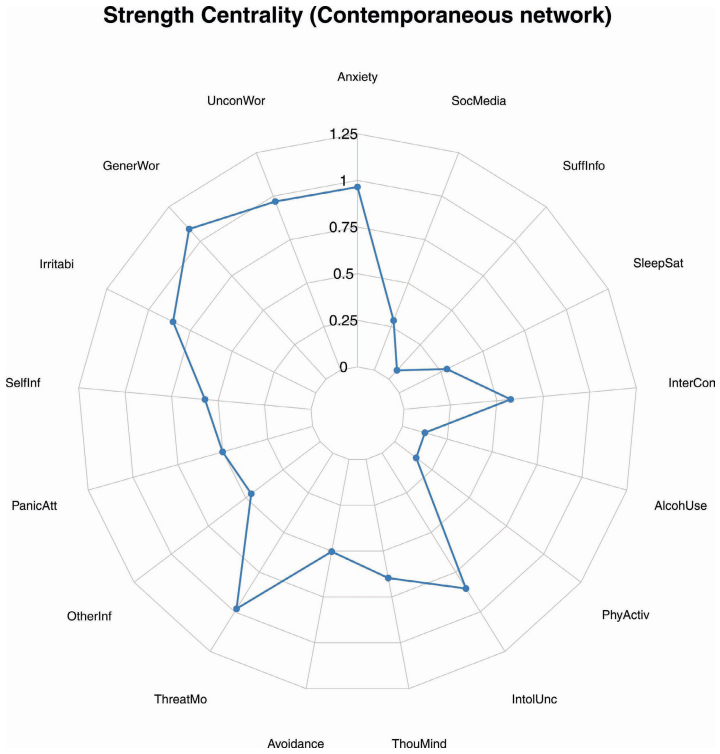


Figure 6.4. Radar chart showing the strength centrality in the contemporaneous network.[31]

30 Outstrength centrality is the sum of all outgoing absolute edge weights from a node. Instrength centrality is the sum of all incoming absolute edge weights to a node.

31 Strength centrality is the sum of all absolute edge weights connected to a node.

The contemporaneous network in Figure 6.2 shows the associations between the within-person fluctuations on the variables within the same day. 'Anxiety', 'uncontrollability of worry', 'generalized worry', 'irritability', 'threat monitoring', and 'intolerance of uncertainty' had high strength centrality (see Figure 6.4). Among the numerous connections, there were marked connections between the GAD symptom 'generalized worry' and the symptoms 'anxiety' and 'uncontrollability', between the two infection fears, and between 'irritability' and 'interpersonal conflict'. Notably, the GAD symptom of 'irritability' had only tiny connections to the other GAD symptoms.

The between-person network is of secondary interest here and is presented in the supplementary material (Supplement B, Figure S6.2). It shows numerous connections. Of note, the variables alcohol use and social media use were not connected to sleep satisfaction.

6.3.5 Network replicability

Details of the replicability analyses are provided in Supplementary information (Supplement B). The Pearson correlations between edges weights across different subsamples were high, ranging from .906 to .999. Centrality estimates were consistent across subsamples, correlations ranging from .700 to .995. Finally, 90% of the edges with the highest centrality were re-obtained in the subsample analyses across all networks.

6.4 Discussion

The main purpose of this study was to examine the within-person dynamics of anxiety symptoms, theory-derived cognitive and behavioral processes, and situational and psychobiological state variables in a period of lockdown during the COVID-19 pandemic. As a preliminary step, we examined whether the variables changed over the course of the observation period. Some of the anxiety symptoms and processes, including 'fear of being infected', increased, possibly reflecting the gradually increasing infection rate over the period and the duration of being in a period of strict mitigation protocols. All the symptoms and processes were lower on weekends. This may reflect that many participants, especially those in work, were free to adhere more to social distancing protocols during weekends and therefore felt less exposed to possible contagion. Another possibility is that they experienced more well-being during some of the weekends due to a slight lightening of protocols during holiday periods. These holiday periods coincided more with weekends than with weekday periods.

The temporal network represents the predictive relationships between the components, controlled for each other. 'Uncontrollability of worry', 'generalized worry', 'fear of being infected', 'fear of significant others being infected', and 'threat monitoring' had the highest outstrength within the network, indicating that these components were the most predictive of other components. 'Uncontrollability of worry', 'generalized worry', and 'threat monitoring' also had the highest instrength centrality within the network, indicating that they were predicted by other components. In addition, the core GAD symptom 'anxiety' had high instrength. The balance of a node's out- and instrength suggests its role in the network. For instance, 'fear of being infected' had a high outstrength and a medium instrength, indicating that it may be more a driving mechanistic variable than a passive outcome variable. For 'anxiety', on the other hand, the balance was opposite, suggesting that it may be mainly an outcome variable.

Overall, the obtained pattern of temporal connections aligns with the mechanisms in the MCT model of GAD. An external stressor (COVID-19) elicits fear of being infected and fear of significant others being infected, and these two fears reinforce each other. They both influence the more general processes of threat monitoring and generalized worry about many things. In turn, generalized worry engages in positive feedback loops with threat monitoring, anxiety, and uncontrollability of worry. A positive feedback loop also emerges between the uncontrollability of worry and the belief that thoughts could lead to loss of mind. One deviation from the expectations derived from the MCT model occurred: 'avoidance' did not predict anxiety symptoms. The findings are consistent with the finding in the study of students that COVID-19 worry led to generalized worry (Fried, Papanikolaou, et al., 2022).

The temporal network also aligns with the mechanisms in the intolerance of uncertainty model of GAD. 'Intolerance of uncertainty' predicted 'anxiety', 'uncontrollability of worry', and 'generalized worry'. Conversely, more 'generalized worry' predicted more 'intolerance of uncertainty'.

The contemporaneous network represents the associations of features within the same-day window of measurement, controlled for each other and the temporal effects. This network also depicts within-person relationships and thus reflects effects occurring within individuals. However, these effects occur faster than from day to day, and the effects are bidirectional. Of marked connections, the belief that thoughts could lead to loss of mind was related to 'panic attacks', which reasonably reflects the fast occurrence of the effect of this belief. Other marked contemporaneous connections – between the three core symptoms of GAD, between 'anxiety' and 'intolerance of uncertainty', between 'fear of self' and 'fear of others being infected', and between 'irritability' and 'interpersonal conflict' – underlined obtained temporal connections. Notably, the GAD symptom of 'irritability' had only tiny connections to other symptoms in the contemporaneous network an no connections to other symptoms in the temporal network. This raises the question of whether irritability reasonably belongs to the GAD diagnosis. However, this may reflect the nature of the sample as anger and irritability has been found influential in a clinical sample (Fisher et al., 2017a).

Although the main focus of this chapter lies on investigating within-person relationships, we additionally estimated a between-subjects network to inspect between-person effects in an exploratory fashion. Interestingly, we could not find any between-person relationship between 'social media use' and 'alcohol use' with 'sleep satisfaction'. Negative between-person associations are well-established in the literature (Levenson et al., 2016; Stein & Friedmann, 2006). A reason for this discordance may be that several other variables, for instance anxiety symptoms, were controlled in the present study. Interestingly, 'alcohol use' negatively and 'social media use' positively predicted 'sleep satisfaction' on the within-person level. This pattern of findings underlines the need to study psychological variables at both levels.

A strength of this study was that the network focus on observables instead of latent constructs led to more differentiated findings. From such findings, more specific clinical recommendations can be inferred. A further strength was that dynamic within-person rather than between-person relationships were examined. This is an asset because theories of symptomatic change concern how within-person change in a process variable relates to subsequent within-person change in an outcome variable. Such knowledge informs clinicians which variables could be affected to achieve client improvement. Further strengths include the careful and theoretically based selection of

studied variables, the large sample size and number of repeated observations, and the relatively representative sample composition.

Several limitations need to be mentioned. We used a statistical VAR model with certain assumptions (e.g., linear lag-1 dependencies), and it is uncertain to what extent this model captures causal interactions (Bringmann, 2021). Only four symptoms of GAD were measured, and despite the careful selection of variables, it is probable that many causally relevant variables were missing. The more numerous connections in the contemporaneous than in the temporal network suggest that many predictive relationships within days were not captured using the present day-to-day time scale. Future dynamic studies should therefore use a time scale of several measurements per day to further identify directed patterns of interaction between clinical processes and symptomatology.

In sum, the estimated temporal network revealed that 'uncontrollability of worry', 'generalized worry', 'fear of being infected', 'fear of significant others being infected', and 'threat monitoring' were the most predictive of other components. The finding of independence of these anxiety network components from situational and state variables may suggest that the findings apply for non-pandemic as well as pandemic situations, but this should be investigated in future studies. The pattern of temporal connections suggested that the processes of threat monitoring and generalized worry were largely responsible for the propagation of infection fears to a more severe anxiety state. Thus, targeting these processes could prevent infection fears from escalating to a full-blown GAD. Notably, the GAD symptoms generalized worry and anxiety were predicted both by processes proposed by the MCT model (uncontrollability of worry and threat monitoring) and by a process (being stressed by unforeseen events) proposed in the intolerance of uncertainty model. This suggests that there may be several routes to changing GAD symptoms. However, before more definite clinical recommendations are given, the present findings should be followed up with testing experimentally whether interventions on variables lead to changes consistent with the estimated network structures (Fried, Papanikolaou, et al., 2022).

CHAPTER 7

NETWORK ANALYSIS OF DEPRESSION SYMPTOMS DURING THE COVID-19 PANDEMIC

CHAPTER 7

Abstract

In order to understand the intricate patterns of interplay connected to the formation and maintenance of depressive symptomatology, repeated measures investigations focusing on within-person relationships between psychopathological mechanisms and depressive components are required. This large-scale preregistered intensive longitudinal study conducted 68,240 observations of 1,706 individuals in the general adult population across a 40-day period during the COVID-19 pandemic to identify the detrimental processes involved in depressive states. Daily responses were modeled using multi-level dynamic network analysis to investigate the temporal associations across days, in addition to contemporaneous relationships between depressive components within a daily window. Among the investigated psychopathological mechanisms, helplessness predicted the strongest across-day influence on depressive symptoms, while emotion regulation difficulties displayed more proximal interactions with symptomatology. Helplessness was further involved in the amplification of other theorized psychopathological mechanisms including rumination, the latter of which to a greater extent was susceptible toward being influenced rather than temporally influencing other components of depressive states. Distinctive symptoms of depression behaved differently, with depressed mood and anhedonia most prone to being impacted, while lethargy and worthlessness were more strongly associated with outgoing activity in the network. The main mechanism predicting the amplifications of detrimental symptomatology was helplessness. Lethargy and worthlessness revealed greater within-person carry-over effects across days, providing preliminary indications that these symptoms may be more strongly associated with pushing individuals toward prolonged depressive state experiences. The psychopathological processes of rumination, helplessness, and emotion regulation only exhibited interactions with the depressed mood and worthlessness component of depression, being unrelated to lethargy and anhedonia. The findings have implications for the impediment of depressive symptomatology during and beyond the pandemic period. They further outline the gaps in the literature concerning the identification of psychopathological processes intertwined with lethargy and anhedonia on the within-person level.

This chapter has been adapted from: Ebrahimi, O. V., Burger, J., Hoffart, A., & Johnson, S. U. (2021). Within-and across-day patterns of interplay between depressive symptoms and related psychopathological processes: a dynamic network approach during the COVID-19 pandemic. *BMC medicine*, *19*(1), 1-17.

7.1 Introduction

The global pandemic caused by the SARS-CoV-2 virus has been accompanied by substantial augmentations in psychiatric symptoms in the general population, with scholars denoting this homologous co-occurrence as a parallel pandemic of detrimental psychiatric symptomatology (Yao et al., 2020). Among the studied symptom domains, the cross-continental elevations in depressive symptoms have been deemed an area of concern warranting further investigation (Ebrahimi, Hoffart, et al., 2021; Ettman et al., 2020; Salari et al., 2020; Winkler et al., 2020). To date, the preponderance of the pandemic literature has concerted its efforts toward the identification of prevalence estimates and demographic risk factors accompanied by the alterations in depressive symptom levels (Salari et al., 2020; Wade et al., 2020; Xiong et al., 2020). Consequently, knowledge remains exiguous concerning the psychopathological mechanisms that are interconnected with psychiatric symptom expressions during the pandemic (Wade et al., 2020).

Psychopathological mechanisms refer to processes which contribute to the amplification and maintenance of psychiatric symptomatology. Within the processes encapsulated in this phenomenon, behavioral and cognitive-affective mechanisms connote a prime category of interest, as they are loanable to manipulation by a wide range of psychiatric treatment modalities aimed at alleviating depression. Notably, such mechanisms (e.g., rumination) entail processes that are tied to fluctuations in symptoms *within* individuals. By contrast, risk factors provide information about the likelihood of experiencing detrimental symptoms *compared to peers* in the population with other dispositional or circumstantial disparities. Accordingly, investigations of mechanistic processes versus risk factors of depression yield distinctive pieces of information not necessarily compatible with the other, with their separation requiring the deployment of the appropriate level of analysis to disaggregate between what is referred to as *within-person* and *between-person* relationships, respectively (Curran & Bauer, 2011). As reflected by recent research calls (Demakakos, 2021; Wade et al., 2020), however, much of the pandemic literature encompasses of study designs and analytical tools that are precluded from appropriate separation of these pivotal relationships.

Several scholars have denoted the substantive necessity of disentangling within- from between-person relationships (Bos et al., 2017; Curran & Bauer, 2011; Hamaker et al., 2015; Hoffart, 2014; Kievit et al., 2013), with an example from the field of medicine highlighting its importance. Although the risk of heart attack is lower among physically active people (i.e., a *between-person* relationship), the chances of an individual having a heart attack is higher while exercising (i.e., a *within-person* relationship). Consequently, the presence of these opposing effects with the same set of variables (termed Simpson's paradox, e.g., Kievit et al., 2013) accentuates the importance of their appropriate and distinctive investigation. From this perspective, knowledge concerning the formation of depressive symptoms and their patterns of interconnection with psychopathological mechanisms warrants investigations at the within-person level of analysis, presenting a key step toward the identification and impediment of the escalatory processes tied to the aforementioned increases in detrimental depressive symptomatology during the present pandemic. Mapping out such interrelations is further of utility beyond the pandemic period, as more knowledge is needed concerning the multitudinous processes involved in the maintenance of deleterious mental health states in non-clinical populations. As such, calls have been made for the adaptation of multi-level

dynamic network approaches using longitudinal designs and time series data (Ebrahimi, Hoffart, et al., 2021; Nature Medicine, 2020; Skjerdingstad et al., 2021; Wade et al., 2020; Wang et al., 2020), yielded with the aptitude of detecting the different components involved in the maintenance of depressive symptomatology while appropriately separating within- from between-person effects across time.

A suitable dynamic network approach incorporating these properties includes the use of the multi-level vector auto-regressive (VAR) model, further allowing investigations of relationships among variables occurring across specific time lags and within a given time window (Bringmann et al., 2013; Epskamp, Waldorp, et al., 2018; Fried et al., 2017). These patterns of interaction may further be interpreted through the lens of the network theory of mental disorders (Borsboom, 2017; Borsboom & Cramer, 2013), conceptualizing psychiatric symptoms and related components as networks of causally interacting entities. The time-lagged relationships in such dynamic network models are indicative of *Granger causal* relationships (Granger, 1969), denoting a variable's ability to predict another variable at the consecutive time point, yielding important information about which variable temporally precedes another in a system. Simultaneously, such network models provide information concerning interactions between variables occurring within a given time window, providing information about processes that may unfold at a faster rate than the studied temporal window of measurement (Epskamp, van Borkulo, et al., 2018). In summary, the adaptation of dynamic network models allows for investigations of within-person relationships between symptoms and mechanisms, while providing information about their temporal order and preliminary indications concerning the time windows which they interact on.

The present preregistered study uses multi-level VAR networks to investigate the day-to-day and within-day fluctuations of depressive symptoms during the COVID-19 pandemic, with the aim of identifying the mechanistic processes involved in the amplification and maintenance of deleterious depressive symptomatology in the general adult population. In adapting a multi-level approach, the study further disentangles within-person from between-person relationships to identify and separate between processes of change and risk factors, respectively. Such investigations represent tests of theorized connections between depressive symptomatology and its constituents, advancing the insight concerning the patterns of interplay present among symptoms and mechanistic and contextual variables in detrimental depressive states.

As detailed in the preregistered protocol of this study, a comprehensive range of psychopathological mechanisms and contextual variables were investigated, with the aim of advancing the insight concerning how these theorized variables interact with specific symptoms of depression. Several psychopathological theories predict rumination to be a key process involved in depressive dynamics. As proposed by metacognitive theory (Wells, 2009), rumination may arise as an attempt to understand the reasons of depressed mood, only to operate as a maintaining mechanism of depressive symptomatology with individuals remaining stuck in the depressed state through engagement in repetitive cognitive processes rather than functional problem solving. Among other psychopathological mechanisms, helplessness may play a particularly prominent role in maintaining depressive states during pandemic periods, with learned helplessness theory predicting depressive symptomatology to arise when individuals perceive to have limited influence over the circumstances they are exposed to (Miller & Seligman, 1975). Additionally, emotion regulation difficulties are theorized as

a maintaining mechanism in depressive states, with increased proneness of employing maladaptive emotion regulation strategies presenting greater difficulties of recovering from negative emotions, sustaining the depressed mood (Joormann & Gotlib, 2010; Solbakken et al., 2023). Finally, contextual variables previously tied to depressive states in pre-pandemic periods were investigated, including loneliness (Fried, Bockting, et al., 2015), physical activity (Camacho et al., 1991), social media use (Aalbers et al., 2019), interpersonal conflict (Keser et al., 2020), sleep quality (Çelik et al., 2019), relatedness needs (Vansteenkiste et al., 2006), and productivity (Heiligenstein et al., 1996). As the preponderance of these aforementioned variables has been subject to fluctuation during the present pandemic, an investigation of their relevance in the maintenance of depressive states is important. Examples include fluctuation in loneliness levels tied to social distancing protocols (Hoffart et al., 2020), changes in productivity related to transitions from work to home office, and sleep disturbances connected to perturbations in daily routine (Cheng et al., 2021). Finally, access to information (Ebrahimi, Hoffart, et al., 2021) and social contact (Benke et al., 2020) was investigated, both of which have been related to depressive symptoms in pandemic settings.

7.2 Methods

The preregistered protocol of this study can be found at the online repository of the Center for Open Science (https://osf.io/trf2y). All elements of the submitted study adhere to its preregistered protocol. Ethical approval for this study was granted by the Regional Committee for Medical and Health Research Ethics (reference: 125510).

7.2.1 Study design and time period description

The present study comprises an intensive longitudinal design conducting daily measures of depressive symptomatology and related mechanistic and contextual constituents for 40 consecutive days during the COVID-19 pandemic. This data collection method is referred to as a diary study and falls under the area of ambulatory assessment techniques (Mestdagh & Dejonckheere, 2021), which also encompass the experience sampling method (ESM) and ecological momentary assessment (EMA). In the clinical empirical literature, these terms are often used interchangeably and commonly referred to as the sampling of intensive longitudinal data in the participant's real life using portable devices.

The measurement period (i.e., February 17 to March 28, 2021) was characterized by several periodic-specific events, encompassing (1) three longer and continuous periods of national holidays (i.e., days 6 to 12, days 13 to 19, and day 38 onward) and (2) a consecutive and uninterrupted period with implemented viral mitigation protocols where no modifications in national protocols occurred (i.e., days 20 to 37). This uninterrupted viral mitigation period was characterized by a stable set of protocols, such as quarantine upon contact with infected individuals, isolation upon infection, closure of schools and universities, restriction on social gatherings, public activities and events, and visitation restrictions. Several of these implemented protocols (e.g., social gatherings, domestic travel, and visitation restrictions) were slightly lightened during the three holiday intervals encompassed in the study period (i.e., the two winter and the Easter holidays).

7.2.2 Participants and procedure

This study is part of the Norwegian COVID-19, Mental Health and Adherence Project, a large ongoing longitudinal investigation of psychiatric symptomatology in the general adult population. Eligible participants included all adults (i.e., age ≥ 18 years) residing in Norway. Prior to the aforementioned daily measurements conducted for the present study, the participants provided responses at four measurement waves since the onset of the pandemic. Upon initial recruitment to the project (i.e., the first wave of data collection, March 2020), the participants responded through an online survey disseminated to a random selection of Norwegian adults through a Facebook business algorithm, in addition to systematic dissemination of the survey via national, regional, and local information platforms (i.e., television, radio, and newspapers). This procedure is elaborated in detail elsewhere (Ebrahimi, Hoffart, et al., 2021). The same participants were recontacted at each wave of measurement. At the fourth wave of data collection (i.e., January 2021), the participants were queried concerning their interest in participating in an upcoming 40-day study about mental health (i.e., the present study). A total of 2,383 participants expressed interest to partake in the study, of which 1706 individuals formally enrolled in the study. Daily measures were conducted across a 40-day period, encompassing of a 24-h sampling frequency with the participants receiving the set of time-variant items each evening at 6:30 PM. The sampling frequency was held constant throughout the measurement period, and daily measures were conducted to investigate temporal effects (i.e., relationships across days) and contemporaneous effects within the same time window (i.e., relationships within a day) (Epskamp, van Borkulo, et al., 2018). The daily sampling frequency was deemed as appropriate given its direct relation to the assessment of depressive symptom endorsement in the Diagnostic and Statistical Manual of Mental Disorders (DSM-V), querying about the presence of symptomatology during and across days (American Psychiatric Association, 2013).

7.2.3 Measurement

7.2.3.1 Time-invariant variables

The participants reported their age, sex, education, civil status, preexisting mental health status, and region of residence.

7.2.3.2 Time-variant variables: item selection procedure and response scale

The item selection procedure in the present study was designed to accommodate for critical topics in the dynamic network analytic literature. First, all items were selected with the aim of avoiding topological overlap and thus possible inflation in centrality estimates (Fried & Cramer, 2017). Second, this theoretically grounded selection was proceeded by a data-driven approach, affirming the correlation matrix to be positive definite and that the included items were not linear combinations of one another. Subsequently, the goldbricker algorithm (Jones et al., 2021) was used to search for pairs of highly intercorrelated items, in addition to items displaying similar behavioral patterns with the other items in the network. Dependent correlations were investigated using the Hittner method (Hittner et al., 2003). The data analytical approach was congruous with the theoretical selection, identifying no redundant items.

Another topic that has received notable attention in the (dynamic) network literature includes utilization of validated items, which were predominantly adapted in this study (cf. preregistration protocol). Finally, these aforementioned topics were coupled with selections of theorized psychopathological mechanisms and contextual variables of potential relevance to depressive symptom dynamics. Overall, the item selection process followed a consensus procedure consisting of six meetings between the authors, yielding the following preregistered study protocol (https://osf.io/rekzm) containing the full details of each investigated variable and the theoretical rationale underlying item selection.

The full list of items measuring the depressive symptoms and related mechanistic and contextual constituents is provided in Table 7.1. All items were adapted to capture daily patterns of interplay. Following Fried and colleagues (2022), the items were measured on a 5-point response scale, with all variables and their full measurement details presented in the table note of Table 7.1.

Table 7.1. Items measured daily over the course of 40 consecutive days.

No.	A bbreviation	Item
1	Depressed mood	"Today, I felt down, depressed or hopeless."
2	Anhedonia	"Today, I had little interest or pleasure in doing things."
3	Lethargy (energyless)	"Today, I felt tired or that I had little energy."
4	Worthlessness	"Today, I felt bad about myself or felt like a failure."
5	Rumination	"Today, I thought negatively about things that have happened in the past."
6	Emotion regulation deficits	"Today, it has been difficult to cope with my emotions."
7	Helplessness	"Today, I felt helpless with regard to my problems."
8	Loneliness	"Today, I felt lonely."
9	Sleep satisfaction	"Today, I was satisfied with my sleep."
10	Productivity	"Today, I felt productive or useful."
11	Relatedness	"Today, I felt close to other people."
12	Sufficient information	"Today, I received enough information on how to deal with the pandemic and its associated protocols."
13	Interpersonal conflict	"Today, I argued or had negative discussions with someone."
14	In-person social contact	"Today, I spent... minutes/hours on physical social gatherings (i.e., meeting others face-to-face, offline)."
15	Digital social contact	"Today, I spent... minutes/hours on digital social gatherings."
16	Social media	"Today, I spent... minutes/hours scrolling social media just to make the time pass."
17	Physical activity	"Today, I spent... minutes/hours physically exercising to the extent that it lead to increased pulse or at least minimal sweating."

Note. All Items were measured on a five-point scale (1–5). *Items 1–13*: 1 (Not at all), 2 (Slightly), 3 (Moderately), 4 (Very), 5 (Extremely). *Items 14–16*: 1 (0 min), 2 (1–15 min), 3 (15–60 min), 4 (1–2 h), 5 (Over 2 h). Item 17: 1 (0 min), 2 (10–15 min), 3 (15–30 min), 4 (30–60 mi), 5 (over 1 h).

7.2.4 Statistical analyses

7.2.4.1 Time series analyses and data pre-processing for network models

All statistical analyses were performed using R version 4.1.0 (R core team, 2013). The R code and the correlation matrices necessary to regenerate the estimated models may be found here https://osf.io/trf2y/. Period-specific patterns across the different periods of the study (i.e., holiday periods and uninterrupted period of viral mitigation) were investigated using multilevel models, with a two-sided alpha level of .001 set as the inference criteria. Along with the time series visualizations, these auxiliary analyses provide descriptions of the investigated variables across the 40-day measurement period to be briefly presented in the 'Results' section.

Prior to the estimation of the main analyses of the study (i.e., estimation of networks), pre-processing steps common for dynamic network models were performed. First, these analyses require a minimum number of observations per person. Because the procedure is based on within-person centering using sample means per person, it is generally not recommended to include individuals with less than 20 measurements (Epskamp, Waldorp, et al., 2018; Jordan et al., 2020). To find an optimal balance between including participants with minimal missingness and retaining as many participants as possible, the number of completed diaries was visualized as a function of the cumulative number of participants (see Figure S7.1). The plot indicated that any more lenient cutoff for completed diaries than about 30 would not lead to substantially larger numbers of included participants. Accordingly, participants who completed at least 30 out of 40 diaries were selected. This resulted in including 1,368 out of 1,706 participants.

Second, the presence of trends in the data may lead to lower specificity or sensitivity in the resulting networks (Epskamp, van Borkulo, et al., 2018). Accordingly, a linear trend analysis was performed for each variable using two components; a cumulative linear trend over the assessment period, and a weekday versus weekend trend. Such trends can be identified by performing a regression of the item scores on the assessment time (linear trend), as well as on a dummy variable coding week-days versus weekend-days (weekend trend). For a detailed, reproducible work flow of the trend removal, the reader is directed to the R-code found in the 'Code availability statement' section.[32] In the subsequent analyses, these trends were removed from each variable by subtracting the linear trends and weekend effects from each observation. Note that the time series visualized in Figure 7.1 portray the data prior to the detrending procedure.

---

32 The materials necessary to regenerate the estimated models of the present study may be found at the online repository of the Center for Open Science (https://osf.io/trf2y). As our received ethical approval from the Norwegian Centre for Research Data (NSD) precludes submission of raw data to public repositories, the matrices underlying the model estimation are provided. Access to the data can be granted from the principal investigators Omid V. Ebrahimi and Sverre Urnes Johnson following ethical approval of a suggested project plan for the use of data granted by NSD and REK. Code availability: All code for the present study is uploaded at the online repository of the Center for Open Science (https://osf.io/m2zhu/). We also provide a step-by-step guide for conducting radar plot visualizations of centrality metrics, readily available in the code.

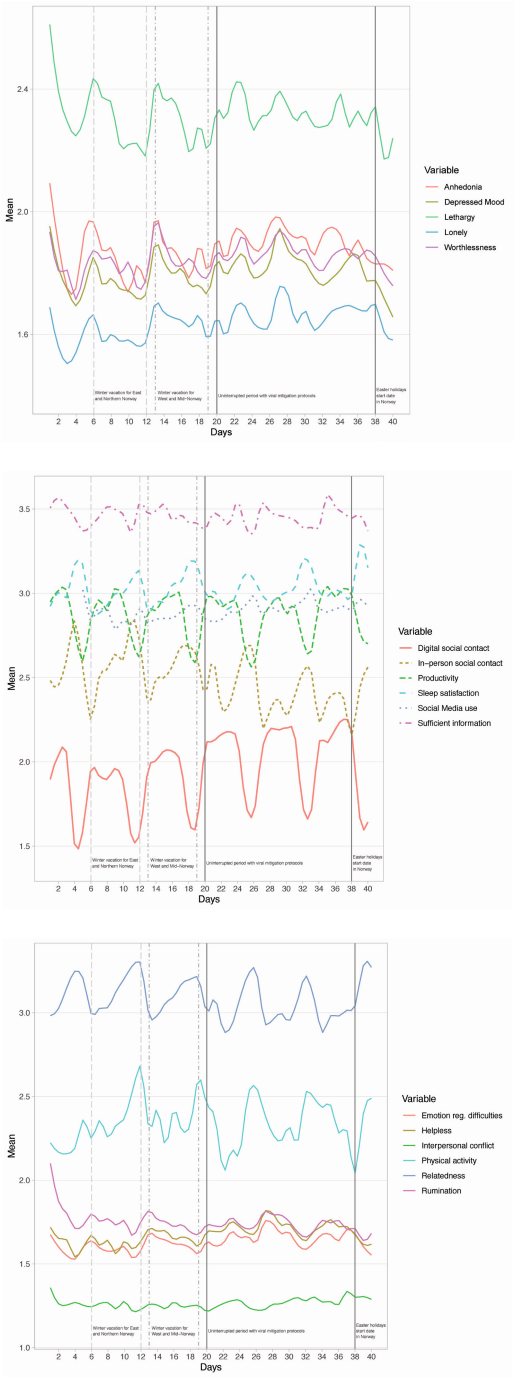


Figure 7.1 Nomothetic time series visualizations of all investigated variables through the measurement period, further depicting period-specific patterns across the 40-day study period.

7.2.4.2 Main analyses

We used the multi-level vector auto-regressive model implemented in the *mlVAR* package in R (Epskamp et al., 2019) to estimate the network models from the data. The algorithm implemented in *mlVAR* is based on a two-step procedure. First, (*within-person*) temporal and *between-subjects* effects are computed based on a node-wise multi-level regression, and second, (*within-person*) contemporaneous effects are obtained by performing a subsequent node-wise multi-level regression from the residuals in step 1. In line with the recommendations for networks with more than six nodes (Epskamp et al., 2019; Epskamp, Waldorp, et al., 2018), *orthogonal* estimation was chosen for both the temporal and contemporaneous networks.

This results in three types of networks, visualized in Figure 7.2. (1) A *fixed-effect temporal network* (top panel of Figure 7.2), in which average within-person effects indicate predictions of different nodes at the consecutive time point (i.e., lag-1), capturing the potential across-day temporal interactions between depressive symptomatology and related components. The temporal network provides directed statistical relationships (i.e., one-headed arrows) that are interpreted as Granger-causal (Granger, 1969), representing whether a node at time t predicts another at the subsequent time point (i.e., $t + 1$), while controlling for all other nodes in the network. (2) A *fixed-effect contemporaneous network* (middle panel of Figure 7.2), indicating average within-person effects between variables that are not captured in the temporal network, which estimates the unique interactions between all nodes within the same time window. In the dynamic network literature, these effects have been interpreted as dynamics that are potentially faster than those captured in the lag-1 temporal effects (Epskamp, van Borkulo, et al., 2018), indicative of interactions between nodes within the same day in the present study. (3) The *between-subjects network* (bottom panel of Figure 7.2) indicates relationships between variables based on the person-wise means of each variable. The between-subject network concerns average *between-person effects*, revealing how higher average levels on a variable *compared to peers* (i.e., compared to other subjects) is related to the mean levels in another variable compared to others in the population (e.g., people who on average are more physically active compared to their peers also likely have lower average heart rate than their peers). The temporal and contemporaneous networks concern average *within-person effects* across and within measured time windows respectively, both revealing how people displaying higher scores on a variable *compared to their own average* may display average within-person level changes on another variable (e.g., when individuals exert more physical activity than their own average, they also experience higher heart rate than their own average). The within-person effects provide insight into the patterns of interplay between symptoms and mechanisms of change in a depressive system, while the between-subject effects provide information concerning risk factors associated with depressive symptoms across subjects.

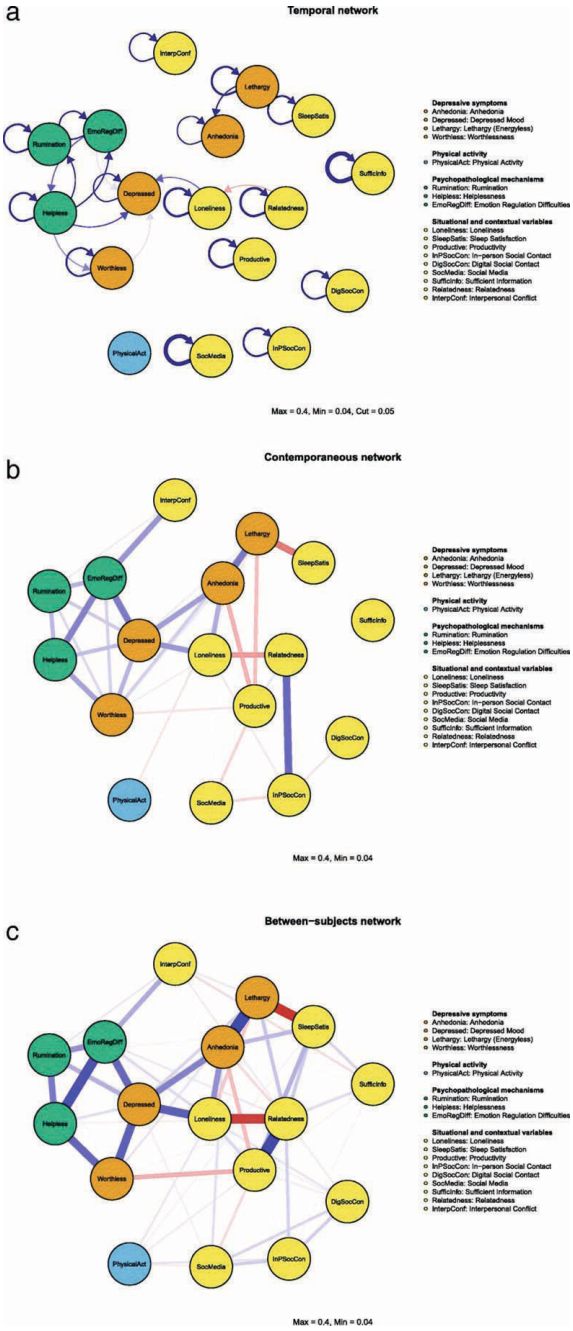


Figure 7.2. Temporal (top), contemporaneous (middle), and between-subject (bottom) networks derived from the multi-level vector auto-regressive (VAR) model showing the connection between nodes while controlling for all other nodes in the network. The strongest edges include the temporal network coefficient = 0.27 ('SufficInfo' → 'SufficInfo'), the contemporaneous network coefficient = 0.26 ('InpSocCon' – 'Relatedness'), and the between-subject network coefficient = 0.45 ('Anhedonia' – 'Lethargy').

Each network consists of sets of nodes (i.e., variables) listed in Table 7.1 and sets of edges describing the relationships between nodes. Blue and red edges portray positive and negative relationships, respectively. Importantly, each network model estimates the unique relationships among nodes while controlling for all other variables in the network. The main focus of the present study includes the average *within-subject* relationships (i.e., temporal and contemporaneous networks).

Centrality metrics (Opsahl et al., 2010) aim to quantify the role of individual nodes for the overall information flow in the networks. Strength centrality enhances the interpretation of network models through highlighting how strongly a node is directly connected to other nodes in the network. As a directed graph, the temporal network model enables estimation of the outstrength and instrength centrality, quantifying the sum of all outgoing and incoming absolute edge weights (i.e., excluding the autoregressive effect) from and to a node, respectively. Instrength thus reveals which nodes are more likely of being influenced by fluctuations in other nodes in the network at the previous day, while outstrength centrality quantifies the magnitude of a node in influencing other nodes in the network at the consecutive day. The undirected between-subject and contemporaneous networks provide estimations of strength centrality, computing the sum of all absolute edge weights connected to a node to quantify the overall weighted connectivity of the node in the network. All aforementioned strength centrality metrics reflect the average conditional associations between a node and the other nodes in a network. In the present study, we introduce a novel approach in *visualizing* centrality metrics using radar charts in order to enhance visual comparisons of centrality indices in a given network (i.e., outstrength versus instrength centrality in the temporal network) and across networks containing the same nodes. In line with the recommended reporting standards for network studies presented in chapter 5, we use raw centrality scores as opposed to standardized estimates, as the latter may inflate dissimilarity between centrality indices.

7.2.4.3 Sensitivity to demographic composition

The proportion of all demographic characteristics was investigated and compared to their known rate in the population. All characteristics not fully representative of the Norwegian adult population were adjusted in sensitivity analyses encompassing of a random selection of participants fully matching the population characteristics. The similarity and degree of replicability between the results from the main sample and the adjusted proportional subsample representative of the population were compared through correlating the respective matrices containing any estimated effects in the study, with its range reported at the beginning of the 'Results' section.

7.2.4.4 Robustness and replicability of networks

Additional analyses were performed to assess the robustness and replicability of the network models. These analyses were conducted across four subdivisions of the dataset. First, all participants were randomly separated into two groups prior to re-estimation of the network models and assessment of the replicability of the findings across the two subsamples. Second, two additional subdivisions of the dataset were created, separating the data into an early subsection consisting of all participants using the first half of the time series and a second subsection encompassing of data of all participants using the latter half of the time series. Thus, the network models were further re-estimated to assess the replicability of the findings across the time-specific subsamples.

In each of the four aforementioned subsamples, three main analyses were conducted to assess the robustness of the findings, with each subsample compared to its respective counter-subsample as detailed above. First, following previous research (Funkhouser et al., 2021), the global replicability and consistency of edges of each of three estimated networks (i.e., temporal, contemporaneous, and between-subjects network) was assessed through correlating the estimated edge weights in each subsample. Second, to assess the stability of centrality values, estimated centrality indices were compared through correlations across each respective pair of subsamples. Finally, the rate of consistency among the nodes with the highest centrality was assessed through comparing the total number of times the most central nodes identified by the main analyses were replicated across all four subsamples described above, used as a proxy to obtain estimations approximating the rank-order stability of the centrality indices.

In using such proposed assessments of robustness across subsamples, a previous study (Funkhouser et al., 2021) found moderate replicability across subsample networks through correlations of .61 when comparing edge weights, toward which the present findings will be benchmarked against. The range of correlations derived from these robustness analyses is to be presented in the 'Results' section labeled *network replicability*.

7.2.4.5 Network visualization

All networks have been visualized using the *qgraph* package in R (Epskamp et al., 2012). The maximum edge weight across the three networks was set to correspond to the largest edge weight across the networks (i.e., *partial r* ~ .4). Correspondingly, to filter out weaker from more notable effects, the minimum edge weight was set to one-tenth (i.e., .04) of the maximum value. Note that the set minimum merely hides edges in the network figures for visualization and interpretation-enhancing purposes and does not remove them from the model. As common across dynamic network studies (Fried, Papanikolaou, et al., 2022; Levinson, Vanzhula, et al., 2018), the temporal network generally exhibited smaller effects than the other two networks. Therefore, for visualization purposes, a cut value of .05 was set to more clearly separate the effects above and below this threshold. The arrangement of the nodes is based on the average layout of the three networks that have been established via the Fruchterman-Reingold algorithm (Fruchterman & Reingold, 1991). The matrices containing all edge weights and the raw networks displaying all edges (i.e., including the weaker effects) can be found at the online repository of the Center for Open Science (https://osf.io/trf2y).

7.3 Results

A total of 1,706 participants enrolled in the study. The age of the participants ranged from 18 to 86 years (M_{age}=37.30), with 1336 (78.54%) of the participants being female, 962 (56.89%) having a university degree, and 830 (49.43%) being married or in a civil partnership. A total of 1,368 of the 1,706 (80.19%) participants provided sufficient data to be included the study, with no pattern of difference identified between initiators and those with sufficient data. The percentage of individuals with preexisting mental health conditions in this sample was 16.62%, representative of the known rate of psychological disorders in the adult population of Norway, which is between 16.66 and 25.00% (Tesli et al., 2016). The sample was further geographically representative of Norway,

with the quota of participants sampled from each region being proportional to region size. With the exception of sex and education (i.e., oversampling females and those with a university degree), the preponderance of demographic characteristics were representative of the Norwegian adult population. To fully match all demographic characteristics (i.e., including sex and education) to the known proportions in the population, sensitivity analyses were conducted on a randomly drawn set of 598 individuals fully matching the population parameters. These sensitivity analyses replicated the results from the main sample across all analyses below, with the correlation between the matrices containing the results of the representative sample and main sample ranging from .96 to .99.

7.3.1 Time series analyses and time-specific patterns across the study period

Figure 7.1 provides a visualization of the time-specific patterns of depressive symptomatology and related constituents across the 40-day study period. Overall, mental health-promoting associations were identified during the holiday periods where pandemic protocols were lightened (i.e., days 6 – 12, 13 – 19, and 38 and onward), while detrimental associations were found during the period encompassing of uninterrupted viral mitigation protocols (i.e., days 20 – 37). Specifically, all unfavorable variables (e.g., 'loneliness', 'depressed mood', 'interpersonal conflict', 'helplessness') revealed linear decreases during holiday periods (*ps* <.001) while increasing during the continuous viral mitigation period (*ps* <.001). All favorable variables (e.g., 'relatedness') revealed linear increases during holiday periods (*ps* <.001), while decreasing during the uninterrupted viral mitigation period. The only notable exceptions from these patterns included (1) 'productivity' (i.e., increasing during uninterrupted viral mitigation period, decreasing during holidays, *ps* <.001) and (2) 'lethargy', 'information access needs', 'sleep satisfaction', and 'rumination' which did not reveal any significant fluctuations during the continuous viral mitigation period (*ps* >.05).

In-person (i.e., offline face-to-face) and digital social contact demonstrated opposite patterns, with 'in-person social contact' decreasing during the continuous viral mitigation period and increasing during holidays, while 'digital social contact' decreased during holidays and increased during the continuous viral mitigation period (*ps* <.001).

7.3.2 Patterns of interplay between depressive symptoms and related components

The *within-person* patterns of interplay obtained in the temporal and contemporaneous network models (Figure 7.2) provide insight concerning the potential processes involved in the maintenance and amplification of depressive symptomatology.

Figure 7.2 (top panel) displays the temporal network revealing the average *within-person* connections between nodes *from one day to the next*, with the radar plots in Figure 7.3 depicting each variables' outstrength and instrength centrality. The radar plots depicting outstrength and instrength displayed distinctive patterns, indicating differences in the extent to which nodes were associated with having outward influencing roles versus susceptibility of being influenced on an across-day basis. 'Loneliness', 'helplessness', and 'in-person social contact' had the greatest outstrength centrality. 'Depressed mood', 'anhedonia', and 'emotion regulation difficulties' had the greatest instrength centrality. Concerning node connections, specific across-day patterns unfolded between 'lethargy' and 'anhedonia', with greater within-person levels of 'lethargy' temporally predicting increases in within-person levels of 'anhedonia' at the consecutive day, and 'anhedonia' further reinforcing itself

across days in a vicious self-loop. This pattern of interwovenness also involved an autoregressive carry-over effect in 'lethargy', in which low energy levels carried over across days. The across-day interplay among depressive symptomatology was coupled and separated, with 'lethargy' and 'anhedonia' representing one pair, while 'depressed mood' and 'worthlessness' represented the other. 'Helplessness' was among the nodes revealing the highest outstrength centrality across days, with higher within-person levels of 'helplessness' being involved in the amplification of other detrimental mechanistic processes (i.e., increases in 'rumination' and 'emotion regulation difficulties') in addition to key symptoms of depression (i.e., increases in 'depressed mood' and 'worthlessness'), all further involved in detrimental self-loops across days. A vicious cycle was identified between 'helplessness' and 'emotional regulation difficulties', with higher within-person levels of each predicting greater increases in the other at the consecutive day. Examples of across-day patterns with smaller magnitude included the directed effects from 'relatedness' to 'loneliness' (i.e., higher within-person levels of 'relatedness' at the previous day predicted less 'loneliness' at the consecutive day), greater 'helplessness' predicting more 'worthlessness' the next day, and more 'emotion regulation difficulties' and 'loneliness' predicting higher 'depressed mood' at the following day. Additionally, although having smaller magnitude in its outgoing effects, 'in-person social contact' demonstrated widely distributed across-day influence on the other nodes in the network, as reflected by its position among the nodes with greatest outstrength (Figure 7.3).

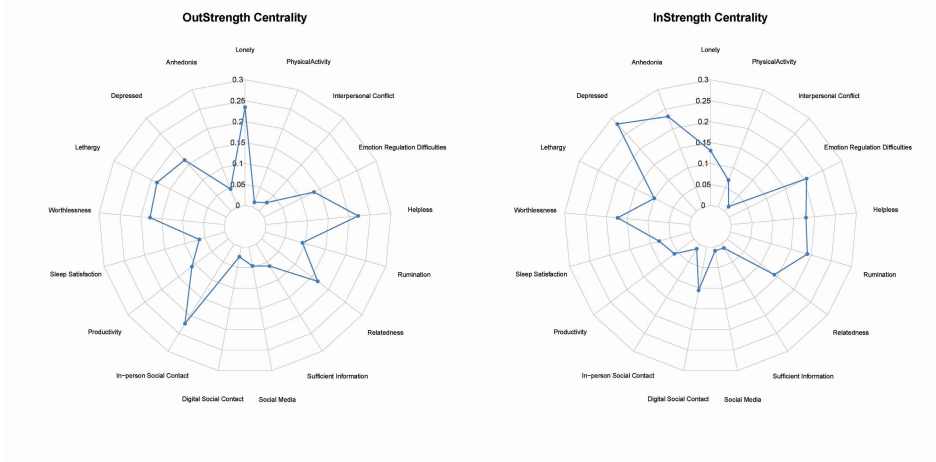


Figure 7.3. Radar chart depicting the OutStrength (i.e., sum of all outgoing absolute edge weights from a node) and InStrength centrality (i.e., sum of all incoming absolute edge weights to a node) of the variables in the temporal network model. The across-day directed involvement of a node is revealed through the extent of which a node influences other nodes (i.e., OutStrength) at the consecutive day or is influenced by other nodes in the network at the previous day (i.e., InStrength).

Inspecting the contemporaneous network (Figure 7.2, middle panel) provides indications of average *within-person* relationships among the investigated nodes occurring *within the same window of measurement*, which in the present study reflects a within-day time window. All abovementioned relationships between depressive symptoms and related constituents were present within the same window of measurement. In contrast to the across-day patterns including separate clusters of in-

teraction among depressive symptoms, all depressive symptoms were related with one another in the contemporaneous network. Notable unique patterns of interconnection were found within a daily window of measurement, with within-person 'sleep satisfaction' inversely related to 'lethargy', within-person increases in 'loneliness' associated with higher within-person levels of 'anhedonia' and lower 'relatedness', and greater 'emotion regulation difficulties' being associated with more 'interpersonal conflict' and 'worthlessness'. Additionally, 'productivity' portrayed negative within-day relationships with both 'anhedonia' and 'lethargy'. Importantly, the relationship between 'rumination' and key depressive symptoms (i.e., 'worthlessness' and 'depressed mood') predominantly occurred within the same window of measurement (i.e., within a day), revealing weak effects across days. Among the contextual variables prominent during the pandemic, 'in-person social contact' and 'relatedness' were further strongly interwoven in the same time window. The most central (i.e., strength centrality; Figure 7.4, left panel) nodes in contemporaneous network were 'depressed mood', 'anhedonia', and 'emotional regulation difficulties', outlining the nodes with the strongest overall connectivity within a day among the nodes in the network.

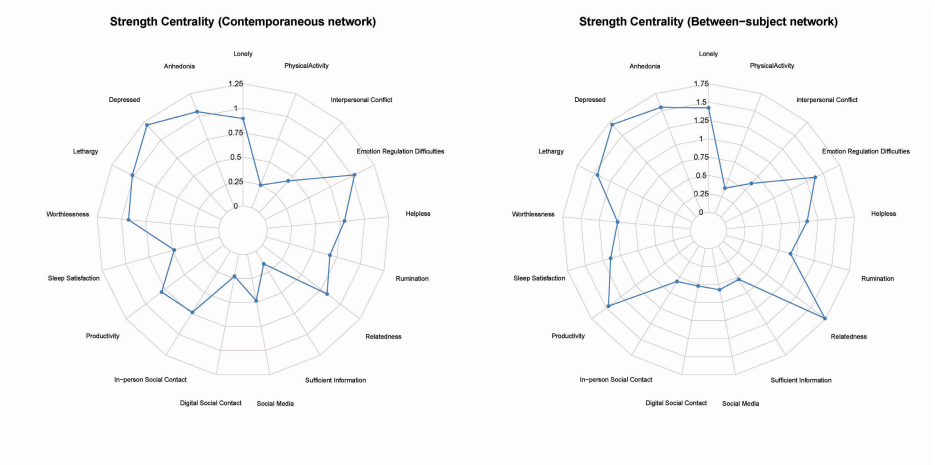


Figure 7.4. Radar chart revealing the strength centrality (i.e., the sum of all absolute edge weights connected to a node), quantifying the overall weighted connectivity of the node in the contemporaneous (left) and between-subject (right) network models.

7.3.3 Risk factors associated with depressive symptoms

The between-subjects network (Figure 7.2, bottom panel) is suitable in the identification of risk factors across subjects in a population. Several distinctive associations were derived from this network, predominantly involving the contextual variables of the study. Particularly, the relationship between 'information access needs' and 'sleep quality' was highlighted, revealing that people who on average feel well-informed about the pandemic also report greater sleep satisfaction compared to other adults in the population. Higher 'relatedness' was associated with greater 'productivity' across subjects. Similarly, there was a negative association between 'productivity' levels and perceptions of 'worthlessness', and a positive association between 'productivity' levels and 'sleep satisfaction'.

Overall, the nodes with the highest strength centrality (Figure 7.4, right panel) in the between-subject networks were 'relatedness', 'depressed mood', and 'anhedonia'.

7.3.4 Network replicability

The estimated network models and their corresponding computed parameters were yielded as robust, replicating the main findings. Specifically, the correlation between edge weights comparing the two random subsamples of participants was $r = .93$ for the temporal network, $r = .99$ for the contemporaneous network, and $r = .97$ for the between-subjects network. Correspondingly, the correlation between edge weights comparing the first half of the time series compared with the latter half was $r = .92$ for the temporal network, $r = .99$ for the contemporaneous network, and $r = .99$ for the between-subjects network. Centrality estimates were further robust across both aforementioned pairs of subsamples, with correlations ranging from $r = .89 - .96$ (i.e., instrength) to $r = .80 - .88$ (i.e., outstrength) for the temporal network, stable at $r = .99$ for the contemporaneous network, and ranging from $r = .88 - .97$ for the between-subject network.

Finally, the edges revealing the highest centrality were consistent across all subsamples, with 96.67% of the edges with the highest centrality re-obtained in the subsample analyses across all networks.

7.4 Discussion

As discussed by multiple scholars (Curran & Bauer, 2011; Hamaker et al., 2015; Hoffart, 2014; Hoffman & Stawski, 2009; Kievit et al., 2013), a study of within-person relationships is required to understand the mechanisms of change in human behavior and psychopathological research. The disentanglement of such within- and between-person relationships are imperative, as conclusions from one level do not necessarily generalize to the other, where in extreme cases, these relationships can convey opposite patterns (Curran & Bauer, 2011; Hamaker et al., 2015; Hoffart, 2014; Kievit et al., 2013). Consequently, understanding the maintaining components involved in depressive states necessitates the study of within-person relationships.

7.4.1 Maintaining mechanisms of depressive symptomatology

The main purpose of the present study was to examine the *within-person* relationships present in the temporal and contemporaneous models of depressive symptoms and its constituents. As these networks model average within-person connections between nodes, they provide insight into potential mechanisms of change involved in the amplification and impediment of depressive symptomatology, providing directions toward further study and identification of targets for interventions aimed at alleviating these detrimental mental health problems.

Although all depressive symptoms were well-connected on a between-subject level and further interacted within the same window of measurement, the findings of present study indicate that interactions between depressive symptoms to a greater extent are separated and uniquely coupled across days. Specific across-day connections were identified between 'anhedonia' and the somatic symptom 'lethargy', while two cognitive-affective symptoms, perceptions of 'worthlessness' and 'depressed mood', were more strongly interconnected on an across-day basis. Additionally, the

relationship between these symptoms were directed, revealing the predominant temporal influence of lethargy on anhedonia, and worthlessness on depressed mood. These findings have implications for efforts aimed at impeding escalations of depressive states, suggesting that lethargy and worthlessness have a greater likelihood of contributing as catalysts in the escalation of deleterious depressive states from one day to the next. As a key feature putting individuals at risk of developing depressive syndrome involves the prolonged constellation and experience of multiple symptoms (American Psychiatric Association, 2013), insight into the specific symptoms that more likely yield carry-over effects across time is of importance from an epidemiological and clinical perspective in more successfully preventing the development of a depressive state. The present study identifies that the two most prominent depressive symptoms that may be involved in such detrimental carry-over effects in the non-clinical population are worthlessness and lethargy. This finding is consistent with cross-sectional network studies identifying worthlessness and lethargy as central nodes in depressive states (Skjerdingstad et al., 2021; Wang et al., 2020), with the present study advancing insight concerning the directed temporal involvement and coupled interaction between these symptoms.

This investigation further extended the applications of network theory through the introduction of relevant psychopathological mechanisms and contextual factors in the networks, yielding novel insights concerning the specific patterns that these processes exhibit in their interactions with depressive symptomatology. 'Loneliness', 'helplessness', and 'in-person social contact' had the greatest outward temporal influence (i.e., outstrength centrality) on the other variables in the network on an across-day basis. Studies during the present pandemic have found undirected associations between loneliness and depressive symptomatology in the general population (Hoffart et al., 2022; Palgi et al., 2020). The present longitudinal study advances the literature by identifying the direction of this association, further identifying that loneliness interacts with depression through its directed association with the depressed mood component of depression, carrying over across days.

The main psychopathological mechanism temporally associated with the maintenance and amplification of depressive dynamics on an across-day basis was 'helplessness'. Accordingly, when an individual reported being more helpless than their own average at a given day, they reported within-person increases in 'depressed mood', 'rumination', and 'worthlessness' at the consecutive day. This finding provides support for helplessness as an important mechanistic variable in the maintenance and change of depressive symptoms in the general population. This is consistent with the learned helplessness theory of depression (Miller & Seligman, 1975), postulating that when an individual comes to believe that their efforts to modify their circumstances are ineffective, developed perceptions of helplessness may incite depressive symptomatology. The finding is further consonant with a central meta-theory of psychopathology proposed by Jerome Frank, suggesting that demoralization (i.e., experienced helplessness or inability to cope) is a key aggravator of psychiatric symptomatology (Frank, 1974). As perceptions of helplessness are theorized by several scholars to be the main reason for individuals seeking psychiatric treatment (Clarke & Kissane, 2002; Frank, 1974), directing efforts toward reducing helplessness may be warranted. The present study provides preliminary indications that such efforts may have the ability to impede deleterious depressive states, although such assertions warrant further investigation using controlled designs.

Aside from being uniquely associated with within-person increases in key depressive symptoms and rumination at the next day, 'helplessness' was further engaged in a vicious cycle with 'emo-

tional regulation difficulties' across days, with emotion regulation problems also associated with heightening of 'depressed mood' from one day to the next within individuals. Combined with the finding that 'emotion regulation difficulties' were the most central psychopathological process in the contemporaneous network, revealing strong interactions with depressive symptoms within a day, this finding suggests it may be important to devote simultaneous attention toward the detrimental role that emotional regulation difficulties may play in depressive mental health states. Notably, this study provides indications that the interaction between depressive symptoms and emotional regulation difficulties may predominantly operate on a faster time scale than helplessness with depressive symptoms. This finding is meaningful, given that emotional regulation problems likely are more situationally contingent and probable of occurring within a more encapsulated time period. Consequently, these findings distinguish between the proximal role that emotion regulation difficulties play in its interaction with depressive symptoms, while identifying helplessness as having a more prominent role in terms of prolonged depressive symptom experience. More granular approaches are called for in future studies to refine the understanding of the possible directed role that emotion regulation difficulties may play within a day.

Among the aforementioned psychopathological mechanisms, 'rumination' was peripheral and did not have any notable interaction with depressive symptomatology on an across-day basis. This finding is consistent with a previous study (Hoorelbeke et al., 2019) identifying rumination to be on the receiving end of predictive temporal relationships in a network of mechanistic variables, in addition to another study not finding any temporal relationship between rumination and depressive symptoms (Lutz et al., 2018). In the present study, the only notable connection with 'rumination' included a directed effect from 'helplessness' predicting 'rumination' at the consecutive day. This finding suggests that helplessness may play a more prominent role in the maintenance and across-day constellation of depressive symptomatology in the non-clinical population, consistent with the goal progress theory of rumination proposing rumination to be a response to failure in achieving a certain task rather than an outgoing mechanistic process (Martin & Tesser, 1989). Consistent with existing studies (Skjerdingstad et al., 2021), 'rumination' revealed undirected associations to some symptoms of depression (e.g., weaker associations with 'worthlessness') on both a between-subject level and within a day. However, the present findings in combination with findings from directed network studies investigating within-day relationships involving depression and rumination (Hoorelbeke et al., 2019; Lutz et al., 2018) provide indications that these associations may to a greater extent be indicative of rumination being an influenced node rather than the influencing node, with implications for interventive efforts aimed at alleviation of depressive symptoms. This finding is further partially consistent with metacognitive perspectives on depression (Wells, 2009), postulating rumination to be a process *ensuing* depressive symptoms as a reactional attempt to understand the reason for their presence and in attempts of identifying solutions to the problem. However, the present study does find indications of rumination subsequently influencing depressive symptoms in turn, which is also postulated by the theory. Still, given the multimodal complexity of rumination (Bernstein et al., 2019; Collins et al., 2021), the literature will benefit from further temporal examinations of depressive symptoms simultaneously investigating rumination along with other psychopathological mechanisms of relevance, to better understand its specific as well as comparative interaction with depressive components.

The findings of the present study further shed some light on the interactions between depressive symptomatology and mechanistic processes that operate on a faster time scale than an across-day basis. In the present study, this reflects the identified interactions in the contemporaneous network, which cautiously provide indications of associations among nodes that may occur within person during a given day. Meaningful connections emerged between 'lethargy' within individuals in its association with reduced 'sleep satisfaction' within the same time window, while being more 'productive' than usual was associated with lower 'anhedonia' and 'lethargy'. 'Loneliness' was a central node with important connections to depressive symptoms and contextual variables across all three networks. On a within-person level, 'loneliness' displayed its largest connectivity within a day, with the findings indicating that while individuals felt greater loneliness than their own average, this was associated with greater within-person intensity of 'depressed mood' and 'anhedonia'. Consistent with a study by Fried and colleagues (2022) on the student population, the present study found higher within-person levels of 'loneliness' to be associated with reduced 'relatedness' and 'in-person contact'. The present study supports and adds to these findings by extending the time period of investigation to later stages of the pandemic and a broader demographic composition of participants, in addition to identifying detrimental associations between loneliness and depressed mood.

Notably, on the within-person level, the three psychopathological processes (i.e., 'helplessness', 'rumination', and 'emotion regulation difficulties') only exhibited interactions with the 'depressed mood' and 'worthlessness' component of depression, being unrelated to 'lethargy' and 'anhedonia'. These findings highlight the connection between these aforementioned cognitive-affective mechanisms with particular depressive components, providing important insights on the patterns of interaction between depressive symptoms and mechanistic processes. Simultaneously, they also leave important gaps in the literature concerning the identification of pathological processes more closely intertwined with lethargy and anhedonia on the within-person level.

7.4.2 Risk factors associated with depressive symptoms across subjects

Across subjects, 'in-person social contact' was revealed as the type of social interaction with the strongest association with 'relatedness', with those who reported being more frequently engaged with such face-to-face contact compared to their peers also reporting greater relatedness. Moreover, individuals who on average felt more connected to their peers during the pandemic reported greater levels of productivity, further mirrored by within-person relationships to outline several beneficial associations of relatedness. However, although 'relatedness' was connected to 'anhedonia' on a between-subject level, this connection was not present in any of the within-person networks (i.e., temporal and contemporaneous network). This demonstrates the importance of separating between- and within-person effects (Curran & Bauer, 2011; Hamaker et al., 2015; Hoffart, 2014; Kievit et al., 2013), with this finding implying that it is unlikely that relatedness is directly associated with anhedonia. Rather, as also revealed by the within-person networks, 'relatedness' is more indirectly connected to depressive symptoms through its association with 'loneliness'.

Between-person associations were further identified between information access and sleep, with those who on average reported 'sufficient access to information' about the pandemic situation reporting greater 'sleep satisfaction' compared to their peers. Still, no within-person relationships emerged for this association. Moreover, no social contact component other than 'in-person social

contact' revealed notable beneficial associations across any of the investigated networks, with other social contact components additionally portraying detrimental associations to depressive states. Specifically, consistent with previous findings (Aalbers et al., 2019; Primack et al., 2021), individuals who compared to their peers who were more engaged in 'passive social media use' had a greater risk of being associated with higher levels of 'anhedonia', in addition to lower 'productivity'. Yet, again, however, no meaningful within-person detrimental association emerged between 'social media use' and 'anhedonia', suggesting the limited likelihood of this factor being associated with within-person fluctuations in depressive states when controlling for all other variables in the network. Additionally, no beneficial within-person associations were identified with 'digital social contact'. Taken together, these findings highlight solely in-person social contact as having a potentially important role on a within-person basis through this variable association with loneliness and relatedness. As loneliness is an important problem in itself (Hoffart et al., 2020) and further was found to be connected to depressed mood across days on the within-person level in this study, this finding implies that attempts to find an optimal balance between strength of viral mitigation protocols and appropriate levels of in-person social contact, the latter of which the present findings reveal to be hard to substitute by other social contact types, may be of utility in combating the concurrently ubiquitous presence of loneliness. Clever behavioral interventions at the population level, including the use of social bubbles, may serve as utile strategies that can simultaneously reap the psychological benefits of reduced loneliness while maintaining control over viral spread (Leng et al., 2020). As for depressive symptoms, however, the present study does not identify any direct within-person relationship between social contact and depressive symptomatology, suggesting that efforts toward alleviation of depressive symptoms may be more fruitful when aimed at other identified mechanistic and contextual variables.

7.4.3 Other notable findings

The social contact components were negatively associated in the contemporaneous network, reflecting that while an individual is engaged in a greater extent of 'in-person social contact' than their own average, they are less involved in 'digital social contact' within the same window of time. This stands in informative contrast with the positive associations between these components in the between-subject network, which highlights that people who on average are more engaged with in-person social contact compared to their peers likely also are people who to a greater extent are engaged in both social media use and digital social contact. In other words, social individuals are sociable, likely to report higher levels of engagement compared to their peers among a wide range of social activities (i.e., between-subject network), but being engaged with one social activity in a given time window reduces the opportunities of being engaged with another social activity within the same time window (i.e., contemporaneous network). This contrasting finding between the two networks highlights the importance and utility of disentangling between within-person and between-person relationships. This is further emphasized through the positive connection identified between 'emotion regulation difficulties' and 'worthlessness' on a within-person level, while this relation was absent across individuals. In other words, while individuals experienced greater emotional regulation difficulties than their own average, this was associated with increased feelings of worthlessness during that day (i.e., a within-person effect). However, individuals who have

greater emotion regulation difficulties compared to their peers are not likely to be individuals who feel worthlessness. Within-person and between-person relationships are not necessarily coherent, and the inappropriate generalizations from the between- to the within-level has been referred to as ecological fallacy (Curran & Bauer, 2011; Robinson, 1950). In its investigation of within-person relationships among multiple theorized detrimental processes, the present study fills the gaps (Wade et al., 2020; Wang et al., 2020) in progressing the understanding of psychopathological mechanisms connected to depressive symptomatology in the general population.

Moreover, 'physical activity' and 'digital social contact' were consistently among the least central and influential node across all networks, outlining their limited relevance and involvement in depressive states when controlling for all other nodes in the network during the present pandemic context. Specifically, as no particularly notable within-person relationship was present between these variables and depressive symptoms, the present findings suggest that future efforts toward identification of variables that may impede deleterious symptoms within subjects best are aimed at other central components of symptom maintenance, such as helplessness and emotion regulation skills building. The findings of the present study thus imply that promising interventive targets warranting investigation in future controlled studies may include testing whether and how techniques such as cognitive restructuring and behavioral activation may temporally interact and impact perceptions of helplessness and lethargy, respectively.

Finally, this study introduces the usage and utility of radar plots in visualizing key information about network centrality metrics, with the results of the temporal network model outlining the comparative extent of involvement of a given node as an outgoing node at an across-day basis versus as a node more strongly tied to being influenced from other nodes at the previous day. Both 'loneliness', 'helplessness', and 'lethargy' had greater strength as outgoing nodes in contrast to being influenced. As relationships in temporal networks are indicative of Granger causal effects, these findings preliminary indicate the greater likelihood that helplessness, loneliness, and lethargy may play in serving as engines in the network, to a greater extent being associated with activation of other nodes. However, as Granger causal effects do not necessarily equate true causal processes and only satisfy its temporal criterion, these findings warrant further investigation in future studies. Other drastic differences were found for in-person social contact in terms of its relative position as an influential node versus being influenced, a finding which is meaningful in the present pandemic setting.

Both 'anhedonia' and 'depressed mood' were more likely to be impacted by other nodes at the previous day than having across-day carry-over effects. Across three of four centrality estimations (i.e., with the exception of outstrength centrality), 'depressed mood' and 'anhedonia' were the most central nodes in the networks, which provides support for their position as the key identifiers of depression (American Psychiatric Association, 2013). Importantly, however, these findings illuminate their more limited outgoing involvement in depressive states, highlighting lethargy and worthlessness to have stronger outgoing impact on other symptoms.

7.4.4 Strengths and limitations

The present chapter consists of several limitations. First, the conclusions of this chapter have to be interpreted in light of the underlying assumptions of the statistical model. More specifically, we interpreted a lack of relationships in the temporal network as indicative of potentially faster interactions between depressive symptoms and related components (Epskamp, van Borkulo, et al., 2018). This interpretation assumes that meaningful interactions can in principle be captured using linear lag-1 models. An alternative explanation for the lack of detected temporal relationships is that these could be nonlinear or time-varying (Bringmann, 2021b; Haslbeck & Ryan, 2022; Hayes et al., 2007; Schiepek et al., 2017), which calls for further investigations using other modeling approaches. Furthermore, although the study investigated some of the most central theorized mechanisms in the psychopathological literature, the edge weights were generally smaller in the temporal network than the other networks, as commonly the case in multi-level network analytic studies (Fried, Papanikolaou, et al., 2022; Levinson, Vanzhula, et al., 2018). This further highlights the necessity of advancing current and building novel theories through formalization and incorporation of the time scales which phenomena may operate on (Fried, 2020b; Haslbeck & Ryan, 2022). Finally, the modeled relationships in the present chapter are on the average within-person level, calling for idiographic efforts (Fisher et al., 2018) in inspecting how closely such within-person aggregations correspond to the level of the individual.

This study consists of several strengths, including that it was pre-registered with a clear rationale preceding the selection of variables. Additional strengths include the use of validated measures, its focused time window of measurement corresponding to the DSM-V depressive symptom endorsement assessment, longitudinal design, broad demographic composition of participants, and conducted sensitivity analyses on a fully representative sample replicating the main findings. Moreover, the robustness and replicability of the network models and their corresponding estimated parameters were assessed across four additional subsamples, revealing high robustness of the results. Importantly, the investigation of psychopathological mechanisms in a non-clinical population provides insight into the processes that may be involved in the formation and maintenance of detrimental depressive states which may turn to more enduring problems. A major strength of the present study includes the focus on within-person rather than between-person relationships. This is an asset because theories in psychopathology concern how within-person change in a mechanism variable relates to within-person change in symptoms. Important differences were identified between these two divergent types of relationships, providing clearer directions concerning promising targets for intervention that should be investigated in future studies. The present study is among the largest intensive longitudinal investigations of psychopathology in the adult population, contributing to the stability of its results. Further efforts to assess the replicability of the presented findings in independent samples and in the clinical population would benefit the literature. Finally, the use of longitudinal data and multi-level approach is powerful and overcomes many of the short-comings experienced in dynamic modeling.

7.5 Conclusions

In identifying psychopathological mechanisms and central symptoms involved in the maintenance of depressive states, investigations of within rather than between-person relationships are needed. This intensive longitudinal study identified helplessness as the main mechanism interwoven with depressive symptomatology on an across-day basis, while emotion regulation difficulties had more proximal associations with depressive symptoms. While depressed mood and anhedonia were identified as symptoms most susceptible toward being influenced by other nodes in the network, the present study identified that the two most prominent symptoms displaying outward temporal influence were worthlessness and lethargy. These symptoms had greater within-person carry-over effects across days, putting individuals at greater risk of prolonged depressive state experiences. This suggests that not all symptoms of depression should be viewed as equal in their role in maintaining this deleterious mental health state. Finally, rumination was to a greater extent susceptible to being influenced rather than temporally influencing other components involved in depressive states. These findings outline several associations between symptoms and mechanisms that are important to investigate further toward advancing the etiological understanding of depression.

Author contributions

Study design (OVE) and feedback on design (AH, SUJ). Study conceptualization (OVE) and preregistration (OVE, JB). Study administration, management, and coordination (OVE). Data acquisition and curation (OVE). Data cleaning and preparation (OVE). Data analysis (OVE, JB). Writing—original draft preparation (OVE). Writing—review and editing (OVE, JB, AH, SUJ). All authors contributed to the interpretation of the data. All authors read and approved the final manuscript.

CHAPTER 8

NETWORK ANALYSIS OF DEPRESSION SYMPTOMS AS A CONSEQUENCE OF SPOUSAL LOSS AND SEPARATION

CHAPTER 8

Abstract

Prior network analyses demonstrated that the death of a loved one potentially precedes specific depression symptoms, primarily loneliness, which in turn links to other depressive symptoms. In this study, we extend prior research by comparing depression symptom network structures following two types of marital disruption: bereavement versus separation. We fitted two Gaussian Graphical Models to cross-sectional data from a Swiss survey of older persons (145 bereaved, 217 separated, and 362 married controls), and compared symptom levels across bereaved and separated individuals. Separated compared to widowed individuals were more likely to perceive an unfriendly environment and oneself as a failure. Both types of marital disruption were strongly linked to loneliness, from where different relations emerged to other depressive symptoms. Amongst others, loneliness had a stronger connection to perceiving oneself as a failure in separated compared to widowed individuals. Conversely, loneliness had a stronger connection to getting going in widowed individuals. Analyses are based on cross-sectional between-subjects data, and conclusions regarding dynamic processes on the within-subjects level remain putative. Further, some of the estimated parameters in the network exhibited overlapping confidence intervals and their order needs to be interpreted with care. Replications should thus aim for studies with multiple time points and larger samples. The findings of this study add to a growing body of literature indicating that depressive symptom patterns depend on contextual factors. If replicated on the within-subjects level, such findings have implications for setting up client-tailored treatment approaches in dependence of contextual factors.

This chapter has been adapted from: Burger, J., Stroebe, M. S., Perrig-Chiello, P., Schut, H. A., Spahni, S., Eisma, M. C., & Fried, E. I. (2020). Bereavement or breakup: Differences in networks of depression. *Journal of Affective Disorders*, *267*, 1-8.

8.1 Introduction

8.1.1 Marital transition and mental health

One of the most well-known wedding vows suggests a long-term perspective on a relationship, with death being the only cause for its termination: "Till death do us part." Demographic data, however, suggest that the end of a marriage is not always marked by the death of a partner. Marital disruption, the termination of a marriage due to separation or divorce, has been well-established as a frequent life event. In the USA, the probability that a first marriage is still intact after 20 years has been calculated at approximately 52% for women and 56% for men aged 15–44 (Copen & Mosher, 2012).

Both spousal loss and separation are associated with major psychological distress, increasing the risk of severe long-term detriments to well-being and health. One of the most frequent consequences of spousal loss and separation are mood-related disorders, and more specifically, depression (Sbarra, 2015; Wójcik et al., 2021). The Diagnostic and Statistical Manual of Mental Disorders 5 (DSM-5; American Psychiatric Association, 2013) characterizes depression through nine criteria, namely, depressed mood, diminished interest/pleasure, weight/appetite increase/decrease, insomnia/hypersomnia, psychomotor agitation/retardation, fatigue, feelings of worthlessness or inappropriate guilt, lack of concentration or indecisiveness, and suicidal ideation. The presence of at least five of the symptoms (at least one of which have to be either sad mood or anhedonia) qualifies for the diagnosis Major Depressive Disorder (MDD). Taking into account all possible combinations of sub-symptoms, this results in over 10,000 hypothetical symptom combinations for the same diagnosis, and empirical studies have observed that many of these are realized in clients with a diagnosis of MDD (Fried & Nesse, 2015; Zimmerman et al., 2015). Crucially, different life events have been associated with differences in depressive symptomatology (Cramer et al., 2012; Fried, Nesse, et al., 2015). Based on this finding, the present study uses a network approach to investigate whether the two types of loss introduced above are differentially related to depression symptoms.

8.1.2 The network perspective to depression following bereavement

The network approach to psychopathology conceptualizes symptoms and other factors of mental health as causally interacting entities (Borsboom & Cramer, 2013). Network analyses have been applied to the field of bereavement, through the study of depression and complicated grief symptoms (Robinaugh et al., 2014, 2016) and their interrelations (Djelantik et al., 2020; Malgaroli et al., 2018). Specifically, as discussed above, Fried et al. (2015) fitted several models to a dataset to compare elderly bereaved versus still-married participants. Loneliness was much more strongly related to spousal loss than other depression symptoms, and in turn was associated with a host of other symptoms. We aim to extend this finding to compare the effects of spousal loss to marital breakup.

8.1.3 Bereavement versus breakup

There are reasons to assume differences in the symptom dynamics of depression following spousal bereavement versus marital breakup. Wrzus et al. (2013) classify widowhood as an expected life event, usually accompanied by a supportive social environment, especially after an initial phase of social withdrawal. Bereavement is predominantly associated with feelings of grief over the loss of the loved person, alongside a variety of related manifestations (Stroebe et al., 2017). While stig-

matizing responses towards bereaved individuals with a diagnosis of prolonged grief disorder have been experimentally demonstrated (Eisma, 2018), conclusive evidence regarding the prevalence of stigmatization in spousal loss is scarce; a systematic review of social support in bereaved individuals found that most studies conducted on this issue face several methodological and sampling limitations (Logan et al., 2018). In a previous network study, Fried, Bockting, et al. (2015) found that people who had lost a loved one primarily developed loneliness over other depressive symptoms; loneliness, in turn, was related to a host of other depressive symptoms. The authors speculated that loneliness might thus be a gateway symptom which prevention strategies for depression could focus on to disrupt relations with other symptoms following spousal loss.

While one can make similar predictions about loneliness following marital breakup (especially perhaps for those who did not initiate the separation, cf. Hewitt & Turrell, 2011), other symptoms of depression would seem likely to be important as well.

Wrzus et al. (2013) noted that separation (specifically: divorce) can be especially stressful due to the reduction in a person's social network, through the partial loss of in-laws and spouse's friends. Given that breakup is associated with adverse interpersonal relationship experience (Sbarra, 2015), items representing the perceived negative opinions and social responses of others might thus be as or even more apparent, compared to loneliness. Measures of depression include relevant items; the CES-D (Radloff, 1977) items "I thought my life had been a failure" and "People were unfriendly" (in the following referred to as *failure* and *unfriendly*, respectively) thus arguably capture the experience of breakup better than bereavement.

Following these contrasts in marital transition, crucial differences in the nature of mental health-related difficulties could be expected: For bereaved individuals, one could argue that loneliness as a consequence of spousal loss (Fried, Bockting, et al., 2015) is accompanied with symptoms related to grief work. Separated individuals on the other hand are more liable to evaluate their life plan as a failure, with their social environment often compounding this due to lack of support and/or understanding (Wrzus et al., 2013).

8.1.4 The current study

We estimated network models and compared symptom levels following widowhood and separation, compared to a still-married sample and tested three hypotheses:

H1. CES-D sum-scores are higher among both bereaved and separated individuals compared to married individuals.

H2. Separated individuals show higher levels of *failure* and *unfriendly* compared to widowed individuals.

H3. Both loss types are primarily linked to *loneliness*, which in turn is associated with other CES-D symptoms.

A note on exploratory analyses. Network analysis at present is largely used to gain exploratory insight into multivariate dependencies. These structures can generate hypotheses about putative causal relations. To this end, we extend our investigation to interesting relations that have not been hypothesized. These exploratory analyses are distinguished from our confirmatory findings

(the latter include the respective hypothesis in brackets). Most importantly, we are interested in how *loneliness* is differentially related to other CES-D symptoms, comparing bereaved with separated individuals.

8.2 Methods

8.2.1 Participants

We analyzed data from the Swiss project "Relationships in later life" (http://www.kpp.psy.unibe.ch/forschung/projekte/nccrlives/index_ger.html). In this project, information on marital transitions and related mental health components were collected over three waves (2012, 2014, and 2016). The Swiss Federal Statistical Office identified a random sample (stratified by gender, age, and marital status) of 6889 married, widowed, divorced and separated individuals aged 40 – 90. These individuals subsequently received letter mail with an invitation to the study and the paper-and-pencil questionnaire. Additionally, advertisements were placed on various platforms (radio, newspaper, and online). Participants were informed regarding the purpose of the prospective longitudinal data-collection (changes and stability of relationships in later life). In total, data on 1,276 married, 566 widowed, 721 divorced, and 250 separated individuals were collected, from which we derived two marital status sub-samples. A schematic overview of the sampling procedure in this study can be seen in Figure 8.1.

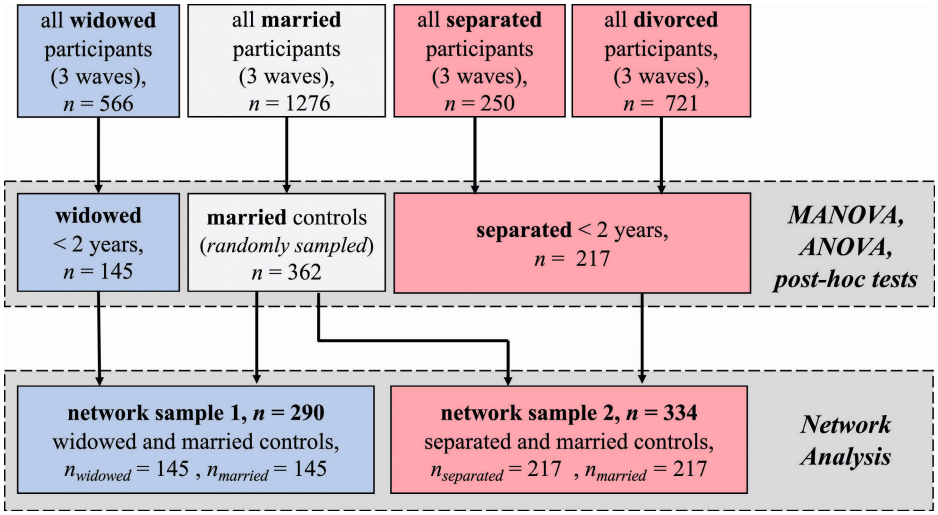


Figure 8.1. Schematic set-up of the samples and analyses used in this study. Inclusion criteria for separated/widowed individuals were a) a maximum time-distance to the respective life event of two years, and b) that the participant was not living in a new partnership. Married controls were randomly sampled from the pool of married participants. In order to be able to model the loss-type in the networks, an equal amount of married controls was added to both samples.

8.2.1.1 Widowed and separated individuals

We sampled widowed and separated individuals from all three waves, if they met two inclusion criteria: First, the loss/breakup occurred within two years prior to assessment, and second, the widowed/separated person did not have a new partner at the time of assessment.

The former criterion was chosen on the basis of two considerations: On the one hand, due to the way data was collected (time distance of two years in between waves), extending the time criterion to more than two years would mean that participants who experienced loss/breakup more than two years prior to wave 2 and 3 would be sampled multiple times (from several waves). On the other hand, decreasing time intervals to less than two years would have led to rather low sample sizes in the present dataset. We therefore faced a trade-off between statistical power and capturing experiences in close approximation to the life event, and opted for a compromise of two years. We hope that future research will investigate effects of different time distances to the life event to capture both, adaptation over longer periods including more complex processes of loss and depression, as well as experiences in close approximation to the life event).

The second criterion was chosen to account for protective influences that a new partnership might have on an individual's grief (Gierveld, 2004). This resulted in 145 widowed and 217 separated individuals.

8.2.1.2 Samples for network analysis

We see two main possibilities for constructing networks to tackle our research questions: a) adding married participants as controls/contrast to both the widowed and the separated sample, and estimating two networks for the respective samples (using a similar logic to Fried, Bockting, et al., 2015), or b) estimating three separate networks for the three groups widowed, separated and married. The main difference between these approaches is that the networks estimated in method a) allow us to include the life event as a node in the network, which is not possible for networks estimated in method b). This is because in method b), the samples are set up in a way that each participant experienced *the same life event* within one sample. The variable 'life event' thus has no variance, consequently making it impossible to estimate (partial-)correlations between the life event and other variables.

Since the focus of our analysis is to examine differences in how widowhood and separation are (differentially) related to depressive symptoms, we estimated two networks according to option a), while providing the networks resulting from the estimation method b) in the supplemental material (Supplement C, Figure S8.1). The networks estimated according to method b) can be relevant in focusing on structural differences of depressive symptoms within each sample, if relations to the life event are not of interest. Accordingly, we randomly sampled 362 married controls who did not previously experience spousal loss or separation/divorce, and constructed two samples that were then used to estimate the networks. The first sample consisted of the 145 widowed individuals introduced above combined with 145 married controls, the second sample of 217 separated individuals combined with the remaining 217 married controls. Table 8.1 compares demographic characteristics across the widowed, separated and married sample.

Table 8.1. Demographics of the widowed, separated and married sample.

	Widowed < 2 years, *n* = 145		Separated < 2 years, *n* = 217		Married controls, *n* = 362		Comparing widowed against separated sample		
	M	*SD*	*M*	*SD*	*M*	*SD*	Difference tests	Significance	Effect size, confidence interval
1. Gender, (% female)	79.31	-	76.04	-	52.76	-	$X^2(1) = 0.53$	$p = .466$	$w = 0.001$
2. Age	71.80	11.90	51.88	8.43	64.69	13.64	$t(238.72) = 17.44$	$p < .001$***	$d = 1.93$, *CI* [17.67, 22.17]
3. Duration of marriage (years)	16.58	9.97	21.86	11.03	11.52	6.72	$t(12.54) = 1.78$	$p = .100$	$d = 0.50$, *CI* [-1.17, 11.73]
4. Time since separation (months)	11.95	7.29	11.23	7.20	-	-	$t(306.15) = 0.93$	$p = .352$	$d = 0.10$, *CI* [-2.26, 0.81]
5. CES-D sum score	11.65	6.72	13.47	9.91	6.67	6.07	$t(306.52) = 1.93$	$p = .055$	$d = 0.21$, *CI* [-3.66, 0.04]

We decided to sample married controls randomly as opposed to making use of matching procedures, since several demographic variables of interest had many missing observations. To ensure that estimated network structures were not dependent on the seed chosen to sample married controls, we repeated the sampling procedure four times with other random seeds, and correlated the adjacency matrices of the resulting network with the one discussed below. Correlations ranged from 0.89 to 0.92 for the widowed, and from 0.92 to 0.94 for the separated network, indicating that the network structures had high consistency for different compositions of the married sample.

8.2.2 Outcome measures

Depressive symptoms were assessed using the German short version of the Center for Epidemiologic Studies Depression scale (German: Allgemeine Depressions-Skala; Meyer & Hautzinger, 2001; CES-D; Radloff, 1977). Participants rated 15 items with respect to the frequency with which they occurred in the last week, with the four response categories "rarely or none of the time (less than 1 day)", "some or a little of the time (1–2 days)", "occasionally or a moderate amount of time (3–4 days)" and "most or all of the time (5–7 days)". The German version of the CES-D has been found to be reliable with Cronbach's Alpha between 0.89 and 0.92 (Hautzinger & Geue, 2016). In line with these findings, we obtained a Cronbach's alpha of 0.90 for our study sample. While the CES-D is used as a screening-tool and does not allow to determine diagnostic status, it provides useful information regarding our proposed differences in comparison to other scales. Specifically, the CES-D items "I thought my life had been a failure" and "People were unfriendly" are relevant to investigate the above discussed differences in social support and evaluation of one's life.

One major challenge in the extant network literature in psychopathology is that some items modeled in networks might measure the same construct (Fried & Cramer, 2017). This poses a

problem for inferences because edges in network models should only be interpreted as putative causal relations if the nodes are indeed distinct entities. At present, there are no clear guidelines to differentiate between a correlation that arises from items measuring *the same construct* and a correlation due to two items being related, but originating from *distinct constructs*. Since purely data-driven approaches cannot account for theoretical considerations, we combined items if they met two criteria. Items were combined if the items showed correlations of $r \geq 0.50$, and if the items could be understood to measure the same construct. Accordingly, we combined the items *mood, upset* and *depressed* into the new item *mood*, and *happy* and *enjoy* into the new item *happy*, resulting in 12 instead of 15 items. The final list of items is presented in the supplemental materials, Table S8.1. The item-pairs *depressed – concentration, concentration – exhausted, lonely – mood, lonely – depressed, sad – depressed, getgo – depressed, getgo – exhausted* and *lonely – sad* all exhibited correlations of $r \geq 0.50$, however, for the purpose of this chapter, we understand them as theoretically separate constructs.

8.2.3 Statistical analyses

8.2.3.1 Symptom level comparison

Prior to the network analyses, widowed, separated and married individuals were compared with respect to differences in the item sum-score using a one-way ANOVA and post-hoc tests. Furthermore, overall differences with respect to specific symptoms were analyzed in a MANOVA and symptoms were examined individually with respect to group differences.

8.2.3.2 Network analysis

Following the group comparisons, we estimated two separate networks. Both networks consisted of the combined set of 12 CES-D items and one node to the life event (network 1: spousal loss versus marriage, network 2: marital breakup versus marriage). We estimated regularized partial correlation networks (Epskamp & Fried, 2018) based on Spearman's rank correlation, due to the ordinal nature of items. We chose Spearman correlations over polychoric correlations, since polychoric correlations led to highly unreliable parameter estimates; as explained elsewhere (Epskamp, Borsboom, et al., 2018), this can happen when the sample size is small, items have few response options, and are considerably skewed. To account for potential spurious relations, we used a regularization approach with the tuning parameter γ (specifying the level of sparsity) set to 0.5 (Foygel & Drton, 2010). Recent literature suggests that non-regularized networks might be preferable in some cases, especially for very large sample sizes (Williams et al., 2019b). Since this is not the case for our sample, we present the non-regularized partial correlation networks in the supplemental material (Supplement C, Figure S8.2).

It is good practice to determine the accuracy and stability of estimates and inferences in the networks. To this end, we conducted the stability/accuracy routine using the *bootnet* package in R described elsewhere (Epskamp, Borsboom, et al., 2018). The networks were estimated using the *bootnet* and the *qgraph* package (Epskamp et al., 2012). Additionally, we compared the two networks using the *NetworkComparisonTest* (van Borkulo, Millner, et al., 2017). Since this procedure might yield biased results if the network samples are unequal in size (van Borkulo et al., 2015), we

additionally correlated the weight matrices to obtain a measure of similarity, and subtracted the weight matrices to examine the largest absolute differences between edge weights.

Contrary to many network analyses conducted in the field of psychopathology, we did not calculate centrality measures for our networks. Most centrality measures are metrics based on summarizing edge weight information in respect to a given node, degree centrality for instance is calculated by summing all absolute edge weights going into a node. Our networks are composed of both, CES-D items and a node coding a life event, consequently making the interpretation of centrality measures as indicative of central to the network of symptoms problematic. This is because centrality metrics in our case would favor items that exhibit large relations to the life event over items that are unrelated to the life event. For that reason, we focused on comparing specific edges rather than centrality measures.

8.3. Results

8.3.1 Symptom level comparison

8.3.1.1 Sum-score and diagnosis of depression

Widowed (n = 145), separated (n = 217) and married (n = 362) individuals differed in their overall CES-D sum-score, $F(2, 609) = 52.93$, $p < .001$, Cohen›s $f = 0.34$. More specifically, sum-scores of married individuals ($M_{mar} = 6.67$, $SD_{mar} = 6.07$) were lower than those of widowed individuals ($M_{wid} = 11.65$, $SD_{wid} = 6.72$; $t(194.50) = 6.98$, $p < .001$, Cohen›s $d = 0.78$, CI [3.58, 6.39]) and separated individuals ($M_{sep} = 13.47$, $SD_{sep} = 9.91$; $t(293.48) = 8.62$, $p < .001$, Cohen's $d = 0.83$, CI [5.24, 8.35]), but the widowed and separated groups did not differ from each other ($t(306.52) = 1.93$, $p = .055$, Cohen›s $d = 0.21$, CI [−3.66, 0.04]), supporting our first hypothesis (H1). While the CES-D does not allow for determining diagnostic status, prior psychometric analyses (Lehr et al., 2008) suggested a score of 18 for a putative diagnosis. Following this cutoff, 6.04% of the married, 17.95% of the widowed and 29.95% of the separated individuals met the screening criterion of the scale.

8.3.1.2 Differences in specific symptoms

A MANOVA revealed overall differences between widowed (n = 145) and separated (n = 217) individuals with respect to specific CES-D items, $T^2(12, 301) = 4.91$, $p < .001$. In particular, as can be seen in Figure 8.2, differences emerged only for specific symptoms.

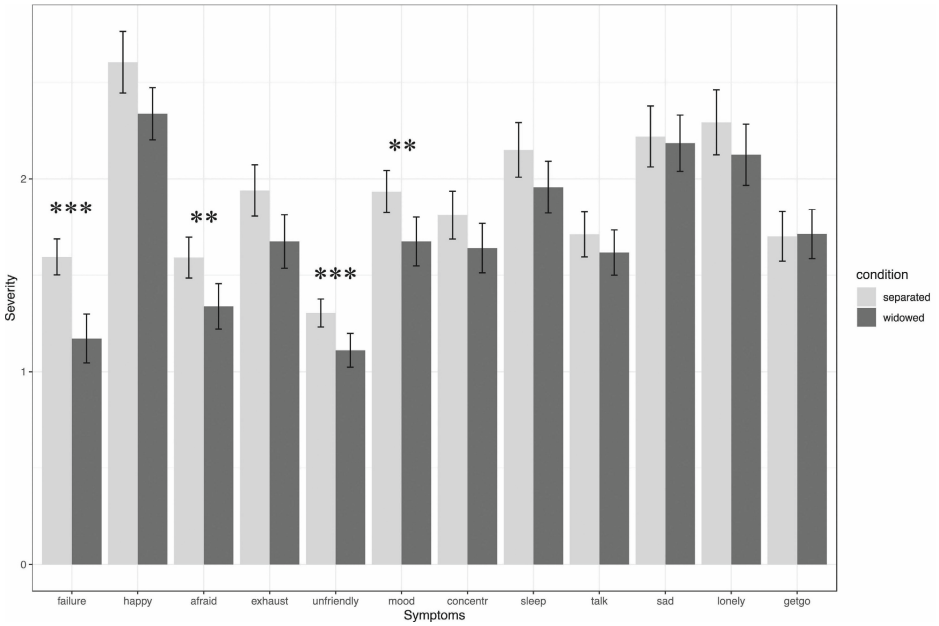


Figure 8.2. Post-hoc comparisons for all CES-D symptoms between separated and widowed individuals, sorted by decreasing mean differences. 95% confidence intervals are indicated. Note that we only indicated significance levels for items that were significant after correcting for multiple testing using the Bonferroni method.

*** **significant** at 0.001; ** significant at 0.01; * significant at 0.05.

As hypothesized (H2), and after accounting for multiple-testing using Bonferroni-correction, separated individuals showed higher levels of *failure* ($t(343) = 5.56$, $p < .001$, Cohen›s $d = 0.58$, *CI* [.27, 0.57]) and *unfriendly* ($t(343) = 3.59$, $p < .001$, Cohen›s $d = 0.36$, *CI* [.09, 0.30]) compared to widowed individuals. Furthermore, there were differences for the symptoms *afraid* ($t(345.98) = 3.17$, $p = .002$, Cohen›s $d = 0.33$, *CI* [.10, 0.41]; separated > widowed) and *mood* ($t(319.35) = 3.03$, $p = .003$, Cohen›s $d = 0.33$, *CI* [.09, 0.43]; separated > widowed).

Some other symptoms indicated significant differences between separated/widowed individuals (*exhaust,* $t(318.96) = 2.78$, $p = .006$, Cohen›s $d = 0.30$, *CI* [.08, 0.45], separated > widowed; *sleep,* $t(321.96) = 2.04$, $p = .043$, Cohen›s $d = 0.22$, *CI* [.01, 0.38], separated > widowed; *happy,* $t(281.39) = 2.60$, $p = .010$, Cohen›s $d = 0.28$, *CI* [.07, 0.47], separated > widowed), however these did not remain significant after controlling for multiple testing. Given that some of these p-values were close to the traditional significance threshold of 5%, we want to call for caution in interpreting these effects as either clear positive or negative effects (Amrhein et al., 2019); more conclusive evidence will require replicating our study.

8.3.2 Network analysis

8.3.2.1 Network accuracy and stability

Graphical results of the stability and accuracy analysis can be found in the supplemental materials (Supplement C, Figures S8.3–S8.5). In general, the edge weights exhibit rather large confidence

intervals, and some of the lower absolute edge weights do not differ significantly from other edges, indicating that the *order* of edges should be interpreted with some caution.

8.3.2.2 Network inferences

Figure 8.3 shows the estimated networks for the widowed/married (a, left) and the separated/married (b, right) sample.

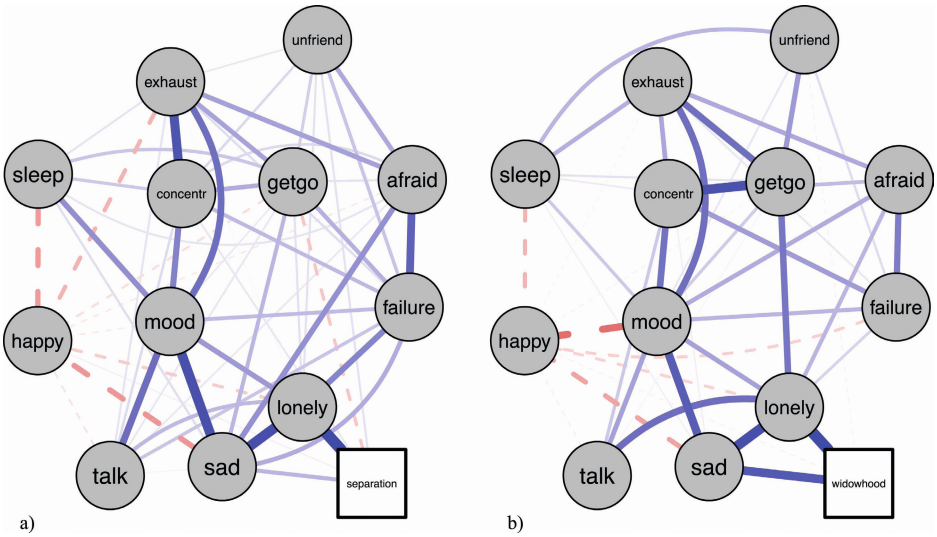


Figure 8.3. Regularized partial correlation network of the combined set of CES-D symptoms and spousal loss (a, 145 widowed individuals and 145 married controls) and marital breakup (b, 217 separated individuals and 217 married controls). Solid blue lines represent positive edges, dashed red lines represent negative edges.

8.3.2.2.1 Widowhood. As hypothesized (H3), and in line with prior findings of Fried et al. (2015), experiencing spousal loss was primarily associated with *loneliness* (partial correlation of $r = 0.30$), and additionally with *sadness* ($r = 0.26$). In turn, *loneliness* was linked to several CES-D symptoms (sorted by decreasing partial-correlation): *talk* ($r = 0.17$), *getgo* ($r = 0.16$), *mood* ($r = 0.11$), *afraid* ($r = 0.09$), *happy* ($r = -0.06$), and *failure* ($r = 0.06$). In contrast to Fried et al. (2015), this analysis additionally revealed a strong direct relation between spousal loss and *sad* ($r = 0.22$) and weaker associations with *unfriendly* ($r = -0.01$) and happy ($r = -0.01$).

8.3.2.2.2 Separation. As hypothesized (H3), and similar to the widowed network, separation was also strongly linked to *loneliness* ($r = 0.33$). *Loneliness* was in turn associated with other CES-D symptoms (sorted by decreasing partial correlation): *sad* ($r = 0.29$), *failure* ($r = 0.16$), *mood* ($r = 0.14$), *talk* ($r = 0.10$), *happy* ($r = -.07$), *getgo* ($r = 0.04$), *unfriendly* ($r = 0.04$), and *exhausted* ($r = 0.01$). Next to *loneliness*, this network also exhibited somewhat weaker direct relations to the life event: *sad* ($r = 0.10$), *getgo* ($r = -.08$), *unfriendly* ($r = 0.04$), and *happy* ($r = 0.02$).

8.3.2.3 Network comparison

To compare the networks globally, we first calculated the correlation of the adjacency matrices to obtain a measure of similarity, and second conducted the *NetworkComparisonTest*. The correlation between the adjacency matrices was $r = 0.75$, indicating that overall, the two network structures were largely similar. The *NetworkComparisonTest* revealed a significant result for the global invariance test ($p = .005$), indicating that there were some differences in the overall structure between the networks.

Of specific interest for our hypotheses (H3) was the extent to which *loneliness* following the two life events was differentially related to other CES-D symptoms. In an exploratory analysis, we investigated for which edges the two network structures showed the maximum difference, through subtracting their weight matrices. We visualized the largest absolute differences between edges in a network (Fig. 4). The largest absolute differences between estimates were obtained for theedges *happy – mood* ($\Delta_r = 0.15$), *exhaust – concentration* ($\Delta_r = 0.15$), *afraid – sad* ($\Delta_r = 0.15$), *getgo – concentration* ($\Delta_r = 0.13$), *separation/widowhood – sad* ($\Delta_r = 0.12$), *afraid – unfriendly*($\Delta_r = 0.12$), *lonely – getgo* ($\Delta_r = 0.12$), *lonely – failure* ($\Delta_r = 0.11$), *sad – failure* ($\Delta_r = 0.11$), and *getgo – failure* ($\Delta_r = 0.11$). With respect to our hypotheses (H3), differential associations with *loneliness* could be found to *failure* and *getgo*.

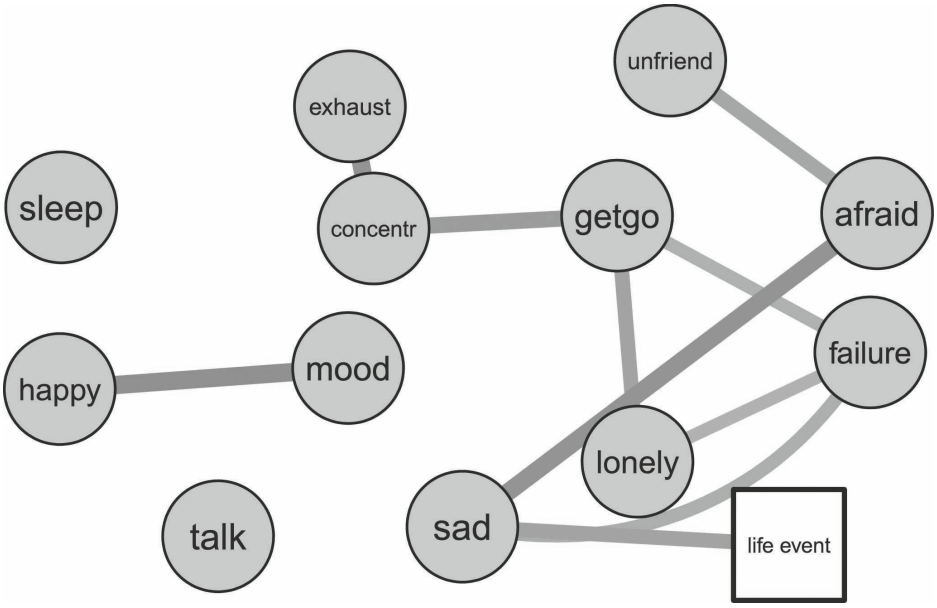


Figure 8.4. Network indicating the ten largest absolute differences in edge weights for the widowed network compared to the separated network, based on the difference scores of the respective weight matrices.

8.4. Discussion

Different life events may lead to different depressive symptoms, not only in overall quantity – some life events have more severe consequences than others – but also in quality. Since episodes of major depressive disorder are often preceded by severe stress or adverse life events (Hammen, 2005), the idea that different life events lead to different symptom profiles could explain a large part of the dramatic heterogeneity of depression symptoms (Fried & Nesse, 2015; Zimmerman et al., 2015).

To our knowledge, this is the first study to investigate potential differences in depressive symptomatology between spousal loss and marital breakup by comparing symptom profiles and modeling the relationship between life events and symptoms via network models. We showed that one of the main differences between the two life events is a stronger feeling of experiencing an *unfriendly* environment and oneself as a *failure* within separated compared to widowed individuals. This finding is consistent with literature regarding consequences of the reduction in social network following separation and its effect on the individual's psychosocial well-being (Wrzus et al., 2013).

The network of bereaved individuals is largely consistent with previous findings of Fried, Bockting, et al. (2015), indicating that spousal loss is primarily connected to *loneliness*, in turn connecting to other depressive symptoms. Additionally, we found a strong link between spousal loss and *sadness*. The present study extends this finding to a different type of marital disruption; similar to spousal loss, marital breakup was also primarily linked to *loneliness*. Overall, the two networks showed largely similar structures, as indicated by a large correlation between their weight matrices.

In an exploratory analysis, we investigated the largest differences in edges between the two networks. Experiencing oneself as a *failure* revealed a stronger connection to *loneliness* in separated compared to widowed individuals. For widowed individuals, we obtained stronger links for lonely – getgo, getgo – exhaust, and getgo – concentration. Keeping in mind the exploratory nature of this analysis, these findings give rise to two hypotheses: 1) Loneliness in separated compared to widowed individuals is more strongly associated with symptoms related to the normative evaluation of the life event (stronger relation of loneliness with experiencing oneself as a failure), and 2) loneliness in widowed compared to separated individuals is more strongly associated with symptoms related to the person's level of activity and cognitive capacities (stronger relations of loneliness with getting going, and getting going with exhaustion and concentration).

8.4.1. Implications for future research and clinical practice

In line with previous research (Cramer et al., 2012; Fried, Nesse, et al., 2015), our study provides further evidence of the importance of contextual information in explaining depressive symptom patterns. In clinical practice, this could provide important information in conceptualizing a client's case, in understanding the etiology of depression, and in identifying potential treatment targets. This study indicates that the main difference in widowed compared to separated individuals might be characterized through a) differences in the *intensity* of specific symptoms (i.e., experiencing oneself as a failure and an unfriendly environment), and b) differences in specific relations to for example loneliness (e.g., failure and get going). These findings can help tailoring treatment approaches to characteristics of a given life event.

For both groups, prevention strategies targeting *loneliness* might be promising. For widowed and separated individuals specifically, one could try to disrupt relations between loneliness and other symptoms, if these can be replicated in other work. For instance, this study suggests that separated individuals would additionally benefit from learning that experiencing loneliness does not mean that their life plan is a failure (i.e., disrupting the association between *loneliness* and *failure*), and widowed individuals could benefit from a stronger focus on helping them "getting going", for instance through behavioral activation (Papa et al., 2013).

8.4.2. Limitations

The results of this study must be interpreted in the light of some limitations. First, we analyzed cross-sectional data, any conclusions regarding dynamics remain thus putative. Further, the time scale on which depressive episodes unfold may differ between participants, depending on the complexity of their depressive patterns. In a follow-up study, it would be important to include several time points to aim to estimate Granger-causal relations between life events and symptoms, and test effects of varying time distances to the life events of interest.

Second, as became evident in the accuracy and stability analysis, many parameters are estimated with at best moderate precision. Our study faced a trade-off between sample size and the time passed since the critical life event, and we opted for a compromise of less than two years. We hope to replicate our finding in larger datasets of bereaved and separated individuals – once these become available – which will allow for stricter screening. This would also allow us to differentiate between potentially meaningful subgroups, such as initiators and non-initiators of separation (Hewitt & Turrell, 2011).

Third, separated individuals were significantly younger widowed individuals in this study. This might be considered a potential confound and limit the extent to which results can be generalized to other age groups. Demographic data (Copen & Mosher, 2012) suggest that separation is indeed more prevalent among younger individuals, whereas elderly individuals are more likely to experience spousal loss compared to separation. The precise role of age in expressing specific symptoms thus remains a topic for future research.

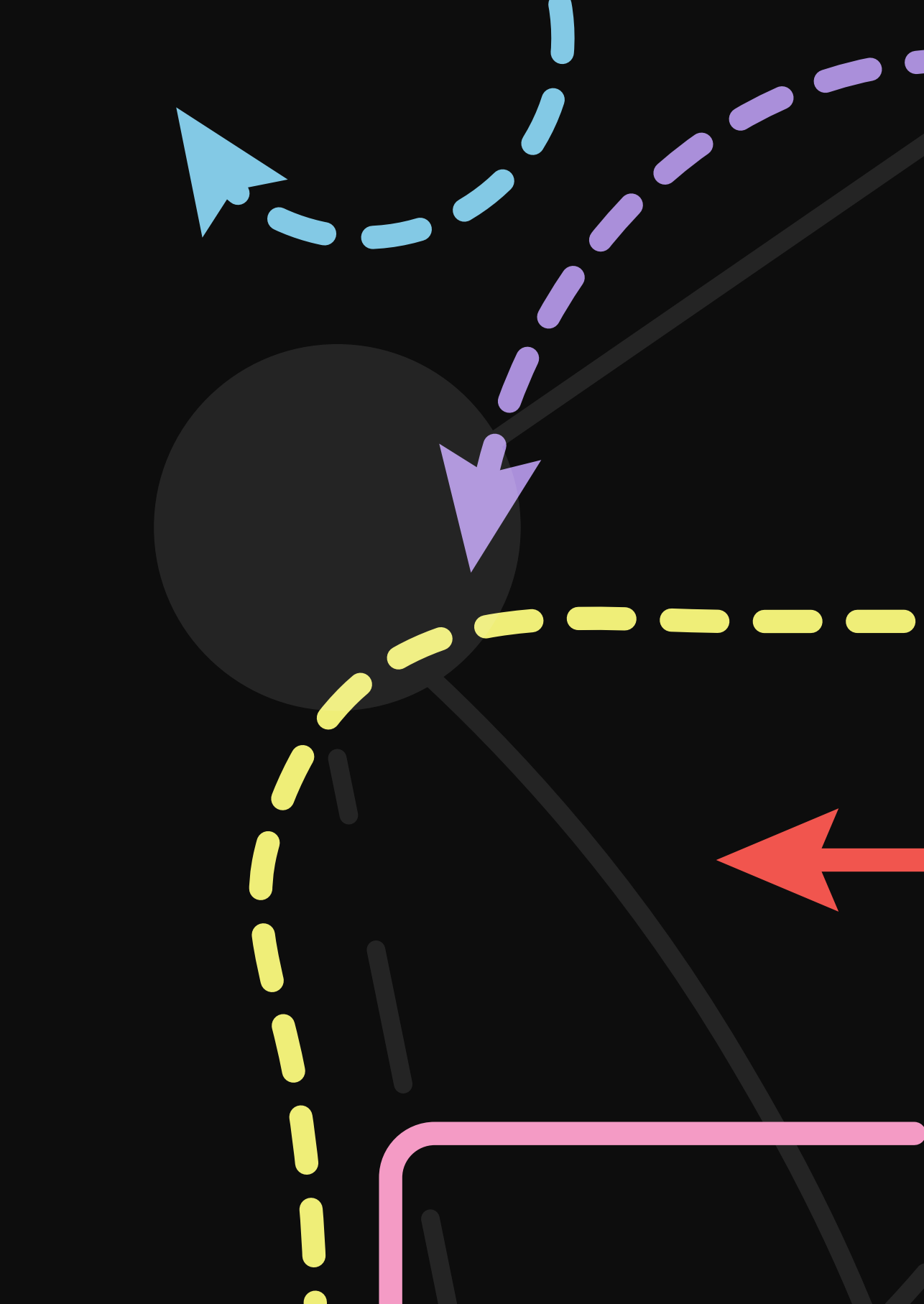
Fourth, when applying network analyses to psychological scales, the choice of the scale and the topological overlap of its items might drastically influence the structure of the resulting network (Fried & Cramer, 2017). In the present dataset, we identified variables that could have been potentially relevant to add to our network investigation, more specifically contextual information regarding the cause of death in widowed participants, reasons for separation, and the Prolonged Grief Disorder-13 (PG-13; Prigerson et al., 2009) tool, however, these variables have unfortunately not been assessed at all three waves, and therefore were not suitable to be included in our analyses. Since reactions to loss experience have been linked to these specific symptoms of Prolonged Grief Disorder (PGD; Prigerson et al., 2009), we encourage to include such variables in future studies. Furthermore, since the network structure is based on partial correlations, excluding or combining items will lead to different network structures. This is why we, unlike most prior studies in the field, decided to thoroughly study item content, and modified the constructs under investigation based on a thresholding rule. However, this issue needs more attention from both clinical theories

and empirical research, and decisions should in the best case be guided by both statistical tests and theoretical considerations.

Lastly, we used the CES-D for this analysis. The CES-D contains the items *loneliness* and experiencing oneself as a *failure*, which were important for our research questions. On the other hand, it is a screening tool for depression but is not used for the actual diagnosis of depression according to the DSM-5 (American Psychiatric Association, 2013), and differs considerably from other depression scales in terms of content (Fried, 2017). It would thus be interesting to model a broader range of depressive symptoms in future studies.

8.5. Conclusions

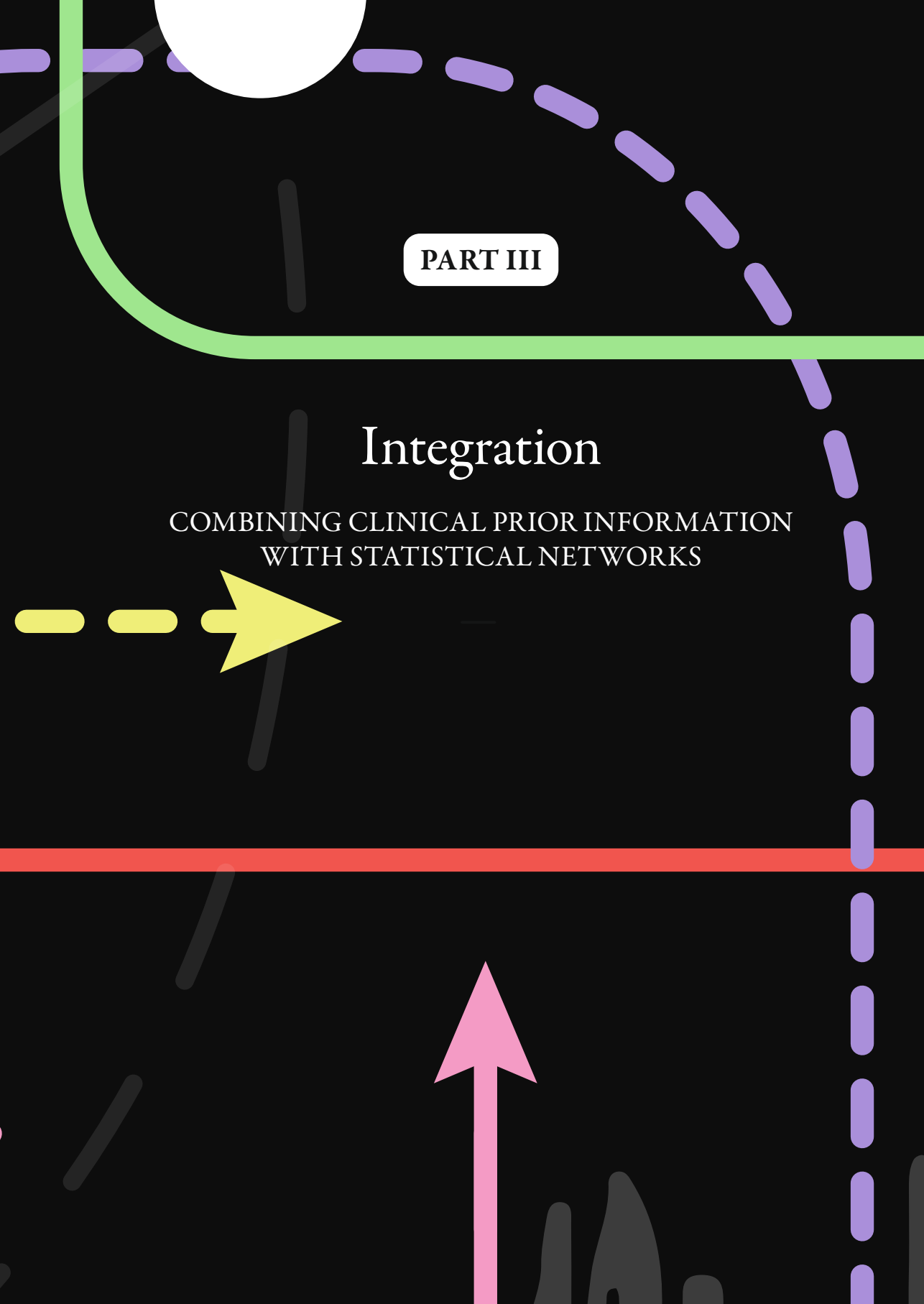
This study provides further evidence for the relation between specific adverse life events and different symptom patterns of depression. Network models are a promising tool in understanding these differential relations, and can be used to compare spousal loss with marital disruption in this regard. A better understanding of these differences can in turn help in tailoring interventions to specific contextual factors.



PART III

Integration

COMBINING CLINICAL PRIOR INFORMATION WITH STATISTICAL NETWORKS



CHAPTER 9

PREMISE: THE PRIOR ELICITATION MODULE FOR IDIOGRAPHIC SYSTEM ESTIMATION

CHAPTER 9

Abstract

Over the past decade, the idiographic approach has received significant attention in clinical psychology, incentivizing the development of novel approaches to estimate statistical models, such as personalized networks. Although the notion of such networks aligns well with the way clinicians think and reason, there are currently several barriers to implementation that limit their clinical utility. To address these issues, we introduce the Prior Elicitation Module for Idiographic System Estimation (PREMISE), a novel approach that formally integrates case formulations with personalized network estimation via prior elicitation and Bayesian inference. PREMISE tackles current implementation barriers of personalized networks; incorporating clinical information into personalized network estimation systematically allows theoretical and data-driven integration, supporting clinician and client collaboration when building a dynamic understanding of the client's psychopathology. To illustrate its potential, we estimate clinically informed networks for a client suffering from obsessive-compulsive disorder. We discuss open challenges in selecting statistical models for PREMISE, as well as specific future directions for clinical implementation.

This chapter has been adapted from: Burger, J., Epskamp, S., van der Veen, D. C., Dablander, F., Schoevers, R. A., Fried, E. I., & Riese, H. (2022). A clinical PREMISE for personalized models: Toward a formal integration of case formulations and statistical networks. *Journal of Psychopathology and Clinical Science*, *131*(8), 906.

9.1 Introduction

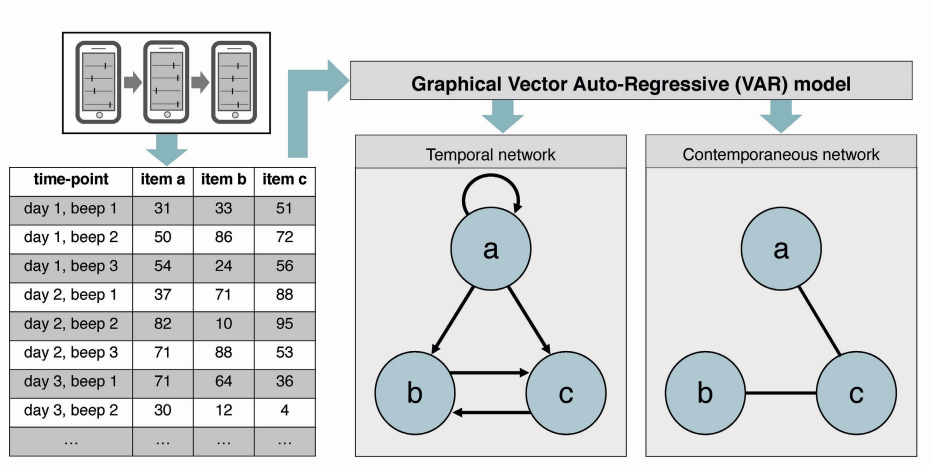
In recent years, the idiographic approach received significant attention in psychopathology research (Barlow & Nock, 2009; Fisher et al., 2018; Molenaar, 2004; Molenaar & Campbell, 2009). Proponents of this approach emphasize that individuals differ considerably in their symptomatology and etiology, even within the same diagnosis. It follows that findings from group-level studies can only be generalized to within-person processes under very strong, potentially unreasonable assumptions. The idiographic approach therefore calls for a stronger focus on processes at the *individual* level (Hayes et al., 2019; Hofmann & Hayes, 2019; Zuidersma et al., 2020), one that aims to identify *the right treatment for the right client at the right time*. This call for personalization in psychopathology research has incentivized the development of new statistical approaches that allow clinicians to estimate personalized models (Piccirillo & Rodebaugh, 2019; Wright & Woods, 2020).

9.1.1 Statistical advances in idiographic research: Personalized networks

An increasingly popular example in the area of idiographic modeling is the use of personalized networks (Epskamp et al., 2018; Wild et al., 2010) estimated from data collected via the Experience Sampling Method (ESM; Myin-Germeys et al., 2018; Shiffman, Stone, & Hufford, 2008). Such networks aim to display dynamic interactions between personalized variables and may guide tailored intervention planning (Henry et al., 2020; Rubel et al., 2018). One commonly used approach to estimating idiographic networks is based on the *vector auto-regressive (VAR) model* (Bringmann, 2021), predicting the current score of each variable by (a linear combination of) the scores of all variables at one (or multiple) previous measurement occasion(s). The VAR model can be used to derive *temporal* relationships (indicating predictive effects over time), as well as *contemporaneous* relationships (indicating effects within the same time frame). Figure 9.1 shows a schematic example of estimating temporal and contemporaneous networks from ESM data of a client.

Idiographic networks have been applied to a vast range of psychological disorders, such as personality disorders (Dotterer et al., 2020), eating disorders (Levinson, Cash, et al., 2020; Levinson et al., 2021), depression (Wichers et al., 2021), and anxiety disorders (Fisher et al., 2017; Lutz et al., 2018)[33]. Van Os and colleagues (2013) emphasized that the use of precision diagnoses via ESM derived personalized networks can increase empowerment in clients, and Kaiser and Laireiter (2018) suggested that these models can provide insight into the interaction between symptoms and therapy processes. Another particularly relevant application of personalized networks is the identification of tailored interventions (Epskamp, van Borkulo, et al., 2018; Henry et al., 2020; Rubel et al., 2018). It should be noted, however, that these methods to identify intervention targets are heuristic and require further scrutiny. This is because network models are statistical models that, by themselves, do not allow for causal inference (Dablander & Hinne, 2019; Pearl et al., 2016; Ryan et al., 2019).

33 Some of these references use multi-level approaches (Bringmann et al., 2013; Epskamp & Bringmann, 2019) that are not truly idiographic in the sense that data from only one individual are used for network estimation (Piccirillo & Rodebaugh, 2019; Wright & Woods, 2020).



time-point	item a	item b	item c
day 1, beep 1	31	33	51
day 1, beep 2	50	86	72
day 1, beep 3	54	24	56
day 2, beep 1	37	71	88
day 2, beep 2	82	10	95
day 2, beep 3	71	88	53
day 3, beep 1	71	64	36
day 3, beep 2	30	12	4
...	...	...	...

Figure 9.1. Illustrating the process of estimating vector auto-regressive (VAR)-based personalized networks from ESM data, see also chapter 3. Temporal and contemporaneous relationships can be calculated via components of the VAR model, which predicts the current score of a variable from previous scores of all variables. The resulting model can be used to construct temporal networks (left, directed conditional relationships) and contemporaneous networks (right, undirected conditional relationships).

9.1.2 From personalized networks to case formulations: The inference gap

Idiographic reasoning is not new to therapeutic practice. Indeed, the *case formulation approach to cognitive behavioral therapy* (Persons, 2012, 2006; Persons & Talbot, 2019; Kuyken et al., 2009) provides a concrete framework to extrapolate individual models from nomothetic theories and tailor evidence-based interventions to the client's specific psychopathology, thinking patterns, and resources. Page and Stritzke (2014) formulated a science-informed model for clinical practice that embeds case formulations within the therapeutic process (see also Page, Stritzke, & Mclean, 2008), which we will draw on in this chapter.

Constructing case formulations can be challenging, and personalized networks could therefore provide supportive exploratory insights into dynamic relationships between variables (von Klipstein et al., 2020). Current efforts to implement personalized networks into clinical practice are primarily focused on using the resulting statistical models to investigate client-specific dynamics. Although the *content* of the ESM items is grounded in clinical considerations, the *relationships* between the items – the connections in the network – are most commonly established using data-driven routines that disregard clinical theory and expertise. For this reason, we here refer to these routines as *agnostic estimation*. In this data-driven approach, clinician and client determine a personalized set of ESM items, the client collects data in between the sessions, typically repeatedly throughout the day, and personalized networks are subsequently estimated from the collected data. The resulting networks can then be used to stimulate a dialogue between clinician and client regarding the identified dynamics, and may provide a rationale for tailored interventions (Rodebaugh et al., 2020; von Klipstein et al., 2020). In doing so, personalized networks promise to provide a powerful tool that can inform the construction of case formulations and therefore overcome the therapist's dilemma–the challenge to tailor nomothetic principles and treatment indications to

the circumstances faced by a specific client (Piccirillo & Rodebaugh, 2019; Piccirillo, Beck, et al., 2019; Rodebaugh et al., 2020).

This promise to use personalized networks in practice has experienced a tempered response in both the clinical and technical literature. In the following, we will summarize some of the main theoretical, technical, and practical limitations to the agnostic approach.

9.1.2.1 Limitation 1: Lack of clinical considerations

As discussed later in chapter 12, clinicians may not see utility in using personalized networks if these fail to acknowledge their intuitions or fail to capture the client's experience. Indeed, the central aim of constructing case formulations is to integrate these considerations by connecting the client's presenting psychopathology with clinical theory, empirical literature, and clinical expertise. As discussed above, personalized networks are in most cases estimated *agnostically*, that is, their parameters are based on a data-driven algorithm that lacks the flexibility to incorporate clinically relevant prior information.

9.1.2.2 Limitation 2: Inaccurate estimation

Statistical networks usually consist of many parameters that need to be estimated, which in turn requires a large number of observations to arrive at reliable estimates. For a relatively simple network of five variables, a graphical VAR network model contains 35 parameters. Note that the number of variables included in personalized ESM assessment is typically much higher, at least 20 (von Klipstein et al., 2020), which would lead to a total of 590 parameters. Reliably estimating such complex models requires a minimum number of observations that is often not realistic to achieve in idiographic research designs. Simulation studies indicate that the length of commonly obtained ESM time series in psychiatric settings lead to networks with low sensitivity, potentially leaving relationships undetected (Mansueto et al., 2020).

9.1.2.3 Limitation 3: Practical considerations and technical skills

Statistical skills to estimate and interpret personalized networks are not routinely taught in training programs for health care psychologists. This barrier makes it hard for these models to be used by practitioners directly, and would require additional statistical consultation and collaboration with researchers. Although such collaborations may be desirable because they stimulate interdisciplinary exchange, they are also time-intensive and might therefore hamper implementation.

9.2 A formal integration of case formulation and personalized networks

The main objective of this chapter is to address these limitations by presenting an approach that formally integrates case formulation with personalized network estimation, and to offer an intuitive and user-friendly tool to apply the presented approach in clinical practice. We introduce the *Prior Elicitation Module for Idiographic System Estimation* (PREMISE) as a first step towards a systematic incorporation of clinical considerations in estimating personalized networks.

The core idea of the PREMISE approach is to use an initial case formulation ("working hypothesis") as the fundament for further statistical modeling routines. The integration of such clinical information with technical estimation routines requires that the case formulation first needs to be translated into a computational model using mathematical equations, a process referred to as the *formalization* of the case formulation. A new line of literature highlights the benefits of such computational accounts of theories for psychological science generally (Borsboom, van der Maas, et al., 2021; Fried, 2020; Guest & Martin, 2021; Haslbeck, Ryan, et al., 2019; Robinaugh et al., 2019, 2021; van Rooij & Baggio, 2021), and also specifically for case formulations as an example of theories on the individual level, illustrated in detail in chapter 12 (Schiepek, 2003; Stoger-Schmidinger et al., 2016; Schiepek, Stoger-Schmidinger, et al., 2016; Schiepek, Aas, et al., 2016). Once formalized, it is possible to investigate the precise implications of a case formulation through computer simulations. This allows one to evaluate to what extent the simulated implications of the case formulation align with clinical observations and to investigate the effects of formalized interventions, see chapter 12. *Verbal* accounts of case formulations (and theories in general), on the other hand, tend to be rather imprecise in their specifications and are therefore fallible in terms of accurate predictions and intervention testing (Fried, 2020a).

The process of formalization is complicated and entails making many technical decisions. To increase accessibility, tutorial papers have been published that guide researchers in formalizing verbal theories (Smaldino, 2020; van Rooij & Blokpoel, 2020). In this chapter, we draw on principles of *prior elicitation* (O'Hagan, 2006, 2019; Stefan et al., 2020) as one approach to make the formalization of case formulations more accessible for clinical practice. Prior elicitation refers to "*the process of extracting expert knowledge about some unknown quantity or quantities, and formulating that information as a probability distribution*" (O'Hagan, 2006). The experts, in our case clinician and client, can formalize case formulations without specifying probability distributions themselves, circumventing the technical limitation of implementing formalization techniques in clinical practice. Prior literature focused on similar approaches to eliciting perceived relationships, for instance using *Perceived Causal Relations* (Deserno et al., 2020; Frewen et al., 2012), and *Perceived Symptom Relations* (Schumacher et al., 2021). These approaches have been extended to the idiographic context, referred to as *Perceived Causal Problem Networks* (Klintwall et al., 2021). Furthermore, there are new approaches that use ESM data to quantitatively assess relationships within functional analysis (Scholten et al., 2021).

Figure 9.2 schematically illustrates differences in the process between the *agnostic estimation* of personalized networks (online version: highlighted in gray, print version: highlighted in light gray), and the *estimation with PREMISE* (online version: highlighted in cyan, print version: highlighted in dark gray). In both approaches, items are established in collaboration with the client (paths A and E in Figure 9.2). The core difference lies in the way these approaches estimate relationships between the ESM items: Whereas agnostic estimation calculates relationships directly from ESM data in a data-driven manner (paths B and C), estimation with PREMISE formalizes an initial working hypothesis via prior elicitation (path F). This clinical prior model is then subsequently updated using ESM data (paths G and H). Finally, the resulting networks of both approaches can then be used to inform case formulation (paths D and I).

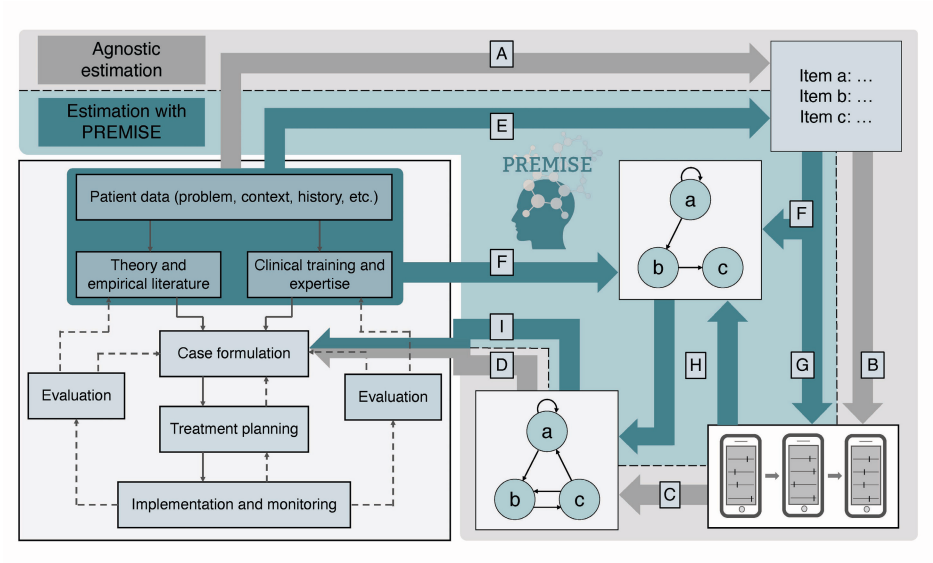


Figure 9.2. Relating two different approaches to estimating personalized networks to the process model of constructing case formulations proposed by Page and Stritzke (2014). In the *agnostic estimation*, ESM items are derived from client data, theory, literature and clinical expertise and training (A). Once items are established, the client collects data in their daily life (B), which can subsequently be used to calculate personalized networks (C). Such networks can stimulate conversations between client and clinician, and inform the construction of case formulations (D). In the estimation with PREMISE (the Prior Elicitation Module for Idiographic System Estimation), ESM items are also first derived from the client data (E). In contrast to the agnostic approach, however, PREMISE formalizes prior beliefs regarding the relationships between items, based on client data, theory, literature, and clinical expertise (F). Once data is collected (G), these clinical networks can then systematically be updated (H) via Bayesian inference. The resulting network can be used to inform case formulation (I).

9.3 The Prior Elicitation Module for Idiographic System Estimation (PREMISE)

In the following, we introduce a first step towards implementing the approach outlined in the previous section. In its current implementation, PREMISE extracts expert information on linear relationships between the selected ESM items via prior elicitation. Depending on the processes of interest, expert information can be extracted for temporal or contemporaneous relationships (Epskamp et al., 2018). The extracted information is then used as so-called *informative prior,* representing the perceived distributions of putative relationships, for the subsequent estimation of a Bayesian VAR model. Doing so allows one to systematically integrate clinical considerations with further statistical modeling. Once ESM data have been collected, the priors can be updated using Bayesian inference. This entails shifting the clinical prior model (i.e., the prior probability distributions derived via prior elicitation in PREMISE) according to the pattern found in the data.

Two principles are important here: First, the more data points are used, the more the initial estimates will shift towards the signal in the data. This means that if only little data are available, the

updated model will be largely based on the initial specification of the clinician and client, whereas with the number of observations increasing, the model will more and more converge to the effects driven by the data. Second, prior information can be assigned weights which determine how much data is required to override the prior information. This means that strong priors (i.e., priors with a narrow distribution) will take more data to be ruled out as compared to weak priors (i.e., priors with a wide distribution).

9.4 Clinical example: client with obsessive-compulsive disorder

To illustrate the principles of PREMISE, we describe the data and case formulation of a 31-year old client diagnosed with obsessive-compulsive disorder. In this example, the clinical prior were derived from a verbal client report, and the models have been estimated with different amounts of available data (i.e., after 2 weeks and 4 weeks), mimicking the updating of personalized networks during bi-weekly therapy sessions. Another example on eating disorders is presented in chapter 10.

9.4.1 Methods

Data on personalized ESM items have been collected three times a day over a period of almost one year, starting in 2017. During this period, the client followed a cognitive-behavioral therapeutic program, which included exposure therapy with response prevention. Data collection was exempted from formal ethical assessment (METc 2015/140). For a more extensive description of the dataset, see the paper by Bringmann and colleagues (2020).

9.4.1.1 Formalizing initial case formulations via PREMISE

During the initial stages of therapy, clinician and client discussed a working hypothesis regarding interaction and maintenance of symptoms. The client reported the following: "*Having intrusions (a), I can encourage (b) myself that they are harmless. This is something I must be able to do myself, independent of others. I can keep doing this but it exhausts me, and it becomes less and less effective, until I come to the point where I can no longer hold on to what I am telling myself. I become increasingly sad and hopeless (c). Passing this certain threshold, I panic and become extremely afraid to lose control (d) over myself.*" The clinician additionally observed that once this fear of losing control became unmanageable, the client usually *contacted (e)* their "safe persons" at the hospital and asked for admission, which made them feel safe from acting out on their intrusions. Other than their reaching out to safe contacts, the client showed no behavioral compulsions. The absence of other overt behavioral compulsions is the result of previous (thus partly effective) intensive cognitive behavioral treatments.

Using this report, we constructed a prior network based on the five ESM items (translated from Dutch): (a) *intrusions* ("How credible are the intrusions?"), (b) *encourage* ("I can encourage myself."), (c) *sad* ("I feel sad, useless, meaningless."), (d) *control* ("I am afraid of losing control."), and (e) *contact* ("Have you thought frequently about contacting 'safe' persons?"). The structure of the established prior network can be seen in the top left panel of Figure 9.3. As the reported process unfold relatively fast, and therefore likely occur within assessments, contemporaneous networks are the more appropriate choice (Epskamp, van Borkulo, et al., 2018).

9.4.1.2 Sampling observations and data preparation

Over the course of one year, the client experienced several relapses. In this example, we therefore focus on a sample of the data that did not coincide with a relapse period, because the VAR model assumes that its parameters do not change over time, an assumption referred to as *stationarity*. For this example, we selected four weeks worth of data collected between June 1st 2017 and June 28th 2017. We estimated a temporal model using the *psychonetrics* package (Epskamp, 2020d), and proceeded to use the residuals as observations for the estimation of contemporaneous networks. Prior to estimating the networks, we conducted several pre-processing steps that are common for time series analyses. For details on pre-processing and statistical estimation, see the supplementary R code[34].

9.4.1.3 Estimation with PREMISE versus agnostic estimation

The prior network structure derived from the client report served as a formalized working hypothesis that was systematically updated in two steps. This resulted in three networks: First, the prior network based on the client report (without ESM data), second, the updated network after two weeks (23 data points; 39 scheduled assessments), and third, the updated network after four weeks (53 data points; 84 scheduled assessments) of data collection. We will refer to these three networks as the *PREMISE networks*. Additionally, we estimated networks without the report-derived prior information, which we will refer to as the *agnostic networks*. These networks are representative for the VAR-based network models that are estimated without clinical input, and serve as a comparison point between the two approaches.

As is common in the field of undirected networks, edges represent the partial correlation structure of the variables (Epskamp, Waldorp, et al., 2018). Here, we used the STAN implementation in R (Stan Development Team, 2022) to model the variance-covariance matrix of the residuals via an inverse-wishart distribution[35]. In the PREMISE approach, we used the case formulation network matrix as informative prior for the inverse-wishart distribution (the so-called *scale matrix*). Furthermore, the degrees of freedom of the inverse-wishart distribution, here set to 30, determine how strongly the prior matrix will be weighed in during the updating process, with larger degrees of freedom centering more probability mass around the prior. In the agnostic approach, we used an uninformative prior set-up further described in a paper by Schuurman and colleagues (2016). Edges were thresholded by only including them if the respective 95% credibility interval did not include zero (Jongerling et al., 2022).

9.4.1.4 Transparency and openness promotion (TOP)

The chapter follows level 2 of the TOP-guidelines on all fundamental aspects of research planning and reporting (i.e., chapter shares materials when legally and ethically permitted). We share all relevant computer code, and provide references that further describe the dataset, including research

34 The R-code and STAN model are accessible in an OSF repository: https://osf.io/dguaj/.

35 Networks are based on the standardized precision matrix, rather than the variance-covariance matrix. The former represents partial correlations, and can be computed by taking the inverse of the variance-covariance matrix, followed by standardization.

material specifications (Bringmann et al., 2020). The example analyses in this chapter were not pre-registered.

9.4.2 Results

All networks are visualized using the *qgraph* package in R (Epskamp et al., 2012), and can be seen in Figure 9.3. The goal of the PREMISE estimation (top row) is to investigate changes to an initially established prior network (the "case formulation network"), which may advance the understanding of the client's psychopathology. In this example, updating the model with two weeks worth of ESM data removes one edge (*control – sad*), but includes additional edges (*control – intrusions*; *sad – intrusions*; *encourage – contact* [negative]). After four weeks, further edges are removed (*intrusions – contact* [negative]; *encourage – contact* [negative]; *sad – intrusions*; *sad – encourage*). In this updated model, the client experiences worries about losing control when intrusions are currently very credible. In turn, they think about contacting the "safe" persons, which makes them feel increasingly sad, useless, and worthless. At the same time, they manage to regulate the credibility of intrusions through self-encouragement.

The agnostic network, on the other hand, is more sparse and misses links specified in the case formulation. For example, in the agnostic approach, the relationship between them being worried about losing control and thinking about contacting the "safe" persons is only detected after four, but not after two weeks. This is most likely because there is not enough evidence (data) yet to establish this relationship after two weeks. In the PREMISE network, this relationship is part of the case formulation network, and is therefore retained throughout the updating process. Furthermore, other features relevant to the case formulation cannot be found in the agnostic network, such as the client's ability to decrease the credibility of intrusions through self-encouragement. Generally, it is important to note that (unexpected) modifications need to be interpreted with caution. These could also arise due to artifacts of the timing of ESM assessment (i.e., there are effects but they are not captured by the assessment, see the discussion), or unmeasured variables that are obscuring effects.

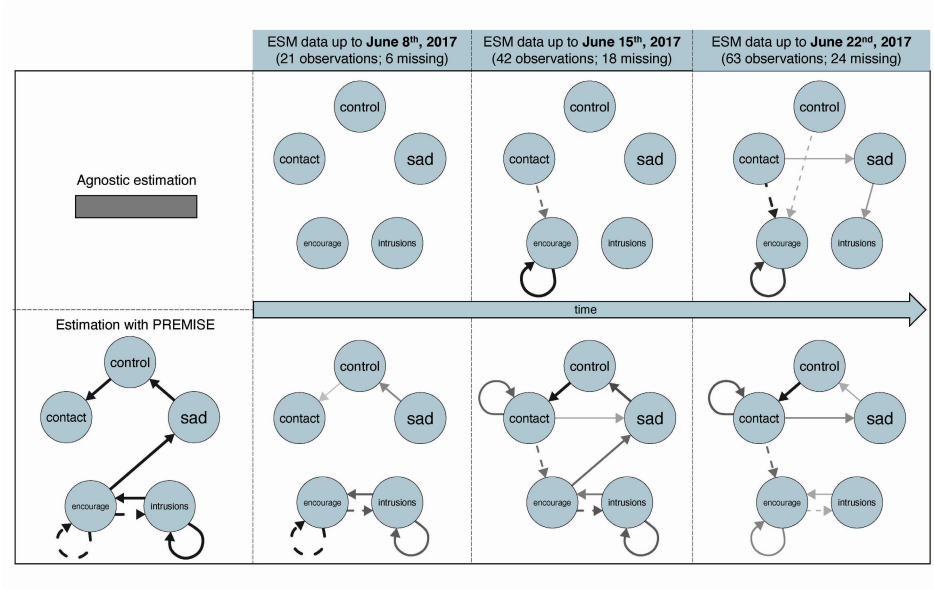


Figure 9.3. Results of contemporaneous networks based on the PREMISE (Prior Elicitation Module for Idiographic System Estimation) approach, using the case formulation network as informative prior (top row), and the agnostic approach, using a default uninformative prior (bottom row). Solid edges denote positive relationships, dashed edges denote negative relationships. The thickness of each edge corresponds to strength of the relationship.

9.5 Discussion

In this chapter, we contrasted different ways in which personalized networks can be used to inform case formulations. We discussed that current approaches to estimating personalized networks are primarily data-driven ("agnostic") and thus lack options to systematically incorporate clinically relevant information, result in models with low sensitivity, and require a level of technical expertise that might hamper clinical implementation. Based on these considerations, we proposed that a formal integration of case formulation and personalized networks, in combination with an intuitive user-interface, could advance clinical utility and implementation. In the following, we provide future directions on how the PREMISE approach can be used to advance our understanding of an individual's psychopathology, and different considerations for implementing it in practice.

9.5.1 Using PREMISE to gain insight into the client's psychopathology

One main question in the context of the PREMISE approach pertains to what we can learn from discrepancies between the clinical prior model and the statistical model based on ESM data. It is unclear at present which of these models better represent the *ground truth* of the client's personalized systems. Bayesian inference conceptualizes the strength of evidence as the amount of information that points towards a certain effect; the more we learn about the client (i.e., more data), the stronger the evidence for the presence or absence of certain symptom relations. As such, in the context of PREMISE, the ground truth reflects a (hypothetical) model that is based on the maximum

amount of data that can be collected within a stationary time unit (in the case of the classic VAR model). If a personalized model then veers away from the prior model in the updating process, this can be attributed to (a) the learning about new aspects of a client's psychopathology that were previously unknown, (b) a mismatch between the type of prior information that is specified and the assumptions of the statistical model that are imposed (e.g., if prior edges reflect a different time scale compared to the ESM sampling scheme, or if prior edges represent non-linear relationships but are applied to a linear model), or both.

It is impossible to infer which of these two explanations can account for discrepancies between prior and posterior model by merely *observing* them in PREMISE. However, behavioral and thought experiments (e.g., Waller, 2009) may help to identify the source of discrepancies, and therefore investigate if changes in the model indeed reflect new insights into the client's psychopathology. If changes to the initial model seem inappropriate or unreasonable following these experiments, clinician and client may discuss different aspects to the ESM data collection, such as changes to the sampling scheme or the inclusion or exclusion of items. Of note, both outcomes help us to learn more about the individual's psychopathology, either by *directly* providing insight into their experienced symptom relations (explanation a), or by *indirectly* pointing towards changes in the research design that may in turn reveal more valid inferences in the future (explanation b).

To give a clinical example, suppose a clinician and their client establish a positive relationship between them *staying in bed* and experiencing *depressed mood* in the prior model, but the updated model does not contain this relationship. Given theory and experience, this seems surprising, and clinician and client therefore decide to manipulate this pathway in a small experiment: The client is instructed to purposefully stay in bed versus get out of bed on different days, and to specifically monitor the effects on depressed mood throughout the day. If outcomes of this experiment support the pathway *staying in bed – depressed mood*, changes to the sampling scheme should be discussed (ruling out explanation a, support for explanation b). In this example, *depressed mood* potentially operates at a different time scale compared to *staying in bed*, which can only be assessed once a day, and changing the sampling frequency for this variable would therefore not solve the problem. An alternative could be to collect data on related variables that can vary throughout the day, such as feeling tired.

The fact that clinical and statistical predictions may differ, and, indeed, compete with one another (Meehl, 1954), does not mean that one model is *generally* preferable over the other. The different assessment strategies discussed in this chapter have their unique benefits: The *clinical prior models* can be established relatively quickly (about 22.7 minutes were needed for a comparable method by Klintwall et al., 2021), because they are based on a combination of readily available information, such as clinical literature, reported client experiences, and clinical training (Page & Stritzke, 2014). Furthermore, the process of establishing a prior model as collaborative effort between clinician and client may also stimulate a more active discussion on symptom relations compared to solely examining statistical output (see also section *On the Importance of Collaboration* below). As such, clinical models may be preferable in the *initial stages* of data collection when insufficient ESM data are available, because they provide an intuitive framework to efficiently formalize symptom relations. The *statistical models*, on the other hand, provide particular benefits in the exploration of symptom relations (von Klipstein et al., 2020; Rodebaugh et al., 2020) that may have been missed (or overestimated) in the prior model. They are therefore valuable especially in *later stages*

when more ESM data are available, allowing new evidence to suggest potential modifications to the clinical model. The PREMISE approach ties together these unique benefits in a systematic way using Bayesian inference. We hypothesize that these models therefore result in more actionable insights for clinical practice compared to either model alone, because they systematically balance clinical judgment with new evidence.

9.5.2 On the importance of collaboration

Although there are no gold standards, the case formulation approach to CBT emphasizes the importance of collaboration between clinician(s) and client (Kuyken et al., 2009; Persons, 2012). Nomothetic theories and treatment guidelines are usually the starting point of a case formulation, but the ultimate goal is to extrapolate an idiographic model by integrating these theories interactively with clinical expertise, observations, and client experience (Zuidersma et al., 2020). This approach has further benefits, for example in regard to compliance and the therapeutic relationship. Specifically in the context of PREMISE, another benefit to collaboration is the fact that interactive reasoning (*explorative talk*) has been found to improve judgment over individual results (Mercier & Sperber, 2018; Resnick et al., 1993; Wegerif et al., 1999). We, therefore, suggest that PREMISE should be used as a tool to aid interactive reasoning about symptom relations that should involve both clinician and client. PREMISE may help to make the process of interactive reasoning explicit by formalizing expertise and experiences into a prior model that can flexibly be integrated with ESM data.

9.5.3 Choosing a statistical model for PREMISE

The PREMISE approach is not tied to the specific elicitation method (i.e., estimates on temporal or contemporaneous relationships) or statistical model (i.e., Bayesian VAR) used in this chapter. As such, it is important to distinguish the *general approach* as highlighted in Figure 9.2 from the current *statistical implementation* of PREMISE. The key idea of PREMISE as an approach to establishing personalized models survives issues of the specific statistical model because these can be replaced by other implementations, should they offer a more intuitive and valid elicitation of clinical prior information. The VAR model currently takes a prominent role in the literature of personalized networks (Bringmann, 2021), which is why we opted for including it in PREMISE. Other statistical models can be used that are simpler or more sophisticated, which impacts how nuanced and intuitive the implications of the model are.

We see three criteria that are relevant to evaluate the utility of a statistical model for implementation in PREMISE: (a) Can the model describe relevant clinical phenomena?, (b) does the model contain quantities that can intuitively assessed via prior elicitation?, and (c) can the model provide actionable insights relevant for psychotherapy? Below we discuss these points in regard to the current implementation and alternative models.

9.5.3.1 Capturing relevant clinical phenomena

A common criticism of VAR-based networks is that they rely on strong and potentially unfeasible assumptions, such as *stationarity*, i.e., the properties of the time series do not change over time. Generally, it is advisable to specifically examine the collected data in light of the research question

and related modeling goals. For example, stationarity can be investigated by visualizing the time series, by performing formal tests such as the Dickey-Fuller test (Dickey & Fuller, 1979), and by employing change point detection algorithms (Aminikhanghahi & Cook, 2017). Although specific deviations from assumptions can be accounted for by transformations (e.g., removing time-related trends; see chapter 3), some research questions explicitly aim at understanding mechanisms related to change, for example modeling the effects of interventions. In such cases, it may be possible to use the VAR model, including the priors discussed in this chapter, for data collected *within stationary time periods* (e.g., prior to the start of an intervention), however, the VAR model does not allow one to model shifts between disorder states typically following interventions (Henry et al., 2020).

In addition to non-stationarity, the VAR model should not be used to answer research questions that aim at capturing higher-order interactions between variables from different levels (Haslbeck, Ryan, et al., 2019), or dynamics between variables that operate at time scales different to the frequency at which ESM is administered (Haslbeck & Ryan, 2021). More sophisticated modeling approaches may be better at capturing these clinical phenomena (Bringmann, 2021; Haslbeck, Ryan, et al., 2019), such as nonlinear (Schiepek, Aas, et al., 2016; Schiepek et al., 2017; Scholler et al., 2019), time-varying (Haslbeck, Bringmann, et al., 2021), and continuous time series models (Driver et al., 2017; Ryan et al., 2018; Ryan & Hamaker, 2020).

9.5.3.2 Intuitive prior elicitation

Prior elicitation techniques infer probability distributions based on quantities that can intuitively be provided by an expert. As such, prior elicitation can benefit clinical implementation because clinician and client do not need to specify technical aspects of statistical models themselves. On the other hand, bias can arise when the elicitation technique imposes additional assumptions which do not align with the expert's intuition. This could be the case if the model in question is too technically advanced.

In the current implementation of PREMISE, clinicians specify estimates for temporal or contemporaneous relationship. It is currently unclear if the specifications provided by the clinician indeed align with the assumptions and specifications of the VAR model. For example, at this moment we do not know whether the elicited quantity is understood by the clinician to be a marginal effect or a conditional effect, whether clinicians consider the specified time lag when indicating a relationship, or how to precisely specify distributions for edges that are not indicated by the clinician. Bias may be reduced if less sophisticated–but more intuitive–approaches such as means or marginal correlations between items are used. These could inform case formulation in a more basic yet potentially very insightful manner. The more sophisticated models discussed in the previous section, on the other hand, would potentially require assessment of quantities that are not very intuitive for the expert, and therefore could be a potential source of bias in establishing prior distributions. Another approach to reducing potential bias in more complex models could be to ask for concrete estimates on the item-level (i.e., symptom scores), rather than estimates on the parameter-level (i.e., edges between symptoms). In the current implementation, we opted for eliciting information on the parameter-level, as this aligns conceptually well with the process of establishing case formulations, where clinician and client discuss dynamic relationships (i.e., parameters) between the different items.

9.5.3.3 Actionable insights for psychotherapy

All statistical models–no matter their level of sophistication–are "wrong" in that they are incomplete *approximations* of reality (Meehl, 1990), and one statistical model is not necessarily more useful than another one simply because it features more sophistication in its modeling approach. The utility of a model for clinical inference is also determined by its ability to provide *actionable insights* (Fried, 2020b) for psychotherapy, which means that simpler, more abstract models could be at least equally meaningful if they qualify as useful thinking tools for clinical practice. Indeed, VAR-based networks have been suggested to serve as a first step towards informing case formulation in an exploratory fashion (von Klipstein et al., 2020). Future research should aim to investigate what model would indeed provide the most useful, intuitive, and actionable insights for case formulations and treatment selection, for instance through focus groups and utility studies.

9.5.4 Clinical implementation via sequential case designs

One of the core aims of PREMISE is to advance the implementation of personalized networks in clinical practice, by embedding the statistical estimation into the context of case formulations. However, integrating personalized models with clinical reasoning is only one aspect relevant for implementation. Another aspect is that these models should be seamlessly integrated with the therapeutic process, answering questions such as "When should we update our networks with ESM data?", and "What can we learn about the individual's psychopathology?". As these questions are inherently idiographic and answers will differ from client to client, they are best addressed using case designs.

In the future, we propose that PREMISE should be implemented using sequential case designs, such as within-person adaptations of the leapfrog design (Blackwell et al., 2019). The leapfrog design compares the efficacy of interventions against the currently most effective treatment (or a waitlist or other control condition, if no treatment has been established yet), by quantifying the evidence of improvement via Bayes Factors. If interventions derived from previously established personalized networks do not lead to substantive improvements (anymore), this could be a sign that the networks should be updated with new ESM data, which in turn may result in a shift in intervention targets and new knowledge on the individual's pathology. In the future, we hope that such implementations will lead to a more systematic dialogue between assessment, statistical modeling, personalized therapy, and the advancement of understanding the individual's psychopathology.

9.6 Conclusion

Formally integrating case formulation and personalized networks could potentially help overcoming current problems in personalized models, such as inaccurate estimation of networks and a disconnect with clinical theory, expertise, and practice. If combined with an intuitive tool for prior elicitation, this approach has promise to bring the benefits of personalized models into clinical practice. Future research should aim to investigate which statistical models are best suited for this approach, work towards providing concrete practical recommendations for implementation, and test if resulting networks can indeed improve therapy outcomes as evaluated by clinicians and clients.

CHAPTER 10

COMBINING CASE FORMULATIONS AND LONGITUDINAL DATA TO ESTIMATE NETWORKS FOREATING DISORDERS

—

CHAPTER 10

Abstract

Eating disorders are serious psychiatric illnesses with treatments ineffective for about 50% of individuals due to high heterogeneity of symptom presentation even within the same diagnoses, a lack of personalized treatments to address this heterogeneity, and the fact that clinicians are left to rely upon their own judgment to decide how to personalize treatment. Idiographic (personalized) networks can be estimated from ecological momentary assessment data, and have been used to investigate central symptoms, which are theorized to be fruitful treatment targets. However, both efficacy of treatment target selection and implementation with 'real world' clinicians could be maximized if clinician input is integrated into such networks. An emerging line of research is therefore proposing to integrate case formulations and statistical routines, tying together the benefits from clinical expertise as well as client experience and idiographic networks. The current pilot compares personalized treatment implications from different approaches to constructing idiographic networks. For two clients with a diagnosis of anorexia nervosa, we compared idiographic networks 1) based on the case formulation from clinician and client, 2) estimated from client ESM data (the current default in the literature), and 3) based on a combination of case formulation and client ESM data networks, drawing on informative priors in Bayesian inference. Centrality-based treatment recommendations differed to varying extent between these approaches for different clients. We discuss implications from these findings, as well as how these models may inform clinical practice by pairing evidence-based treatments with identified treatment targets.

This chapter has been adapted from: Burger, J., Ralph-Nearman, C., & Levinson, C. A. (2022). Integrating clinician and client case conceptualization with momentary assessment data to construct idiographic networks: Moving toward personalized treatment for eating disorders. *Behaviour Research and Therapy*, *159*, 104221.

10.1 Introduction

Eating disorders are severe and chronic psychiatric illnesses, associated with high morbidity, mortality, and societal and personal impairment (e.g., Deloitte Access Economics, 2020). Eating disorders carry one of the highest mortality rates amongst psychiatric illnesses. Anorexia nervosa (AN), in particular, has the second highest mortality rate of any psychiatric illness and is estimated to cost the US alone in one year 64.7 billion dollars in economic costs and 326.5 billion in loss of well-being (Deloitte Access Economics, 2020). Treatments for eating disorders are subpar, with gold-standard treatments (Cognitive-Behavioral Therapy Enhanced; Family Based Therapy) for both adolescents and adults leading to remission in approximately 50% of cases (Chesney et al., 2014; Deloitte Access Economics, 2020; Walsh et al., 2021). These low response rates have led to a push for new and improved treatments for these deadly illnesses.

10.1.1 Heterogeneity of eating disorders

Part of the reason that gold-standard treatments may not work for ~50% of individuals is because of the high heterogeneity present in eating disorders (e.g., Steinhausen, 2009). Recent research has shown that even for individuals with the same diagnosis, symptoms presentations are significantly different (Levinson, Vanzhula, et al., 2018). For example, one individual with AN may present to treatment with restriction, fears of weight gain, depression, and excessive exercise, while another individual with the same AN diagnosis may have symptoms of fasting, binge eating, worry, and low self-worth. Indeed, research shows that while about half of individuals with eating disorders have symptoms characterized by shape and weight concerns, the other half do not (Levinson et al., 2022). Furthermore, most (>50%) individuals with an eating disorder are given a diagnosis of other specified feeding and eating disorder (OSFED), which is essentially a catch-all diagnosis for any eating disorder that does not neatly fit into a diagnostic category (Riesco et al., 2018). As such, researchers are attempting to develop personalized eating disorder treatments that can address such heterogeneity using evidence-based methods.

10.1.2 Network analysis and treatment personalization

One way in which clinical researchers have begun to build personalized treatments is through the lens of network theory (Levinson et al., 2021, in press). Network theory proposes that psychiatric illnesses manifest and maintain themselves through dynamic symptom interactions (e.g., Borsboom & Cramer, 2013). For example, an eating disorder might develop from the symptom fear of weight gain, which directly leads to restrictive behaviors meant to alleviate this fear, which then leads to other symptoms of an eating disorder, such as binge eating and purging. Multiple cross-sectional datasets have been used to illustrate the application of network theory to data, termed network analysis, showing how the structure of eating disorders might be made of statistical relationships between symptoms, with specific *central* (i.e., or most important) symptoms theorized to be logical intervention targets, as central symptoms are hypothesized to be the symptoms that have the most impact on all other symptoms in the network (e.g., Borsboom & Cramer, 2013). Supporting this hypothesis, multiple empirical examples have shown that central symptoms are predictive of short

and long term outcomes both in eating disorders (Elliott et al., 2020; Levinson & Williams, 2020; Olatunji et al., 2018) and related illnesses, such as depression (Levinson et al., 2017).

Cross-sectional networks provide useful insights into statistical relations among symptoms on the between-subjects level. In contrast, aspects of treatment *personalization* may best be studied at the level of the individual, termed idiographic research (Fisher et al., 2018; Molenaar, 2004; Zuidersma et al., 2020). Therefore, moving beyond cross-sectional models, idiographic network analysis uses intensive longitudinal data (typically collected via mobile applications, such as via the Experience Sampling Method [ESM; Mestdagh & Dejonckheere, 2021; Myin-Germeys et al., 2018; Shiffman et al., 2008]) to model the structure of pathology for one individual. This type of analysis is extremely important as it allows for identification of specific central symptoms that might maintain pathology in each specific person (Epskamp, van Borkulo, et al., 2018; Jordan et al., 2020). It has thus been proposed that central symptoms might be matched to evidence-based treatments and that intervention on central symptoms should weaken the overall illness (Levinson et al., in press). Recent work in the eating disorders has demonstrated the clinical utility of idiographic networks (Levinson et al., 2017; Levinson, Vanzhula, et al., 2018, 2020), as well as demonstrated that network-informed personalized treatment is effective at reducing eating disorder severity, eating disorder behaviors, and related anxiety and depression (Levinson et al., in press).

10.1.3 Integrating case formulation and idiographic networks

Importantly, to date, all investigations using idiographic network analysis to inform treatment and psychoeducation have relied solely on intensive longitudinal (i.e., ecological momentary assessment) data collected from clients to build a personalized network and subsequently select treatment targets. However, implementation of such a data-based personalized treatment depends on clinical researchers' ability to bridge the research-practice gap (Bansal et al., 2012), such that clinicians perceive value in data-based personalization. Recent work demonstrated barriers for implementing idiographic networks in clinical practice. For example, as discussed in chapter 12, clinicians and their clients may see limited utility in idiographic networks if these models fail to reflect clinical expertise, intuition, and theory, or client experience. Indeed, a pilot study (Frumkin et al., 2020) showed that clinicians may not see the added benefit of idiographic networks over the use of existing clinical models (e.g., case formulation; Kuyken et al., 2009; Persons, 2012). These reservations have led to the conception of a new approach that aims to *integrate* rather than *contrast* case formulation and statistical network models, see chapter 9 and Scholten et al., (2021). This line of research proposes to use clinical expertise and client experience to construct networks of perceived relationships (Deserno et al., 2020; Klintwall et al., 2021; Schumacher et al., 2021). In a subsequent step, networks based on case formulations can be "updated" via Bayesian inference using ESM data collected by the client. This type of additional modeling strategy has potential not only to bridge the research-practice gap, but also to conceptualize more effective models by integrating multiple perspectives into a data-based algorithm, ultimately improving both the uptake and efficacy of data-based personalized treatments.

10.1.4 Current study

This study aims to systematically integrate clinical information, from both clients and clinicians, with data-driven network estimation routines. Our ultimate goal is to illustrate how these different models can be used to help bridge the research-practice gap, and therefore foster the implementation of idiographic networks in clinical practice, see also the previous chapter. In this chapter, we provide initial empirical evidence for the utility of clinically-informed networks, and showcase implications of this approach in regards to eating disorders. Specifically, we aim to investigate the extent to which centrality-based treatment recommendations differ in idiographic networks that are based on different sources of information: 1) Networks derived from clinician and client case formulations, 2) networks estimated from ESM data provided by the client (the current default in the idiographic network literature), and 3) networks combining case formulation and ESM data, estimated via Bayesian inference using informative priors.

Due to the novelty of the approach, the focus of this study is exploratory, and we do not have specific hypotheses for the extent to which these networks differ from one another. Generally we did expect, based on prior literature showing clinician-judgment may result in different formulations (Pisetsky et al., 2019; Waller, 2016), that there would be differences between models, and that these differences may vary depending on the client and clinician. We also showcase how treatment could be informed by each of the types of models we present.

10.2 Methods

10.2.1 Participants

Participants included in this study were two white self-identified women diagnosed with AN restricting subtype, who were 42 and 31 years-of-age (Client A and B, respectively). Diagnoses were given based on two structured clinical interviews, the Structured Clinical Interview for DSM-5 (SCID-5; First et al., 2015) and the Eating Disorder Diagnostic Inventory (Stice et al., 2008) by a highly trained Master's or PhD clinical psychology student. All diagnoses were double-checked independently by two additional highly trained Master's or PhD clinical psychology students. There were 100% diagnostic agreement between raters for both clients in this present study.

10.2.2 Materials

10.2.2.1 Diagnostic screening measures

10.2.2.1.1 Structured Clinical Interview for DSM-5

The SCID-5 (First et al., 2015) is a semi-structured interview to assess eating disorder diagnoses, severity, subtype, and course. The current study used the eating disorder modules to determine eating disorder diagnoses.

10.2.2.1.2 Eating Disorder Diagnostic Interview (EDDI)

The EDDI (Stice et al., 2008) is a semi-structured interview to assess eating disorder symptoms over the prior year, which has been shown to have acceptable inter-rater reliability for the eating

disorder diagnoses, validity and internal consistency (Stice et al., 2008). This interview was used to confirm diagnoses.

10.2.2.1.3 MINI International Neuropsychiatric Interview

The suicidality, mania/hypomania, and psychosis modules on the semi-structured MINI International Neuropsychiatric Interview (Sheehan et al., 1998), which has been shown to have excellent reliability and good validity, was used for exclusion criteria, which were active suicidal intent, psychosis, or mania.

10.2.2.1.4 Experience Sampling Method (ESM)

The ESM consisted of 56 items spanning a broad range of eating disorder and related (e.g., anxiety, worry) symptoms in which participants rated how intensely they were currently experiencing each symptom (e.g., *I am terrified of gaining weight.*) from 0 (least ever) to 100 (most ever); see Levinson et al. (2021) for full ESM list and more information on the ESM battery.

10.2.3 Procedure

10.2.3.1 Data collection

All procedures were approved by the University of Louisville Human Subjects Protection Program (18.0622). Participants were recruited through advertisements across the United States, completed informed consent, and then were screened for inclusion/exclusion criteria by completing three semi-structured teleconference interviews (see diagnostic screening measures). Individuals were eligible to continue participating in the 10-session online treatment study if they had a current eating disorder diagnosis, and were not actively suicidal, psychotic, or manic. After completing baseline surveys, at the initial teleconference meeting, participants were trained on how to begin an ESM through their mobile phone five times a day for 15 days (75 timepoints; see below for more information) that would be used to guide their treatment plan. Each survey took three to five minutes to complete.

10.2.3.2 Clinician and client case formulation

After the client left session 2, which was a non-structured clinical interview, therapists created a clinician-informed network of symptoms by listing the top eight symptoms they perceived as most important for maintaining the clients' eating disorder (e.g., *fear of weight gain*, *excessive exercise*, *restricting food*, etc.), how these symptoms connect (e.g., *fear of weight gain AND restricting food, fear of weight gain AND excessive exercise, restricting food AND excessive exercise*), and to what strength of connection from 0 to 100 (for an example see Supplement D, Figure S10.1). At session 3, clients worked with their clinician to design their own network of symptoms from the client perspective. Similarly to the clinician-informed network, clients decided what they thought were their top eight most important symptoms, and then how they perceived that their symptoms connected to one another with arrows, as well as how strong they believed the connection was between symptoms from 0 (not at all) to 100 (the strongest). If client had a difficult time numerically rating the symp-

tom strength, then they could instead rate them as *weak*, *medium*, or *strong*. These sessions were completed before beginning any type of treatment.

10.2.3.3 Variable selection

An important question in the context of network analysis is which variables should be included in the networks, see also chapter 5 and Fried and Cramer (2017). This is because network estimates consist of multivariate (partial) effects, and the set of variables therefore has a strong impact on the structure of the network itself: A node that is central within one given set of variables may be at the periphery of a different set of variables. We based the selection of variables on both, theoretical and data-driven criteria, considering the size of the network, topological overlap of items, as well as the variability and stationarity of the time series.

10.2.3.3.1 Statistical power and network size

Statistical network models are highly parameterized models if many variables are used. Therefore, large numbers of observations are usually required to arrive at accurate estimates of connections in the network. Variable selection, and more specifically the number of variables to be included, is therefore directly linked to the question of statistical power and the accuracy of network estimates. To our knowledge, there is currently no principled way to estimate required sample sizes for idiographic networks. Preliminary simulation studies recommend that given the characteristics of data commonly obtained in clinical practice, no more than six variables should be included for network estimation (Mansueto et al., 2022). To this end, for both clients, we first selected items that have been specified by either the clinician or client, and given that all of these items showed sufficient variability (see below), selected the six items with the highest mean of this subset. The means and selected items for both clients are visualized in Figure 10.1.

10.2.3.3.2 Topological overlap

In the context of mental disorders, networks consist of nodes that represent psychological constructs, such as body dissatisfaction, fear of making mistakes, and drive for thinness. In contrast to "real entities", such as individuals or objects, these psychological constructs are not always clearly separable. In the network literature, such conceptual similarities between constructs are referred to as *topological overlap* (Fried & Cramer, 2017). If constructs within one network are not clearly separable, edge estimates conflate the relationship between two nodes with their conceptual similarity. To address this problem, there are algorithms developed to detect redundancies of nodes, such as the goldbricker algorithm (Jones, 2018). These algorithms, however, can currently only be used in the context of data-driven network estimation. In this paper, we derive clinical networks from case formulations, and these networks will likely show different redundancy patterns compared to their corresponding data-driven networks. For example, the algorithm detected two redundancies for client A (*cognitive restraint – drive for thinness*, and *cognitive restraint – body dissatisfaction*). The case formulation network for this client, however, defines these combinations as non-redundant, i.e., the nodes have unique relationships with other nodes in the network. For client B, no item redundancies were identified. One of the main aims of this paper is to compare structures across the different types of networks, and it is therefore important that the networks consist of the same set

of items. To this end, we focused on the items selected by clinician and client as defined above (see *Statistical power and network size*), which reflects a combination of theory- and data-driven variable selection. Identifying node redundancies for the combination of different network structures needs to be investigated in future research.

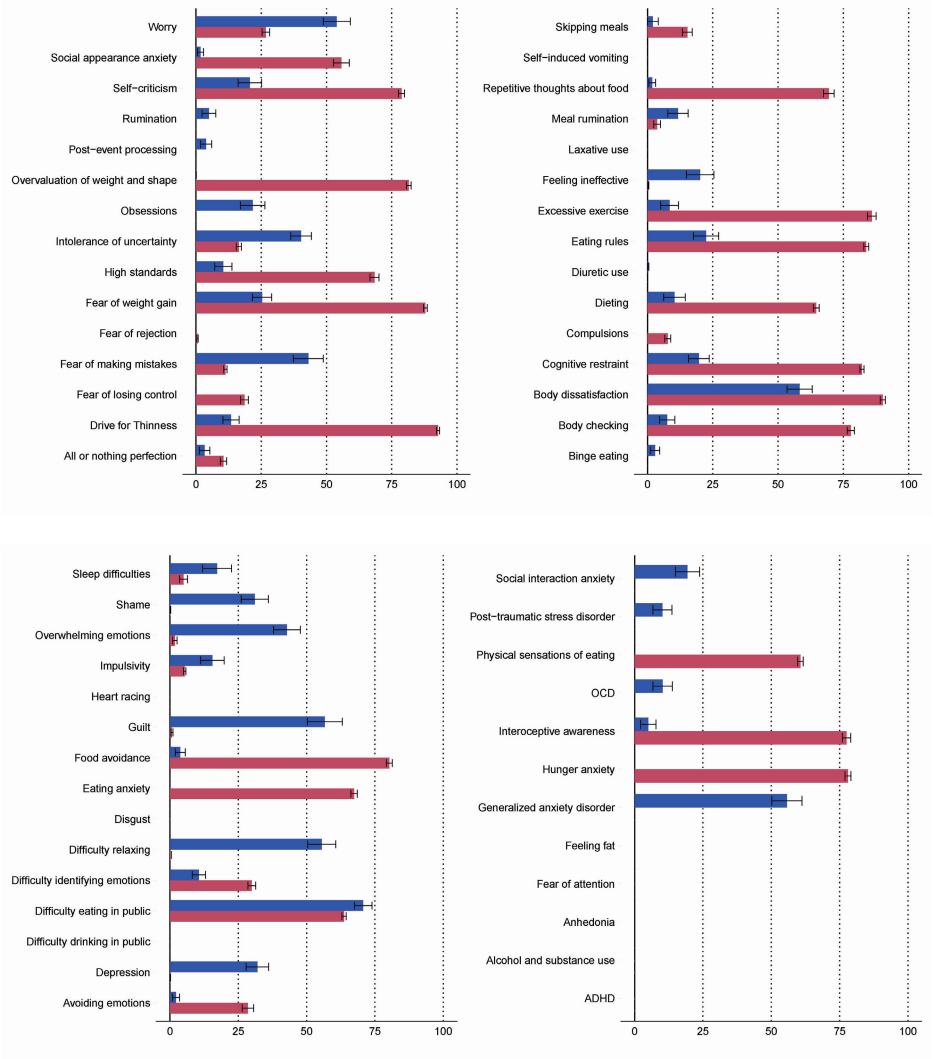


Figure 10.1. Mean scores and 95% confidence intervals for all ESM items and clients (red = client A, blue = client B).

10.2.3.3.3 Stationarity assumption

The models used in this paper assume stationarity, i.e., that the characteristics of the time series do not change over time. Data in this study were collected *prior* to the intervention and over a relatively short period of time (see *Data collection*), which contributes to the feasibility of the stationarity

assumption. In addition, we investigated stationarity visually (see Figure 10.2a and 10.2b), and applied the Augmented Dickey-Fuller (ADF) test to check for non-stationarity (Dickey & Fuller, 1979b). This test investigates the Null hypothesis that a unit root is present, indicating that the time series is not stationary. For client A, all time series were stationary according to the ADF test ($p_{drivethin} = .030$; $p_{bodydiss} = .013$; $p_{fowg} = .032$; $p_{excexercse} = .027$; $p_{cogrestraint} = .010$; $p_{overvalwtshape} = .018$). For client B, the majority of time series was stationary according to the ADF test ($p_{fearmstkes} = .025$; $p_{depression} = .022$; $p_{eatrules} = .043$; $p_{selfcrit} = .013$; $p_{cogrestraint} = .015$; $p_{bodycheck} = .01$;), but there were also time series that are not stationary according to the ADF test ($p_{avoidemo} = .070$; $p_{overwhlmemo} = .085$; $p_{gad} = .065$; $p_{bodydiss} = .405$). Nevertheless, we opted for not excluding them because doing so would result in removing all but two edges from the clinical network (e.g., most of the connections in the case formulation network are with *avoiding emotions*).

10.2.3.3.4 Variability of time series

Network estimation is most commonly based on the analysis of covariance structures, and it is therefore important that items included in networks show sufficient variability around their means. Brose and Ram (2012) suggest two rules of thumb for investigating variability of time series, a) a maximum of 80% of scores within one person and variable being identical, and b) a minimum standard deviation of 10% of the scale. We applied both criteria to all time series to check if they showed sufficient variability. The maximum amount of identical scores for any value and time series was 11.42% for client A (*fear of weight gain*, score of 90; *excessive exercise* score of 100), and 54.93% for client B (*avoiding emotions*, score of 0). For client A, some of the time series had comparably small standard deviations, ranging from SD = 4.39 (*drive for thinness*) to SD = 12.39 (*excessive exercise*). For client B, except for *avoiding emotions* (SD = 8.58), the time series showed variability well above the rule of thumb defined above, ranging from SD = 19.19 (*body checking*) to SD = 38.69 (*fear of making mistakes*). A reason for the somewhat smaller variability of *avoiding emotions* is that this item showed floor effect tendencies, with about half of the assessments (54.93%) marked at 0. This is important to keep in mind, as estimates with variables showing floor effects may be biased (Klipstein et al., 2022). However, as mentioned above, excluding the item *avoiding emotions* would lead to removing most of the edges in the case formulation network for client B, and we therefore opted to include this item in the analyses.

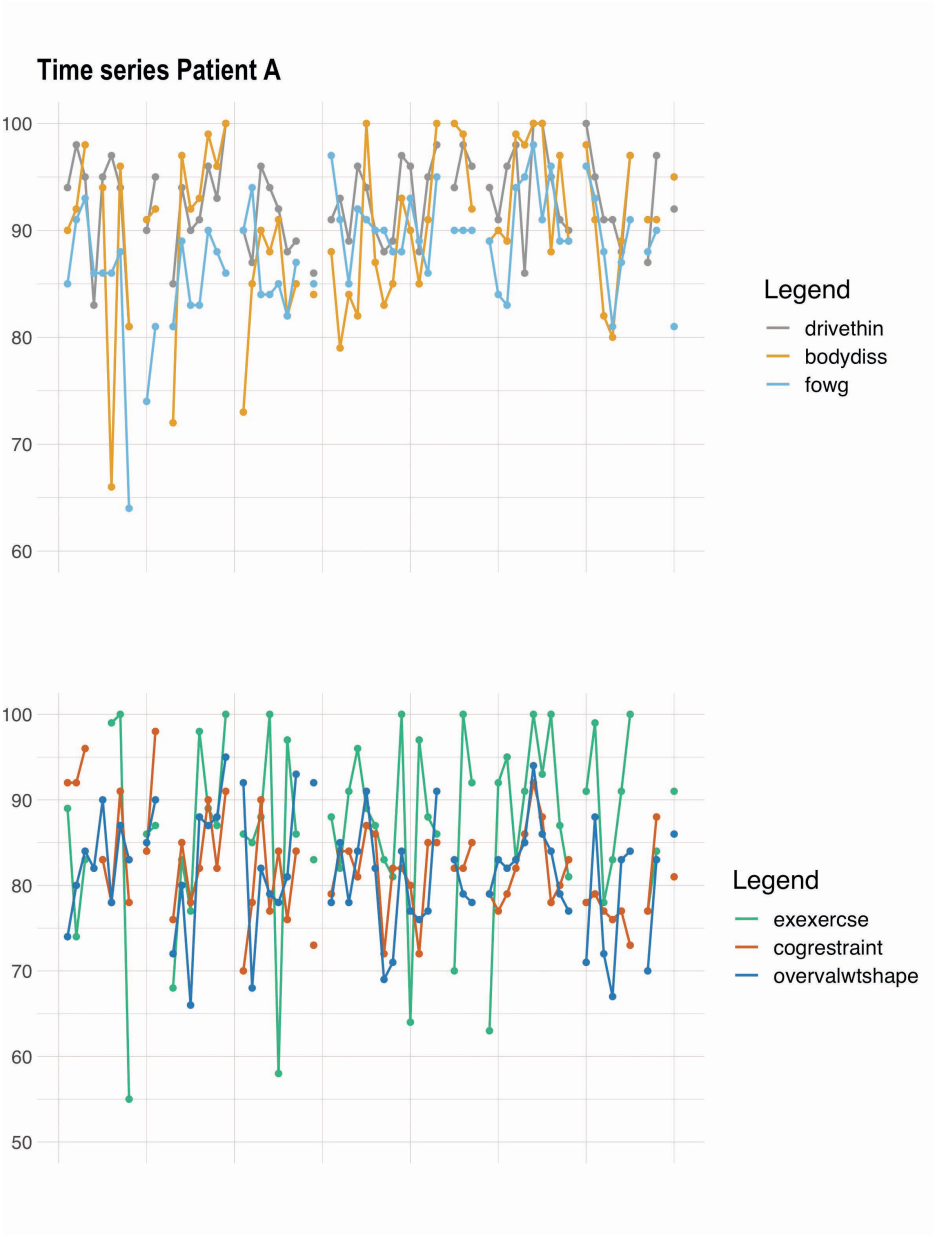


Figure 10.2a. Time series of all ESM items for client A (labels: *drivethin* = drive for thinness; *bodydiss* = body dissatisfaction; *fowg* = fear of weight gain; *exexercse* = excessive exercise; cogrestraint = cognitive restraint; *overvalwtshape* = overvaluation of weight and shape).

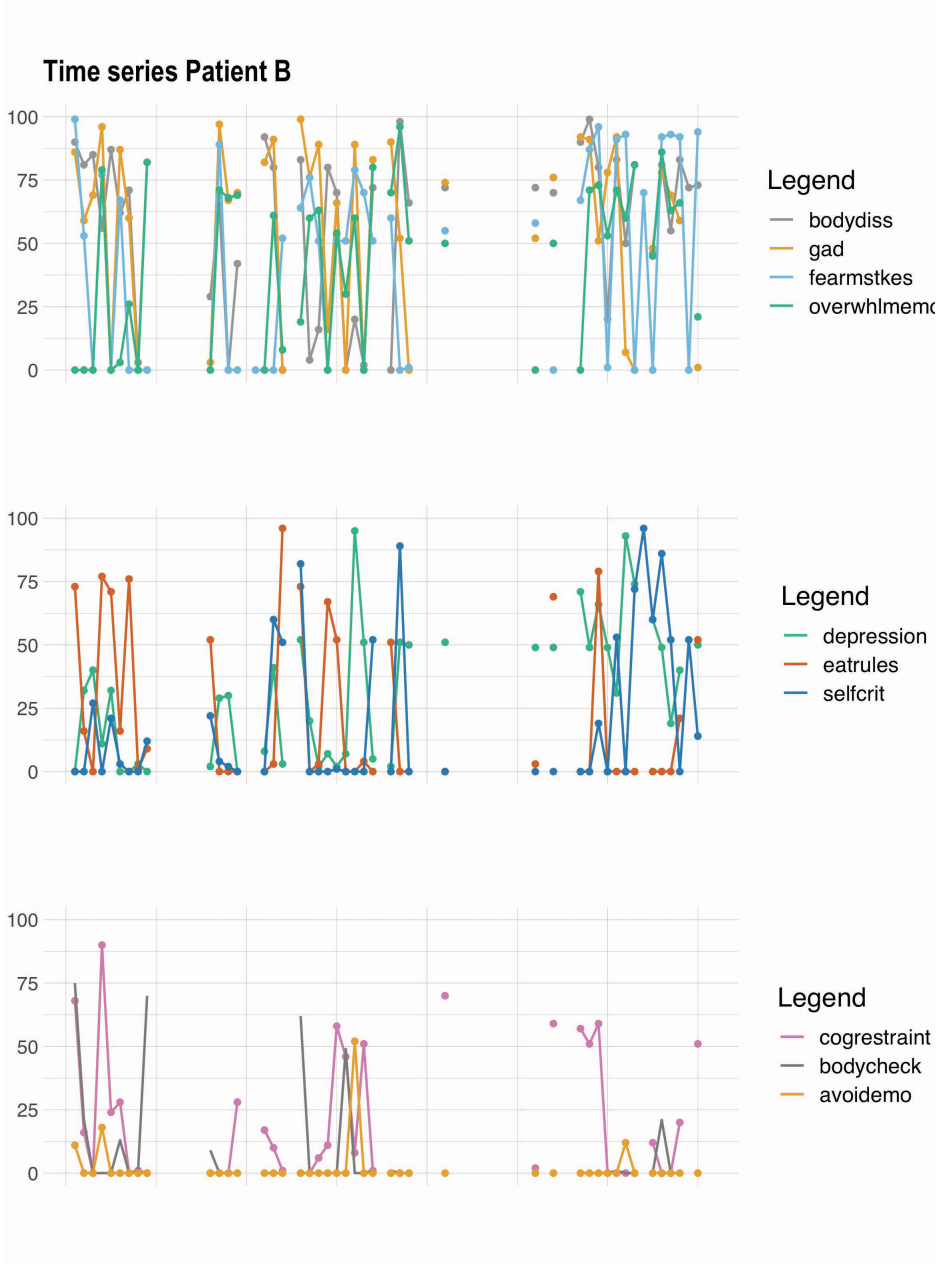


Figure 10.2b. Time series of all ESM items for client B (labels: *bodydiss* = body dissatisfaction; *gad* = generalized anxiety disorder; *fearmstkes* = fear of making mistakes; *overwhlmemo* = overwhelming emotions; *depression* = depression; *eatrules* = eating rules; *selfcrit* = self-criticism; *cogrestraint* = cognitive restraint; *bodycheck* = body checking; *avoidemo* = avoiding emotions).

10.2.4 Network estimation

For both clients, we used the clinician and client case formulation, as well as momentary assessment data, to construct three types of networks. We investigated the case formulation networks separately for the clinician and client, as well as their combination. In the main text of this paper, we use the average of the separately reported clinician and client case formulations (the "combined" case formulation), and provide the separate networks in the appendix (Supplement D, Figure S10.2 and S10.3).

10.2.4.1 Clinician and client case formulation (case formulation network)

First, we constructed a network based on the clinician and client case formulation (in the following referred to as the *case formulation network*), by combining (averaging) the respective clinician and client case formulations, and subsequently making the relations undirected. We opted for using undirected relationships, which align with the notion of contemporaneous ("instantaneous") networks. We did so because (undirected) contemporaneous networks are considered to better capture rapid processes commonly found in psychopathology as compared to (directed) temporal networks, which are restricted to fixed time intervals (Epskamp, van Borkulo, et al., 2018). Other considerations to choosing priors and a statistical model are addressed in the discussion section.

10.2.4.2 Client ESM data network (ESM network)

Second, we estimated a network from the ESM data provided by the client (in the following referred to as the *ESM network*). The estimation was based on Bayesian partial correlation networks using a default prior without regularization (Schuurman et al., 2016; Williams, 2021). For more details, see the R-code in the online supplemental materials.[36]

10.2.4.3 Integration case formulation and ESM data (PREMISE network)

Third, we estimated a network integrating case formulation and ESM data using the PREMISE approach (Prior Elicitation Module for Idiographic System Estimation, see chapter 9; in the following referred to as the *PREMISE-network*). The PREMISE approach estimates Bayesian partial correlation networks, which has been proposed as a fruitful alternative to frequentist estimation (Williams, 2021; Williams & Mulder, 2020). A particular advantage is the explicit incorporation of available prior information, which allows to formally implement the approach outlined in this paper.

10.2.4.4 Network analysis and comparison

Network models can quantify the relative influence of specific nodes in relation to the overall network, referred to as the *centrality* of a node (Opsahl et al., 2010). In the applied network literature, centrality metrics have been used to generate hypotheses on optimal treatment targets (Elliott et al., 2020; Levinson et al., 2017; Rodebaugh et al., 2018). Here, we focus on one-step *Expected Influence (EI)* as a measure of centrality, which is defined as the sum of the weighted edges connected to a

36 The R-script to run all analyses, including the model specifications discussed here, can be found in an OSF repository: https://osf.io/wu7qk/ .

given node (Robinaugh et al., 2016). For both clients, we compared the network-implied EI of the items as proxy for personalized treatment recommendations.

10.2.4.5 Model specifications and uncertainty of estimates

In this section, we briefly describe the main model specification settings. Note that this section is intended for the reader interested in technical details, and for researchers interested in applying the methodology in their own designs. The reader primarily interested in the results of this particular study may wish to skip these sections and continue with *Network analysis and comparison*.[37]

10.2.4.6 Model estimation

For the estimation of the ESM and PREMISE networks, we first estimated a Vector-Autoregressive lag-1 model (VAR; Epskamp et al., 2018; Wild et al., 2010) that accounts for the temporal dependencies in the data. We then modeled the scaled and centered residuals of the VAR estimation as multivariate normal distributions (MVN). The MVN distribution consists of two parameters, the location vector μ and the covariance matrix Σ, the latter encoding relationships between variables. In the network literature, the inverse of Σ is commonly used to construct partial correlation networks (Epskamp, Waldorp, et al., 2018). The difference in the estimation of the ESM and PREMISE network lies in the specific way Σ is modeled: For the ESM networks, we used an inverse-wishart distribution to model Σ, with parameter-settings resembling an uninformative prior (i.e., with scale matrix set to be the identity matrix, and the degrees of freedom set to the number of nodes in the network; Schuurman et al., 2016). For the PREMISE networks, we used the case formulation network as the scale matrix in an inverse-wishart prior distribution, with degrees of freedom set to 30. Increasing the degrees of freedom puts stronger prior probability on the scale matrix. There is potential to inform the setting for the degrees of freedom by the confidence in prior estimates, which we address in more detail in the discussion section.

10.2.4.7 Edge and centrality accuracy

A particular advantage of Bayesian estimation is that the uncertainty of the estimated models can be directly obtained from the resulting posterior distributions (Jongerling et al., 2022). Network studies, such as the present paper, are often interested in centrality estimates, which are generated from previously established network structures (e.g., by summing incoming/outgoing relations). The quantification of uncertainty for metrics such as centrality is somewhat less straightforward compared to the uncertainty of edges, and can result in bias (Epskamp, Borsboom, et al., 2018). To counter this bias, a recent simulation study proposed a new approach to quantify uncertainty for centrality estimates, termed *post-processing shift estimation* (PPS-estimation; Jongerling et al., 2022). To assess the extent to which centrality scores of symptoms are indeed different from one another, we applied the PPS-estimation and checked if the 95% credibility intervals of the posterior distributions for the differences between each centrality score included 0, which may indicate that the difference in centrality between the two symptoms in question may be negligible.

---

37 The R-script to run all analyses, including the model specifications discussed here, can be found in an OSF repository: https://osf.io/wu7qk/ .

10.2.4.8 Software

All analyses have been conducted in R (R core team, 2013) on 03/29/22, using version 4.1.0. We used the *psychonetrics* package version 0.9 (Epskamp, 2020b) to estimate the GVAR models. We then used the STAN implementation *rstan* package version 2.26.6 (Stan Development Team, 2022) to model the multivariate normal distributions and construct the contemporaneous networks. Networks are visualized using the *qgraph* package version 1.6.9 (Epskamp et al., 2012), and edge and centrality uncertainty are visualized using *ggplot2* version 3.3.5 (Wickham, 2016).

10.3 Results

10.3.1 Descriptive statistics

Figure 10.1 shows the mean scores and 95% confidence intervals for all ESM items and clients.

10.3.1.1 Client A

Following the item selection criteria, the top six symptoms specified in the perceived network for client A are: *drive for thinness* ($M = 92.73$, $SD = 4.39$), *body dissatisfaction* ($M = 90.14$, $SD = 7.50$), *fear of weight gain* ($M = 87.88$, $SD = 5.58$), *excessive exercising* ($M = 85.92$, $SD = 12.39$), *cognitive restraint* ($M = 82.14$, $SD = 6.03$), and *overvaluation of weight and shape* ($M = 81.55$, $SD = 6.99$).

10.3.1.2 Client B

For client B, we had to extend the number of nodes from six to ten. This is because from the 10 symptoms that clinician and client used in the case formulation, the six symptoms with the highest mean were unrelated. In fact, only when all 10 symptoms were included did the network show any relations, as most of the symptoms were related to the items with the lowest mean, i.e., *cognitive restraint*, *body checking*, and *avoiding emotions*. Removing these items would result in an empty prior network, which is why we extended the number of nodes to 10 for client B. The included symptoms are, in descending order of the respective means, *body dissatisfaction* ($M = 58.25$, $SD = 31.86$), *generalized anxiety disorder* ($M = 55.70$, $SD = 36.09$), *fear of making mistakes* ($M = 43.04$, $SD = 38.69$), *overwhelming emotions* ($M = 42.79$, $SD = 32.06$), *depression* ($M = 31.98$, $SD = 26.91$), *eating rules* ($M = 22.42$, $SD = 31.73$), *self-criticism* ($M = 20.67$, $SD = 30.27$), *cognitive restraint* ($M = 19.65$, $SD = 25.85$), *body checking* ($M = 7.53$, $SD = 19.19$), and *avoiding emotions* ($M = 2.16$, $SD = 8.58$).

10.3.2 Network estimation and visualization

The networks for both clients are visualized in Figure 10.3a and 10.3b. In addition, we visualized the edge estimates and respective 95% and 50% credibility intervals and added the plots to the Supplementary materials (Supplement D, S10.4 – S10.9). Note that for the perceived networks (top panel of each accuracy plot), no credibility intervals can be computed, because clinician and client only provide point estimates for the perceived relations but no distributions. In this section, we report general characteristics of the networks and their accuracy as discussed in chapter 5, and in the next section we specifically focus on comparing the network-implied centrality scores.

10.3.2.1 Client A

For client A, edges in the case formulation network range from $r = .45$ (*fear of weight gain – cognitive restraint*) to $r = .95$ (*drive for thinness – fear of weight gain*; *drive for thinness – excessive exercise*; *fear of weight gain – overvaluation of weight and shape*). Edges in the ESM network range from $r = .29$ (*drive for thinness – cognitive restraint*) to $r = .43$ (*body dissatisfaction – cognitive restraint*). Edges in the PREMISE network range from $r = .23$ (*fear of weight gain – overvaluation of weight and shape; excessive exercise – overvaluation of weight and shape*) to $r = .46$ (*body dissatisfaction – cognitive restraint*).

10.3.2.2 Client B

For client B, edges in the case formulation network range from $r = .70$ (*self-criticism – cognitive restraint*; *body dissatisfaction – body checking*; *depression – avoiding emotions*) to $r = .95$ (*generalized anxiety disorder – avoiding emotions*). Edges in the ESM network range from $r = -.44$ (*self-criticism – cognitive restraint*) to $r = .58$ (*depression – avoiding emotions*). Edges in the PREMISE network range from $r = -.29$ (*body dissatisfaction – avoiding emotions*) to $r = .57$ (*depression – avoiding emotions*).

For all networks, the 95% credibility intervals showed, on average, relatively large overlap with one another. This means that we cannot be certain about the *specific rank-order* of edges (i.e., one edge being particularly stronger than another edge in the same network). However, we can still interpret the *overall structure* of the networks irrespective of their weight, as the edges are selected based on a pruning procedure (Jongerling et al., 2022) which only includes edges whose 95% credibility intervals do not include 0.

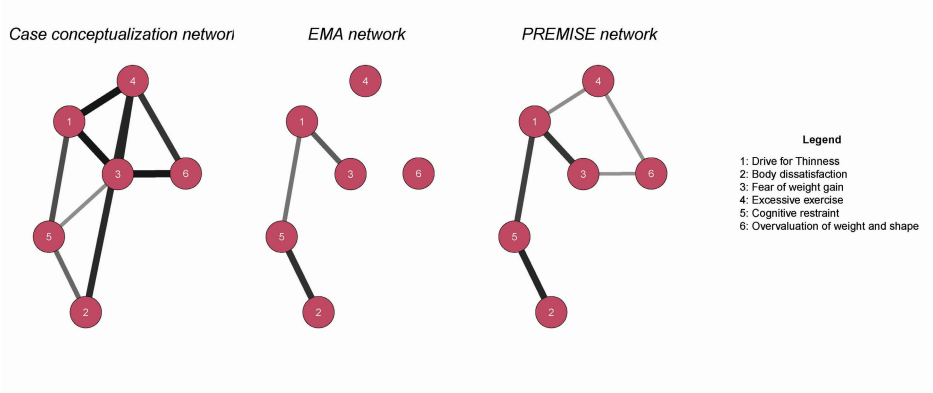


Figure 10.3a. Clinician and client case formulation (case formulation network; left), client ESM data network (ESM network, middle), and combined case formulation and ESM data network (PREMISE network, right) for *client A*.

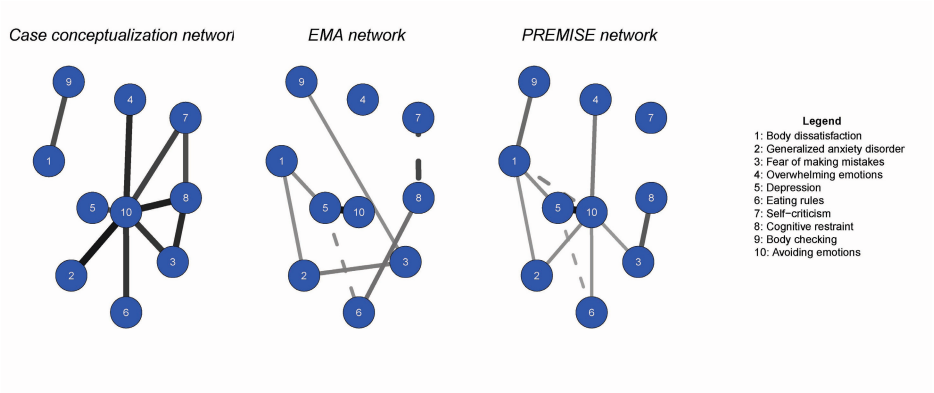


Figure 10.3b. Clinician and client case formulation (case formulation network; left), client ESM data network (ESM network, middle), and combined case formulation and ESM data network (PREMISE network, right) for *client B*.

10.3.3 Network inference: Centrality-based treatment recommendations

Figures 10.4a and 10.4b show a comparison of the network-implied centrality rank orders for the case formulation network, the ESM network, and the PREMISE network for both clients. When comparing centrality scores of symptoms, it is important to consider the width of the posterior distributions, which inform us about the uncertainty of the estimates. For the ESM and the PREMISE network, the left panels show point estimates, as well as 50% and 95% credibility intervals. In addition, the right panels indicate for each combination of symptoms if their respective centrality scores are meaningfully different from one another (i.e., if the 95% credibility interval of the posterior distribution of their difference score does not include 0). As with the edge accuracy plots, no posterior distributions, and therefore no credibility intervals can be computed for the case formulation network, because the centrality metrics are directly inferred from the provided point estimates. In the following, whenever we refer to a centrality *rank order* of the symptoms, we treat symptoms whose difference posterior distribution do not meaningfully differ as defined above as a tie. Based on these comparisons, we present a list of the most central symptoms for each client and network in Table 10.1.

10.3.3.1 Client A

For client A, the most central symptom in the case formulation network was *excessive exercising*, followed by *fear of weight gain*, and *drive for thinness*. In the ESM network, the symptoms *cognitive restraint*, *drive for thinness*, *body dissatisfaction*, and *fear of weight gain* were tied for the most central symptoms. *Overvaluation of weight and shape* was less central than *cognitive restraint* and *drive for thinness*, but not the remaining symptoms. *Excessive exercising* was less central than *cognitive restraint*, *drive for thinness*, and *body dissatisfaction*, but not *fear of weight gain* and *overvaluation of weight and shape*. In the PREMISE network, in contrast, the symptoms *drive for thinness*, *cognitive restraint*, and *fear of weight gain* were tied for the most central symptoms. *Excessive exercising* and *overvaluation of shape* were both less central than *drive for thinness* and *cognitive restraint*, but not

less central than the remaining symptoms. *Body dissatisfaction* was only less central than *drive for thinness*, but not less central than any of the remaining symptoms.

10.3.3.2 Client B

For client B, the most central symptom in the case formulation network was *avoiding emotions*, followed by *cognitive restraint* and *fear of making mistakes*. In the ESM network, none of the symptoms differed from one another in terms of centrality, except for *self-criticism*, which was less central than *generalized anxiety disorder*, *fear of making mistakes*, *depression*, *avoiding emotions*, and *body dissatisfaction*, but not less central than the remaining symptoms. In the PREMISE network, *avoiding emotions*, *fear of making mistakes*, and *depression* were tied for the most central symptom, however, out of the three, only *avoiding emotions* was more central than the remaining symptoms in the network.

Table 10.1. Most central symptoms for the case formulation networks, ESM networks, and PREMISE networks for both clients.

	Case formulation network[a]	ESM network	PREMISE network
Client A	Excessive exercise (C: 3, P: 2) Fear of weight gain (C: 2, P: 1) Drive for thinness (C: 1, P: *n.r.*)	Cognitive restraint[b] Drive for thinness[b] Body dissatisfaction[b] Fear of weight gain[b]	Drive for thinness[b] Cognitive restraint[b] Fear of weight gain[b]
Client B	Avoiding emotions (C: 2, P: *n.r.*) Cognitive restraint (C: 1, P: *n.r.*) Fear of making mistakes (C: 3, P: *n.r.*)	*No rank order*[c]	Avoiding emotions[b] Fear of making mistakes[b] Depression[b]

[a] For the case formulation network, we present the rank-order based on the average clinician-client network, and in brackets the rank-order of the symptom for both clinician (C) and client (P) separately. *n.r.* ("not ranked") indicates that the symptom was unconnected.

[b] Shows the rank order of the point estimates, however, the 95% credibility intervals of the posterior difference distributions included 0. The rank order of these symptoms should therefore be interpreted with caution.

[c] None of the centrality scores of any of the symptoms was meaningfully different from one another, according to the 95% credibility intervals of the posterior difference distributions, except for *self-criticism*, which was less central than most other symptoms.

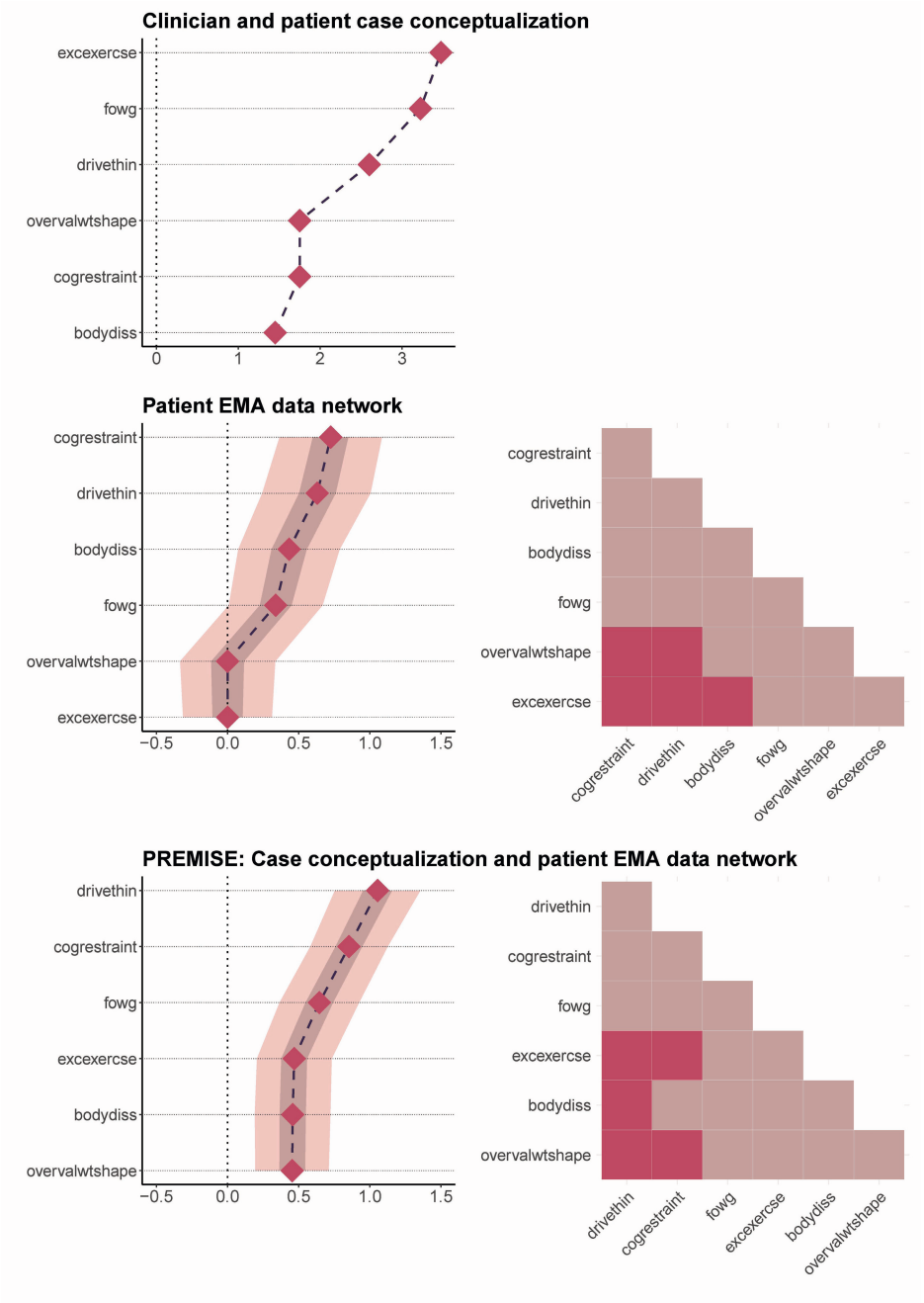


Figure 10.4a. *Client A:* Centrality scores (EI) for the case formulation network (top panel), as well as centrality scores and accuracy estimates based on 95% and 50% credibility intervals (left), and centrality difference test (right) for the ESM network (middle panel) and the PREMISE network (bottom panel). In the right panels, red boxes indicate meaningful differences in the centrality score of the two symptoms.

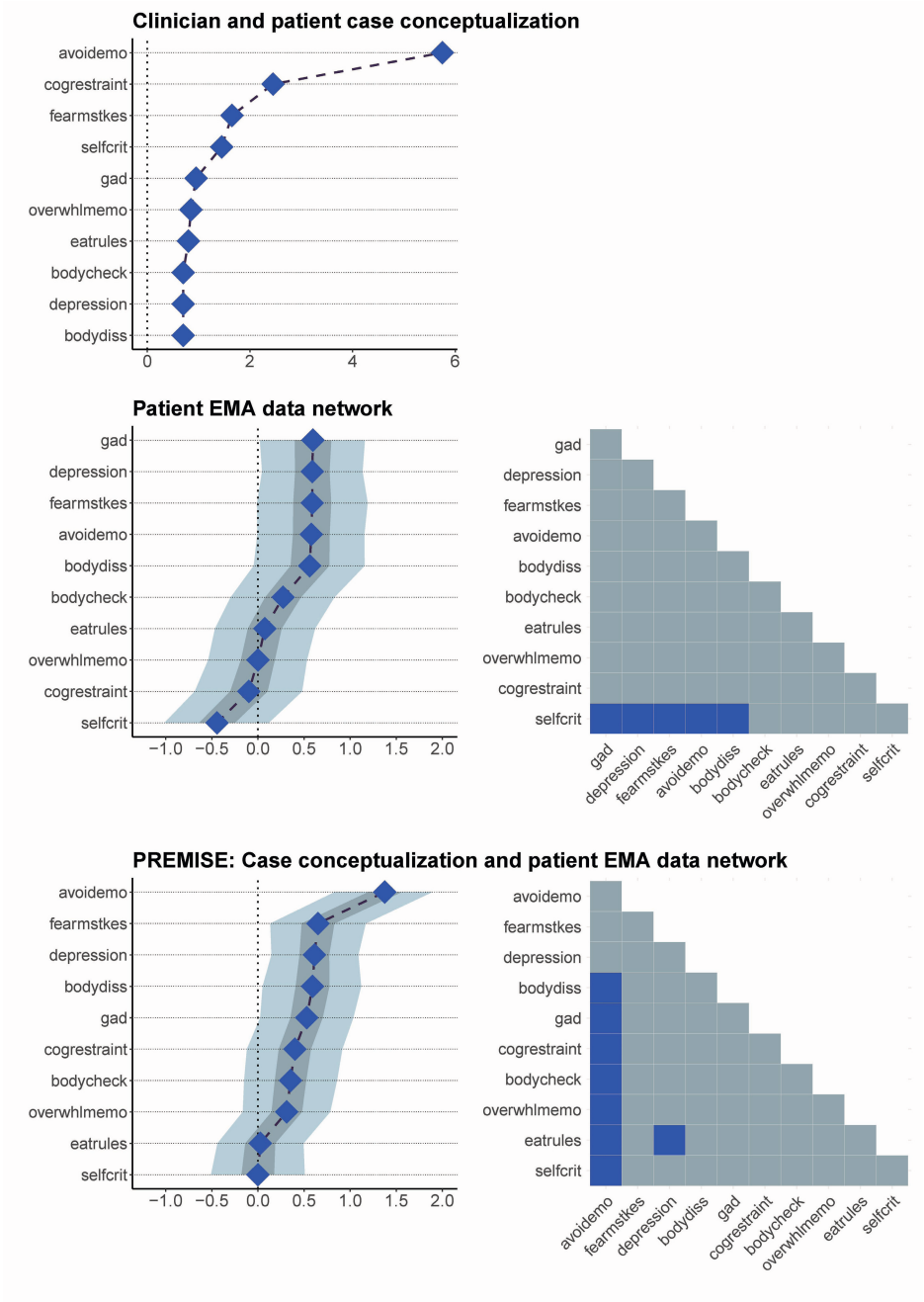


Figure 10.4b. *Client B:* Centrality scores (EI) for the case formulation network (top panel), as well as centrality scores and accuracy estimates based on 95% and 50% credibility intervals (left), and centrality difference test (right) for the ESM network (middle panel) and the PREMISE network (bottom panel). In the right panels, blue boxes indicate meaningful differences in the centrality score of the two symptoms.

10.3.4 Clinical treatment example

We provided two case examples of three different approaches (case formulation, ESM data only, combined case formulation and ESM) for arriving at idiographic network models. For client A, treatment recommendations would defer based on type of algorithm. For this case formulation, the top two targets were *excessive exercise* and *fear of weight gain*, which could be matched to evidence-based treatments such as CBT for reducing *excessive exercise* (Mathisen et al., 2018) and imaginal exposure for *fear of weight gain* (Levinson, Christian, et al., 2020). Alternatively, in the ESM network, the top two central symptoms were *cognitive restraint* and *drive for thinness*, both of which might be best addressed by Cognitive-Behavioral Therapy Enhanced (CBT-E), specifically modules on regular eating and thought challenging (Fairburn et al., 2003). Finally, in the combined network (case formulation plus ESM data) the network was very similar to the ESM network, with drive for thinness and cognitive restraint as the top two central symptoms, which would again lead to similar treatment recommendations of CBT-E modules. However, we should note that *fear of weight gain* was the third most central symptom, also replicating most central symptoms from the clinician network (minus the *excessive exercise* symptom). As such, dependent on the model, treatment modules and ordering would vary.

10.4 Discussion

Current treatments for eating disorders are subpar, with only about 50% of adults responding to evidence-based treatments (Chesney et al., 2014; Deloitte Access Economics, 2020; Walsh et al., 2021), and no treatments currently in existence for other specified feeding and eating disorders (the most common eating disorder) or for AN (Riesco et al., 2018). Part of the reason that treatment may not work effectively for a large subset of clients is that heterogeneity is extremely high in the eating disorders (e.g., Steinhausen, 2009). As such, personalized and evidence-based treatments are needed.

In this chapter, we compared different approaches to constructing and estimating personalized networks of eating disorder symptoms. We estimated networks based on clinician and client case formulations, networks estimated from client ESM data, and networks that combine case formulations and ESM data via Bayesian inference. Using two cases of clients with AN, we highlighted how using these different approaches can influence the results of subsequent centrality analyses, and therefore, potentially impact the choice of personalized treatment targets. We also demonstrated how clinicians might use each of these types of network to inform treatment selection. The current chapter shows how to build personalized networks from intensive longitudinal data (collected via ESM), and how these can be integrated with case formulations. This approach is especially important because the incorporation of both client and clinician data into the models has the potential to provide more effective algorithms than client data alone, and to help bridge the research-practice gap by encouraging clinician engagement with network models. Further, a specific benefit to *integrating* clinical and statistical models over using either of them alone is that the integration via Bayesian inference systematically weighs new evidence against the current case formulation. The systematic integration ties together the benefits of *clinical networks*, that are especially relevant in the early stages when insufficient ESM data are available to reliably estimate models, and the benefits of *statistical networks*, which may generate exploratory insight in later stages of the case formulation (von Klipstein et al.,

2020; for a detailed discussion on the advantages of integrating clinical and statistical networks, see chapter 9). The Bayesian updating routine can stimulate a dialogue and reveal discrepancies between prior and posterior models that may suggest behavior or thought experiments (see chapter 9).

Overall, the extent to which we observed discrepancies between which symptoms were most central for clients varied for the two cases. Specifically, for client A there was a somewhat large discrepancy in central symptoms, especially between the case formulation and ESM network, with clients and clinicians possibly overemphasizing the importance of excessive exercise behaviors. This difference is extremely interesting and may derive from the fact that traditional treatments for eating disorders very strongly emphasize problematic behaviors as key targets for intervention (e.g., Fairburn et al., 2003). However, recent research has suggested that cognitive-affective symptoms of eating disorders may be more important for the maintenance of active illness states (Levinson et al., 2021; Levinson, Vanzhula, et al., 2018), which is a shift in the way in which treatments might be built and delivered. Future research is needed to identify not only which type of algorithm is most effective and most easily accepted by clinicians, but also why some models may have more or less overlap and what that overlap might mean for effective treatment.

There are many possible future clinical implications from this research. The ability to derive a personalized algorithm that identifies symptoms to target in treatment can lead to evidence-based personalized treatment for eating disorders, as well as additional psychiatric illnesses. Crucially, these types of algorithms need input from both clinicians and clients and these data demonstrate how to create such a model and how that type of model can be used to pinpoint treatment targets that can be matched with existing or novel evidence-based treatment modules. We provide an example of how we could match treatment targets such as fear of weight gain and drive for thinness in the results. Future research can turn these types of algorithms into clinician-friendly software to make an easy-to-use guidance system for clinicians.

10.4.1 Limitations

There are some limitations of this research. First and foremost, there were only two participants' data presented in the case-series design. However, we want to strongly emphasize that while we did not include many participants, the amount of data per person was large and consisted of intensive longitudinal data and clinician and client perspective data. With a shift to more personalized types of treatment, clinical researchers must also shift their viewpoint from considering that the size of the dataset refers to the number of observations per person, rather than the number of participants. In fact, "truly" idiographic research is not necessarily concerned with identifying generalizable features across individuals, but rather a model that works for a given client, and should therefore focus on the length of time series and not the number of individuals in a study. If clinical researchers want to build truly personalized evidence-based treatments, we must first develop and test the types of algorithms presented in the current study. We need this type of research, which develops algorithms with the potential to personalize and improve treatment, to build truly evidence-based personalized treatments.

Second, for the construction of the case formulation and combined networks, we used the average of the clinician and the client perceived relations as a proxy for the *collaborative* case formulation. It could be argued, however, that establishing these case formulations as a true collaborative effort between clinician and client in conversation may yield more valid priors. For example, for client A,

for whom we observed large discrepancies between the case formulation network and the ESM/PREMISE network, there were also large differences between the individual clinician and client networks (see Supplement D, Figure S10.2). In fact, the most central symptom in the clinician's prior network was drive for thinness, which was also highly central in the ESM network. In turn, the client's prior network implied relatively high centrality for cognitive restraint, which was the most central symptom in the ESM network. While these important aspects get lost in the statistical averaging of the two networks, they could have been discussed and incorporated in a collaborative prior network. Indeed, it has been found that interactive reasoning ("explorative talk") improves judgment compared to individual results (chapter 9; Mercier & Sperber, 2018; Resnick et al., 1993). More specifically, as discussed in chapter 9, differences in the formulations between clinician and client could stimulate a dialogue and suggest thought and behavioral experiments to test relationships. Finally, such collaboration between clinician and client aligns with the principles of case formulation, and additionally have other benefits, for example positive effects on the therapeutic relationship (Persons, 2012).

Third, it is currently unknown which statistical model best aligns with the type of relationships that clinician and client specify in the case formulation networks. Some preliminary guidelines for selecting a statistical model are presented in chapter 9, for example by choosing a model that can capture the clinical phenomena of interest, and by asking for estimates of quantities that are intuitive for clinician and client. In this chapter we used the VAR model, which currently takes the most prominent role in the field of personalized networks, but is not without limitations (for an overview, see Bringmann, 2021; Haslbeck & Ryan, 2021). One particular limitation relates to the strong assumptions of the VAR model, such as stationarity. While we could show that the assumptions were largely met in the context of this data, there were some reasons for concern (the floor effects of *avoiding emotions* for client B), calling for caution in interpreting effects with this variable. Further, we currently do not know if the provided clinical information are better used for the estimation of temporal or contemporaneous networks. In this chapter, we opted for contemporaneous priors for two main reasons: First, most of the relationships between variables are better reflected on relatively short time scales. For example, it can be assumed that cognitive symptoms, such as self-criticism and body checking are interacting rather rapidly, and not on the lag-1 scales specified in the assessment of this data collection. Second, participants have undergone a training phase in which they were shown contemporaneous networks, and we therefore assumed that their estimates align with the notion of contemporaneous effects. However, there are also limitations to using contemporaneous networks, such as the fact that they do not only reflect contemporaneous effects but also model misspecification.

10.4.2 Future research

The approach used in this chapter implies new areas of research, as it is truly a crucial first step in the personalization of evidence-based treatment for eating disorders. First, a randomized controlled trial is needed to test which type of algorithm leads to the most effective and efficient treatment or if there are comparable results regardless of type of algorithm. Taking this reasoning to the individual level, single case designs could reveal that different clients may benefit from different personalization approaches discussed in this chapter. One hypothesis would be that client groups with strong insight into their own pathological processes, or disorder groups with strong theoretical background, may benefit more from the case formulation network or the PREMISE approach, as these put more emphasis on theory,

clinical expertise, and client experience in deriving treatment targets. Client (groups) with limited insight, or who only recently experienced symptom onset, and disorder groups with weaker theoretical background may in turn benefit more from a focus on data-driven modeling (i.e., the ESM networks presented in this chapter). We suggest that these approaches could be implemented in a sequential within-person design, where the treatment implied by one personalization approach (e.g., targeting the most central symptom in the case formulation network) is used until no further (or no satisfactory) improvement can be achieved. The clinician can then resort to other personalization approaches, for example targeting the most central symptom in the PREMISE or ESM network. Such designs can be built around Bayes factor criteria (e.g., in the leapfrog design; Blackwell et al., 2019), formally indicating when a treatment switch may be appropriate, i.e., when no satisfactory improvement is achieved compared to the status-quo (either the control condition or the most successful treatment up until that point). In the future, implementing such designs in clinical practice could inform the administration of personalized treatments in real-time.

Second, this type of personalization should be extended to additional forms of psychiatric illness. While eating disorders are a very relevant example, given the lack of effective treatments (Chesney et al., 2014; Deloitte Access Economics, 2020; Walsh et al., 2021), many psychiatric disorders are heterogeneous and have less than ideal treatment response, including but not limited to depression, post-traumatic stress disorder, and personality disorders (e.g., Hofmann et al., 2012). Of course, future research is also needed with larger sample sizes and with clinician-friendly software to test the ability to implement such algorithms in clinical practice. Future research integrating clinician input on using this type of model and how to best integrate these algorithms into clinical practice is needed.

Finally, there are several aspects of prior elicitation and statistical estimation that need further investigation. Next to the questions regarding the statistical models introduced above (see *Limitations*), one especially relevant question pertains to how strongly clinical priors should be weighed against ESM data. In this chapter, we model prior information via an inverse-wishart distribution, where the degrees of freedom reflect how strongly the model draws on the specified prior. Because we had no information on the confidence in the prior, we set the degrees of freedom to be 30, in line with the example analysis of chapter 9. Given the amount of data available in this study, setting the degrees of freedom to 30 led to a reasonable balance between prior and posterior model. In the future, these settings could also be informed by eliciting how confident clinician and client are in their priors. Future research should aim to develop anchors for setting these confidence estimates empirically, and incorporate confidence elicitation in the assessment of PREMISE.

10.5 Conclusion

In conclusion, we applied three different approaches for personalizing eating disorder treatment and demonstrated these with data of two clients. We also provided examples of how these models can be used to inform clinical practice by matching evidence-based treatments to identified treatment targets. Overall, we found there were some clients who had similar treatment targets, regardless of type of algorithm, whereas for other clients' treatment targets varied. Future research is needed to continue to expand upon these work in additional eating disorder clients and in additional client populations.

CHAPTER 11

LONGITUDINAL PERCEIVED CAUSAL RELATION NETWORKS

CHAPTER 11

Abstract

Personalized networks of psychological symptoms aim to advance therapy by identifying treatment targets for specific clients. Statistical relations in such networks can be estimated from intensive longitudinal data, but their causal interpretation is limited by strong statistical assumptions. An alternative is to create networks from client perceptions, which comes with other limitations such as retrospective bias. We introduce the Longitudinal Perceived Causal Relations (L-PCR) approach to address both these concerns. 20 participants screening positive for depression completed up to four weeks days of brief daily assessments of symptoms and perceived symptom interactions. Quality criteria of this new method are introduced via a bootstrapping algorithm, answering questions such as "Which symptoms should be included in networks?", "How many datapoints need to be collected to achieve stable networks?", and "Does the network change over time?". Accordingly, about 40% of respondents achieved stable networks and only few respondents exhibited network structure that changed during the assessment period. The method was time-efficient (on average 7.4 minutes per day), and well received. Overall, L-PCR addresses several of the prevailing issues found in statistical networks and therefore provides a clinically-meaningful method for personalization.

This chapter has been adapted from: Burger, J., Andikkhash, V., Jäger, N., Anderbro, T., Blanken, T., & Klintwall, L. (2022). A Novel Approach for Constructing Personalized Networks from Longitudinal Perceived Causal Relations. *PsyArXiv Preprint, under review.*

11.1 Introduction

In network theory (Borsboom, 2017) psychiatric disorders are conceptualized as psychological states that arise from the causal interplay of symptoms. Network theory holds particular promise for the field of mental health because it provides a framework for explaining clinical phenomena such as resilience, comorbidity between disorders, and heterogeneity within disorders. In recent years algorithms to estimate statistical network structures from empirical data have been developed (Borsboom, Deserno, et al., 2021; Borsboom & Cramer, 2013; Epskamp, Waldorp, et al., 2018) and applied across a broad range of psychopathologies (Robinaugh et al., 2020). Notably, while many of these studies are conducted at the *group* level, most of the concepts mentioned above are relating to the *individual* level. Therefore, one challenging aspect in the current landscape of network theory pertains to the construction of idiographic network structures (Bringmann, 2021; Epskamp, van Borkulo, et al., 2018). One way to create such idiographic networks is to estimate statistical models (see chapter 3, as well as Wright & Woods, 2020; Jordan et al., 2020c; Epskamp, van Borkulo, et al., 2018) from dense longitudinal assessments such as data collected via the Experience Sampling Method (ESM) data (Myin-Germeys et al., 2018; Shiffman et al., 2008). Such approaches are based on statistical assumptions (e.g. linearity and stationarity) and inferences are constrained by characteristics of the collected data (i.e., the chosen time scales in the context of ESM). As a consequence, statistical network models are heuristic and do not lend themselves towards a direct causal interpretation of systems (Dablander & Hinne, 2019; Pearl et al., 2016; Ryan et al., 2019).

An alternative to statistical estimation is to construct causally interpretable networks by eliciting information from either clinicians or clients via so-called *Perceived Causal Relations* (PCR) (Deserno et al., 2020; Frewen et al., 2012; Klintwall et al., 2021). In PCR assessments, the respondent rates the extent to which relevant symptoms are causally interrelated. From a clinical perspective, the notion of constructing causal networks from clinical knowledge and client experience aligns with the idea of case formulation (Kuyken et al., 2009; Persons, 2012) and process-based therapy (Hofmann & Hayes, 2019). In line with the rationale of these approaches, PCR networks may be used to identify highly influential symptoms or interactions, and thus guide selection of intervention targets.

11.1.1 Current approaches to assessing perceived causal relations

The first method to assess how clients perceive symptoms' influence on each other was developed by Frewen et al. (2012). In this method, participants are first asked to select which items were present in the previous month from a predetermined list of 40 symptoms. Participants then rate every combination of the selected items regarding the extent to which each item has influenced every other in the past month. This question is phrased "How much do you think your problems with X cause your problems with Y?" and is answered using a 0-10 Likert scale. Extending this work, Klintwall, Bellander, and Cervin (2023) developed the PECAN questionnaire (PErceived CAusal Networks), which aimed to be a simplified and more clinically-relevant PCR assessment. Similar to Frewen's method, participants are asked to select symptoms from a predefined list, although the list is only 26 items and the referenced timeframe is two weeks. The causal ratings are done slightly differently in that the PECAN method does not assess every possible combination. Instead, the questionnaire asks about each selected symptom and the respondent can select up to three of the other selected symptoms

as causes. If at least one symptom is selected, the respondent is asked to distribute percentages across these selected symptom(s) and an extra option labeled "other causes," which must total 100.

11.1.2 Current problems with assessments of perceived causal relations

The current ways to assess PCR face several challenges. A main problem is that, although clinicians rated the PECAN method as clinically useful, test-retest reliability is low (Klintwall et al., 2023 found an within-session reliability of 0.53; and unpublished data from the same group where assessments were made two weeks apart showed a retest reliability of .28). We see two reasons for this problem: First, the retrospective nature of the questionnaire, and second, asking about the strength of the causal relationship expressed as percentages. Both of these points could be addressed by using longitudinal (e.g., daily) assessment of PCR, which we will term L-PCR.

First, clients may find it hard to report causal relations retrospectively. Resulting errors may be especially prominent in traditional PCR methods given that clients are not only asked to remember which symptoms they experienced but also how these are causally linked. This issue might be addressed by using L-PCR assessments on a daily basis. It is widely assumed that data collected closer to the recall event increases ecological validity, and is used in other types of longitudinal assessment, such as in ESM (Myin-Germeys et al., 2018; Shiffman et al., 2008).

Second, traditional PCR assessments ask the client to quantify the perceived strength of a causal relationship, either by attributing percentages (Klintwall et al., 2023) or ratings on a 10-point Likert scale (Frewen et al., 2012). Such questions are remote from how clients usually talk about the way in which their problems might influence each other and, thus, are likely to reduce the reliability and validity of answers. As we will discuss in more detail later on, L-PCR assessment can be realized using simple yes/no-evaluations and the strength of the causal relation can subsequently be calculated either from how *frequently* or from how *consistently* the relationship has been reported.

In this study we analyze data collected via L-PCR assessments to showcase different metrics that can be calculated, as well as the research questions that can be addressed using these metrics.

11.2 Methods

11.2.1 Participants

We recruited 141 Swedish-speaking participants, at least 16 years of age, via social media. Participants were included in the analyses if they met the following three criteria: first, participants were included only if they had a PHQ-9 score of more than 9 at the beginning of the study period, corresponding to at least scores of "medium depression" (Manea et al., 2012). We set this minimum because participants who fell below this cut-off were likely to experience less persistent and consistent symptoms throughout the study period, therefore limiting the extent to which causal links among symptoms could be established. As will be discussed in more detail later on, this is a necessary (but not sufficient) criterion to obtain stable causal structures. Second, participants were included only if a minimum of three symptoms were reported during at least one third of the assessment days. In line with the previous point, this cut-off was set to ensure that there were at least three symptoms that were somewhat consistently reported. Third, participants were included only if they completed at least 20 assessments of the 28 days for which data was collected. This cut-off

was set because it would have been difficult to quantify the stability metrics discussed below for participants who provided fewer assessments.

Out of the 141 participants recruited for the study, 42 participants completed at least 20 assessments. Of these, 20 participants met both the PHQ-9 and minimum symptom criteria, composing our study sample. Included participants were on average 36.1 years old (range: 25–55). They had an average PHQ-9 score of 16.3 prior to the assessment and 15.6 at completion, both indicating levels of "moderately severe depression" (Manea et al., 2012). The average change in PHQ-9 scores pre- and post-assessment was 2.9. Participants reported that they had been depressed for a median of 15.0 months (range: 1 month – 35 years).

11.2.2 Materials

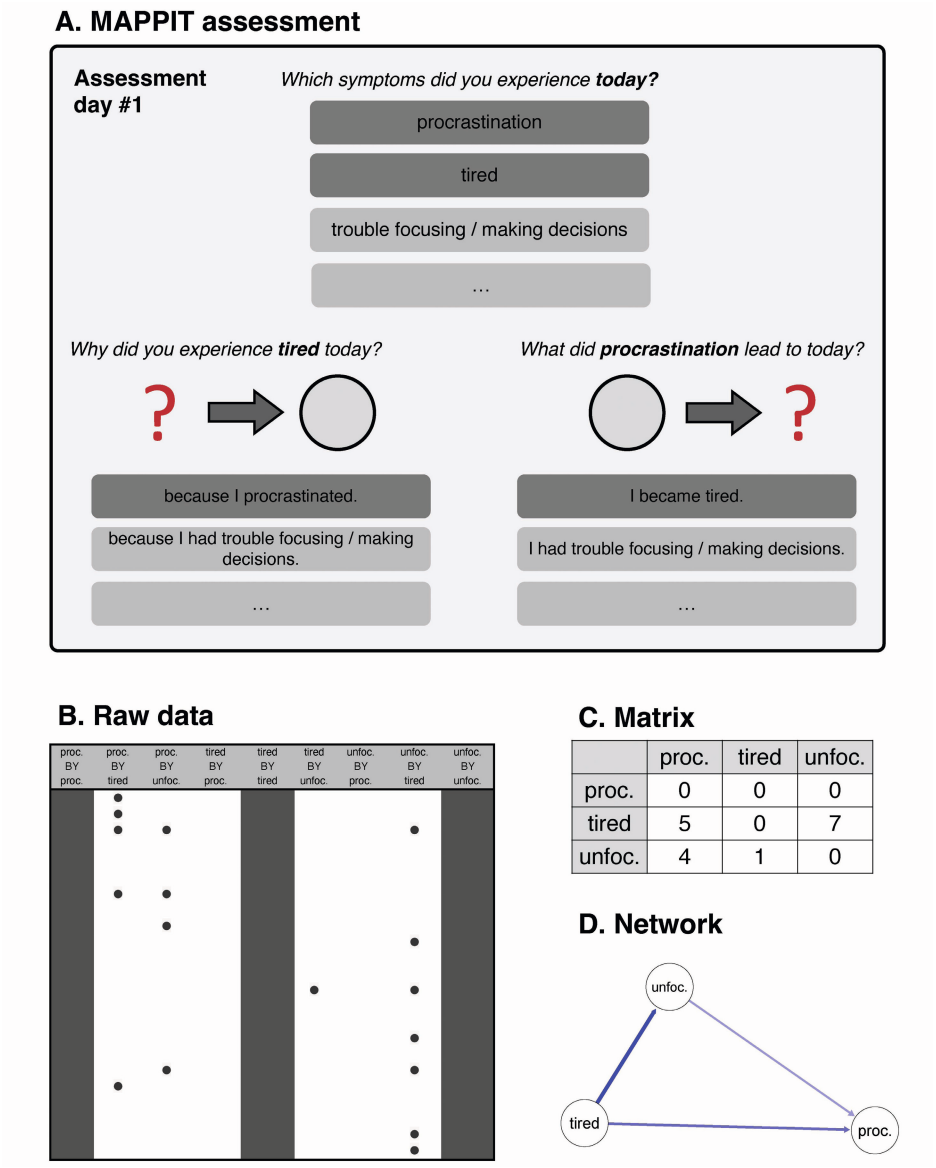
During a period of 28 consecutive days, participants completed daily assessments of the *Momentary Assessment of Perceived Problem Influences Tool* (MAPPIT; for an illustration, see the top panel of Figure 11.1), which is a questionnaire that was implemented in the survey platform Qualtrics to assess L-PCR. The MAPPIT includes 26 symptoms from a list of common problematic behaviors and emotions related to depression and anxiety and was developed to assess L-PCRs on a daily basis. It should be noted that although these items are here described as "symptoms", the included items should perhaps rather be seen as "problems" in a more broad sense (e.g. *procrastination* is not a psychiatric symptom) and future versions of this method could indeed include contextual variables (e.g. *financial problems*) as items. The survey was administered in Swedish. In addition, participants completed the PHQ-9 (Kroenke et al., 2001) before and after the MAPPIT assessment period.

11.2.3 Procedure

The study was approved by the Swedish Ethical Review Authority (ID 2019-06410). Participants were recruited through social media and completed the PHQ-9, followed by a period of 28 daily MAPPIT assessments. During that period, at the end of each day, participants first selected which symptoms (from the list of 26) they experienced throughout that day. For every selected symptom, they answered two multiple-choice questions: "*Why did you experience this problem today?*", aiming to elicit perceived *causes*, and "*What did this problem lead to today?*", aiming to elicit perceived *effects*. We varied the number of causes and effects a participant could choose; either only one, or multiple, to evaluate if this affect the stability of the resulting networks. Accordingly, participants were randomly assigned to the 'single' or the 'multiple' condition. In the 'single' condition participants were able to indicate only a single cause and effect for each symptom they indicated on a given day, whereas in the 'multiple' condition participants could indicate as many causes and effects per symptom as they wanted.

The causes and effects available as options were those symptoms that were selected as present on that day. Individuals only selected perceived causes and effects for each experienced symptom and did not rank or quantify the perceived strength of the relationship. For both cause and effect questions, respondents could also reply "Don't know" or "Other cause / effect," in which case they were given a free-text question asking what the cause or effect might be (this option was only used to evaluate the questionnaire, and was not included in the resulting network). Figure 11.1 shows an example of a participant selected at random for illustration purposes. Finally, participants rated to what extent the current day was representative of their experience, how time-consuming they

found the assessment, and if they experienced negative effects from the assessment. After the 28 assessments, participants completed a post-assessment of the PHQ-9.



	proc.	tired	unfoc.
proc.	0	0	0
tired	5	0	7
unfoc.	4	1	0

Figure 11.1. A) Example of a MAPPIT assessment of a single day for both the cause and effect task. B) Raw data accumulated for the example participant over the entire assessment period. Each row represents an assessment day and each column represents a potential causal link in the network, where the first variable refers to the experienced symptom selected from the list and the second to its attributed cause, e.g. "proc. BY tired" refers to procrastination being caused by tired. In absolute networks (the approach chosen in this paper) column sums are used to construct an adjacency matrix (C), which is visualized as a network (D). For visualization purposes, we only included edges which were reported more frequently than 25% of the weight of the maximum edge in the network.

11.2.4 Data preparation and performance metrics

11.2.4.1 General considerations in constructing causal networks from L-PCR assessments

Once participants have completed enough assessments (for recommendations regarding number of assessments needed, see below), their data can be used to construct idiosyncratic perceived causal networks. One straightforward way to construct such networks is to sum up the recorded causes and effects across all assessment days in one data matrix. This matrix represents all symptoms that have been experienced during the assessment as *causes* in columns, while their corresponding *effects* are in rows (see panel C in Figure 11.1). For example, the cell entry in column *overthinking* and row *anxiety* encodes the number of times overthinking was perceived to have led the individual to be anxious.

The weights of these cell entries refer to how often the causes and effects have been reported. Cell entries can either represent the *absolute* or the *relative frequency* of the causal relationships. The absolute frequency refers to the number of times a specific relationship has been reported in the assessment period. In the example above, the relationship *unfocused* causing *procrastination* is reported on four assessment days. Using absolute frequencies, the strength of the causal relationship *unfocused* to *procrastination* would be four. However, the individual only reported the symptom *unfocused* on nine out of the 23 assessment days. The *relative frequency* takes into account either the base rates of the cause or the effect, scaling the strength of the causal relationship to the base rate. Using base rate of the cause, the corresponding weight in the network would thus be 0.44 (four out of nine days when *unfocused* was present). It is currently unclear which of the two approaches is clinically more meaningful, as both provide unique and relevant information. We have opted for absolute networks in this paper because relative networks might inflate the relevance of edges pertaining to nodes that are infrequent.

Finally, as the questionnaire follows up on each experienced symptom with two assessment questions (causes and effects), there are multiple approaches to construct networks. For example, there are two ways to assess the relationship *unfocused* causing *procrastination*: the cause question targets information preceding the experienced effect ("what was the cause of *procrastination*?"), whereas the effect question targets information following the experienced cause ("what was the effect of *unfocused*?"). Using the two pieces of information, one can construct three types of networks: (a) networks based only on the perceived causes, (b) networks based only on the perceived effects, and (c) networks based on a combination of cause and effect data. Although cause and effect data should align in principle, this is not necessarily the case. Agreement[38] figures between cause and effect ratings range from 16.7% to 69.1% in our example data, with a mean of 37.0%. Interestingly, agreement ratings were significantly higher in the 'multiple' (46.43%) compared to the 'single' (26.49%) condition, $t(16.77) = 3.848$, $p = 0.001$, which indicates that agreement ratings might suffer if a person has to decide on only one cause and effect.

38 We operationalized *agreement* as the number of times that an edge has been reported in both the cause and effect assessments relative to the number of times this edge has been reported in at least either the cause assessment or the effect assessment. For example, an agreement of 37.0% indicates that in 37.0% of the cases when an edge was reported in either the cause or effect assessment, it was also reported in the respective other assessment type.

If one wants to use both sources of data and create a combined network (c), there are different possible ways to handle inconsistent assessments (e.g. on a given day, *unfocused* has been listed as a cause of *procrastination* but *procrastination* has not been reported as an effect of *unfocused*). In the following analyses, the combined networks follow the logic of the OR-rule, meaning that a relationship was counted as present whenever it was reported in either the cause or the effect assessment. The other possibility would be to apply the AND-rule (only counting the relationship when it is present in both the cause and effect assessments), which would be a more conservative approach resulting in sparser network structures. We opted for the OR-rule to obtain denser networks which have a higher chance for stability (see also the discussion section, specifically the paragraph on counterfactual assessment).

11.2.4.2 Defining performance metrics for L-PCR assessments

We calculated several metrics to evaluate the performance of the networks. The metrics were defined to address three important practical questions about specific networks: How many (and which) symptoms should be included in a personalized network? For how long does one need to collect data in order for a specific network to become stable? How can one evaluate if the network properties changed during the assessment period, suggesting that the causal interplay amongst variables was not captured by a single network structure?

11.2.4.2.1 Symptom selection: Stable cores

A key issue in constructing networks pertains to the inclusion of relevant items (Fried & Cramer, 2017). The goal of item selection is to identify symptoms that are playing a key role in maintaining the pathological state. Symptoms often come and go across days, with some symptoms experienced only rarely. Symptoms that are reported infrequently have less potential to exhibit causal relations because relationships with other symptoms are only possible if the symptom is indicated as present during the day. For example, suppose an individual reports *unfocused* only once during the assessment period. They report *procrastination* as a direct effect on that day but they also experience *procrastination* on most other days when they have not been *unfocused*. Even though the individual clearly identified *unfocused* as a cause of *procrastination*, *unfocused* is disqualified as an explanation for the persistent experience of *procrastination* because *procrastination* also occurs independent from *unfocused*. To this end, we focused on the most frequent set of symptoms, which we refer to as the *stable core* – the core of symptoms that is consistently reported.

The chosen minimum frequency of symptoms is directly linked to the size of the network: the higher the specified minimum frequency, the smaller the stable core. There is no general recommendation for setting this cut-off, as individuals likely differ in how consistently they experience symptoms. In this study, for the sake of simplicity, we opted to include symptoms that were present during at least 1/3 of assessments for that respondent. We will address this point in more detail in the discussion section.

11.2.4.2.2 Length of assessment period: Causal saturation

In the context of L-PCRs, single assessment days will not fully represent the causal structure of a participant's overall causal network but rather add a piece to its bigger picture. This introduces the

question of for how long one needs to collect data to reach a point where further data collection no longer contributes to increased understanding the relations. In line with this reasoning, we define *causal saturation* as the moment in data collection when adequate amounts of data are obtained to develop a comprehensive understanding of the causal network. Past this point of causal saturation, additional data points will no longer alter the structure of the network substantially unless the participant's actual causal network structure changes (see below, *causal drift*).

We assessed causal saturation by implementing a bootstrap-procedure. For each participant, this algorithm first randomly samples from that participant's assessments with samples growing larger and larger in steps of 2 (i.e. randomly selecting 2, 4, 6, ... days from the total pool of assessments), second, splits each sample into two equally sized sub-samples, and, third, calculates the similarity between the network structures of the sub-samples (i.e. the spearman correlation of their weight matrices). For each sample size, this procedure is repeated 1,000 times for each participant to obtain an approximation of the sampling distribution of correlations between the sub-samples (Efron, 1979). Based on the results of the bootstrap, we defined two criteria for causal saturation that we evaluated for each participant. First, we determined how many assessment days it took to achieve a mean correlation between sub-sample networks of $r = 0.70$. Second, we calculated the mean correlation between sub-samples when including 20 assessment days as an indication for the consistency of an assessment number that we deemed feasible to collect.

11.2.4.2.3 Network (non-)stationarity: Causal drift

In the previous paragraph we assumed that additional data points will stabilize the causal structure. However, there are also scenarios where more data points will lead to increased dissimilarity within the data. This can happen when the true network structure changes over time, so that new datapoints will reflect a different network compared to earlier datapoints. Conceptually, this relates to the stationarity assumption of networks, which entails that properties of the network do not change over time. If causal networks change over the assessment period, aggregating observations across the period into a single model will fail to adequately reflect this change. We can assume such a time dependency of the model if we assess over the course of therapy (because therapy likely changes the system) but, since depressive symptoms are highly context-dependent, systems may also change in the absence of interventions.

In the context of L-PCRs, we define *causal drift* as the extent to which the distance of two assessment days is related to the correlation between their cause and effect matrices. If there is no causal drift (i.e. the system does not change over time), there should be no relationship between the length of distance between two assessments and the correlation of their matrices. If there is a causal drift detected in client data, caution in interpreting network structures is warranted; this is because the presence of such a drift indicates that the causal system should not be represented by a single structure, but rather that the client's causal network changes over time. To give a concrete example, if the perceived causal mechanisms on the first assessment day are as similar to the second day as they are to a day three weeks later, we could take this as an indication that causal drift is not a problem. By contrast, a situation where the network structures of two assessment days become more dissimilar over time, causal drift is indicated. We calculated correlations between time distance and similarity of networks, as well as visualized scatterplots, to detect potential non-linear relationships.

11.2.5 Research aims and data

The research aims of this paper are two-fold. First, we quantify averages of the size of the stable core (number of symptoms included in networks), causal saturation (assessment days needed to create stable networks), and causal drift (whether actual network structures are time-invariant) for the overall study population. In addition, we include exploratory group comparisons between the questionnaire assessing single versus multiple causes/effects, as well as investigate relationships between such characteristics as size of stable core, days to reach saturation, and symptom severity. These results aim at informing design characteristics for future studies and should be interpreted with caution; as the current sample is too small to derive firm conclusions, results of this part of the study aim should therefore be seen as a pilot study. Second, we select two individuals that represent desirable versus undesirable characteristics in regard to causal saturation and drift, then discuss their individual assessments in more detail. This second part of the analysis aims at researchers and clinical practitioners who wish to use MAPPIT for therapy rather than for group-level studies and are therefore specifically interested in quality assessment and the interpretation of idiographic models.

11.3 Results

11.3.1 Study aim 1: Overall study sample

11.3.1.1 Feasibility and representativeness of assessments

On average, assessments took 7.4 minutes to complete, participants rated the daily MAPPIT as not very time consuming (on a scale of 1 to 5: average of person-wise means 1.4, SD = 0.46), and negative effects from assessment were rated as low (on a scale of 1 to 5: average of person-wise means 1.4, SD = 0.44). Finally, participants indicated that an average of 47.8 percent of assessment days were completely representative for their experience. It should be noted that participants who experienced the MAPPIT as very time-consuming and who experienced negative effects likely dropped out early in the data collection and were thus not included in our present sample.

11.3.1.2 Stable cores, causal saturation, and causal drift

Participants had a mean stable core size of 6.20 symptoms (SD = 3.40), indicating that, on average, participants experienced about six of the listed symptoms during at least one third of the assessment. The symptoms in the stable core were subsequently used to construct the networks per participant. Participants did not differ in their stable core size across the two conditions, $\text{mean}_{\text{single}} = 6.40$, $\text{SD}_{\text{single}} = 5.10$; $\text{mean}_{\text{multiple}} = 6.00$, $\text{SD}_{\text{multiple}} = 2.63$; $t(13.45) = 0.22$, $p = 0.829$. Causal saturation (i.e. reaching a 0.70 correlation between equally-sized subsets of assessment days) was achieved for only eight out of the 20 participants. Within this subset of participants, causal saturation was achieved after an average of about 15 days (mean = 15.25, SD = 3.20). Across all participants, stability at day 20 was on average 0.60 (SD = 0.21).

Finally, we visualized causal drift in scatterplots for each participant, representing the distance between two assessment days on the x-axis and the correlation of the respective adjacency matrices on the y-axis. Figure 11.2 shows causal drift diagrams for three participants, highlighting 1) an example of causal stability, i.e. no systematic relationship between the time distance and matrix correlation

(left panel); 2) an example where a slight causal drift may be present, i.e. increasing dissimilarity between causal structures for larger time distances (middle panel); and 3) an example where causal drift remains inconclusive (right panel). The third case may occur for less severe symptom profiles because a lack of persistent symptoms leads to zero-inflated cause and effect recordings. This, in turn, results in many matrices that have no variance and therefore makes it impossible to establish correlations between many pairs of assessment days. We will elaborate on this specific point in the recommendations section below. Slopes of the causal drift diagrams ranged from $\beta_{min} = -0.12$ to $\beta_{max} = 0.08$, with a mean of $\beta_X = -0.02$ and $\beta_{SD} = 0.04$.

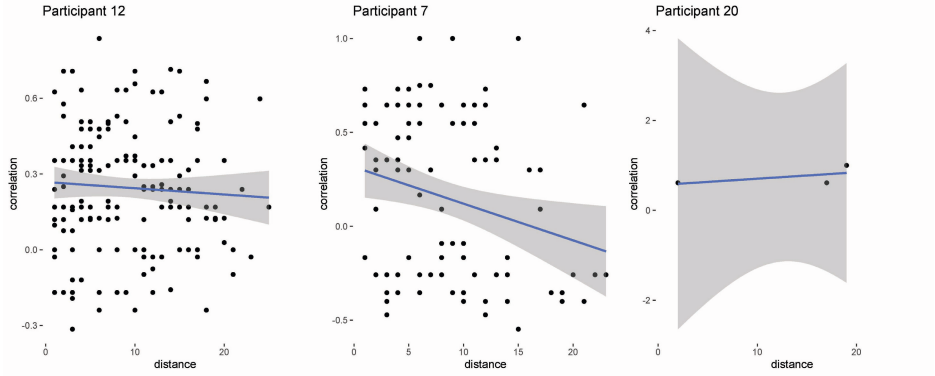


Figure 11.2. Causal drift diagrams for three example participants. The left panel (participant 12) illustrates an example of causal stability because there is no systematic relationship between time distance and correlation of the adjacency matrices. The middle panel (participant 7) illustrates an example of slight causal drift because adjacency matrices become increasingly dissimilar for larger time distances. Finally, the right panel (participant 20) shows an example where causal drift remains inconclusive because the most assessment days are zero-inflated.

11.3.1.3 Correlates with symptom severity and chronicity

As can be expected, participants who reported more persistent symptoms (i.e. participants with a larger core size) also had higher PHQ-9 scores, $r = 0.569$, $t(18) = 2.933$, and $p = 0.009$. None of the remaining pairwise correlations were significant (core size – stability, $r = 0.268$; stability – PHQ9-score, $r = 0.405$; core size – chronicity, $r = -.112$; stability – chronicity, $r = 0.219$), which is unsurprising given the relatively small sample size. Power calculations indicate that, in order to detect small to moderate relationships, sample sizes of about 60 participants would be needed.

11.3.2 Study aim 2: Individual causal network characteristics

11.3.2.1 Example 1: 'Moderately-severe' depressed participant

The first example is a 35-year-old participant with a PHQ-9 score of 19, which can be categorized as 'moderately-severe' depressed according to Manea et al. (2012). They reported a total of 72 months of depression at the start of assessments. The participant was assigned to the 'multiple' condition, completed a total of 27 assessment days, and had a stable core of nine symptoms. Figure 11.3 shows which symptoms have been experienced on each assessment day (top left, A), the causal network

constructed from their data (top right, B), and the causal drift (bottom left, C) and causal saturation diagrams (bottom right, D).

This participant was selected to illustrate an example of good stability, with a correlation between equally-sized subsets of 0.82 at their final assessment day. The initial cut-off of 0.70 stability was achieved already at 12 assessment days. As will be discussed in more detail later, persistence of symptom presence within the stable core facilitates causal saturation,[39] because a symptom-to-symptom relationship can only be stable if the two symptoms are also present for most of the time. In line with this reasoning, this participant is a good example for when the method discussed in this paper can be successful; most of their symptoms in the stable core have been relatively consistent throughout the assessment period (see panel A).

Finally, the participant showed a slight negative slope in the causal drift diagram (panel C), indicating that their causal network might have changed slightly throughout the assessment period. An alternative explanation for this negative slope could be that the causal structure itself did *not* change but rather that other patterns of symptoms were present in the beginning compared to the end of the assessment period. As we only assessed perceived causal relationships for symptoms that were present and not absent, network structures will automatically differ when different symptom patterns are experienced.

11.3.2.2 Example 2: 'Moderately' depressed participant

The second example is a 25-year old participant with a PHQ-9 score of 14, which can be categorized as 'moderately' depressed according to Manea et al. (2012). The participant was assigned to the 'single' condition, completed a total of 25 assessment days, and had a stable core of five symptoms. Figure 11.4 shows which symptoms have been experienced on each assessment day (top left, A), the causal network constructed from their data (top right, B), and the causal drift (bottom left, C) and causal saturation diagrams (bottom right, D).

This participant was selected to illustrate an example of weak stability, with a correlation between equally-sized subsets of 0.36 at their final assessment day. The participant therefore failed to reach stability anywhere near the preset cut-off of 0.70. One possible explanation for this is that, in contrast to the previously described case, this participant had a rather inconsistent symptom pattern (see panel A). Interestingly, although causal saturation was not achieved for this person, their network showed consistence with free-text reports at the beginning of the study. Specifically, they described that the reason they get stuck feeling bad is that worrying about feeling bad makes them feel bad, which is represented in the network (*overthinking* → *anxiety*, *anxiety* → *overthinking*).

This participant had a rather small core of symptoms, which could lead to a situation where some causal effects are established with symptoms that did not make the cut-off of being experienced during at least one third of assessment days. For that reason, several of the assessment days resulted in networks with no causal connections at all, which in turn reduced the number of available observations in the causal drift diagram because correlations with empty assessments could not be

39 Note that symptom persistence is necessary but not sufficient for causal saturation because structures can still differ across persistent symptoms, which hampers causal saturation.

established. Overall, this participant illustrates some characteristics that warrant caution in using the method discussed in this paper, e.g. smaller core sizes and inconsistent experience of symptoms.

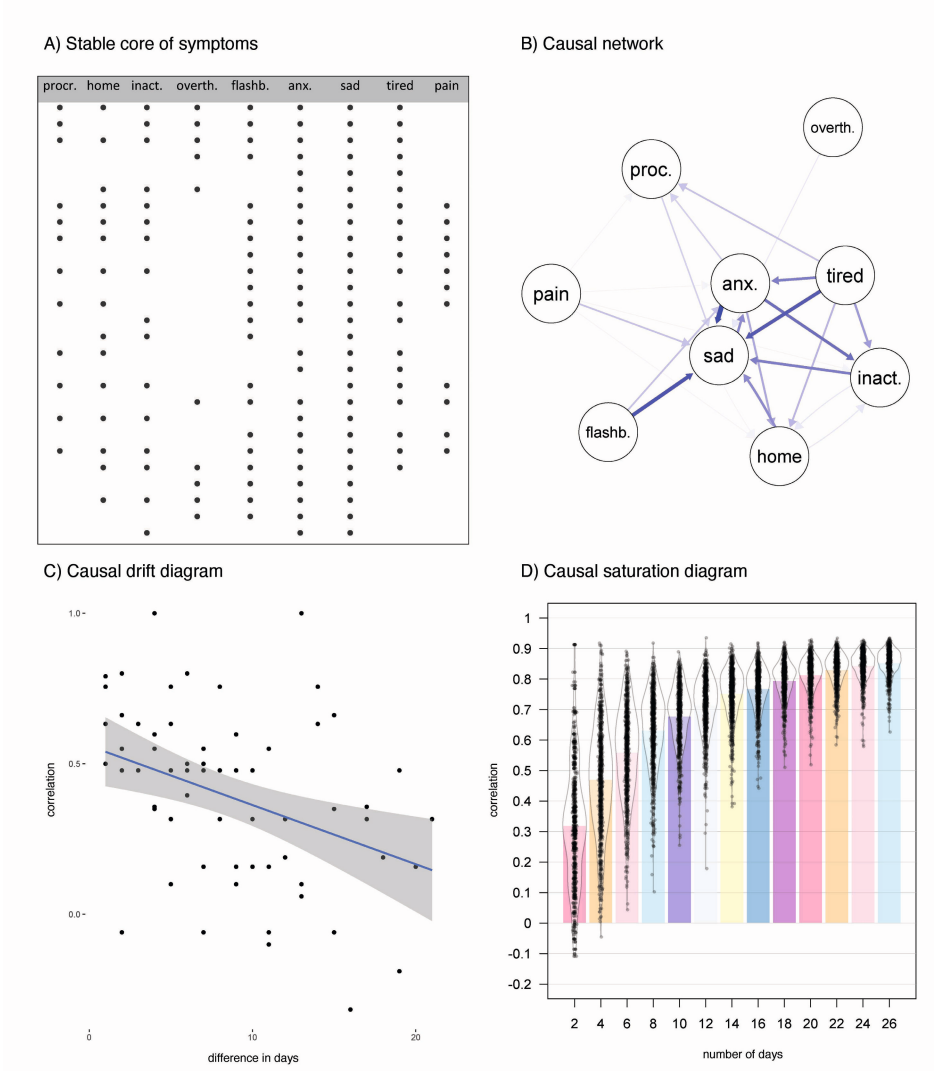


Figure 11.3. Results for example client 1. Top left (A): stable core symptoms experienced on each assessment day, top right (B): causal network based on the combination of cause and effect data (absolute frequencies). For visualization purposes, we only included edges which were reported more frequently than 25% of the weight of the maximum edge in the network. Bottom left (C): causal drift diagram, indicating a slight negative relationship between time distance and similarity of adjacency matrices, bottom right: Causal saturation diagram, indicating saturation reached at 12 assessment days.

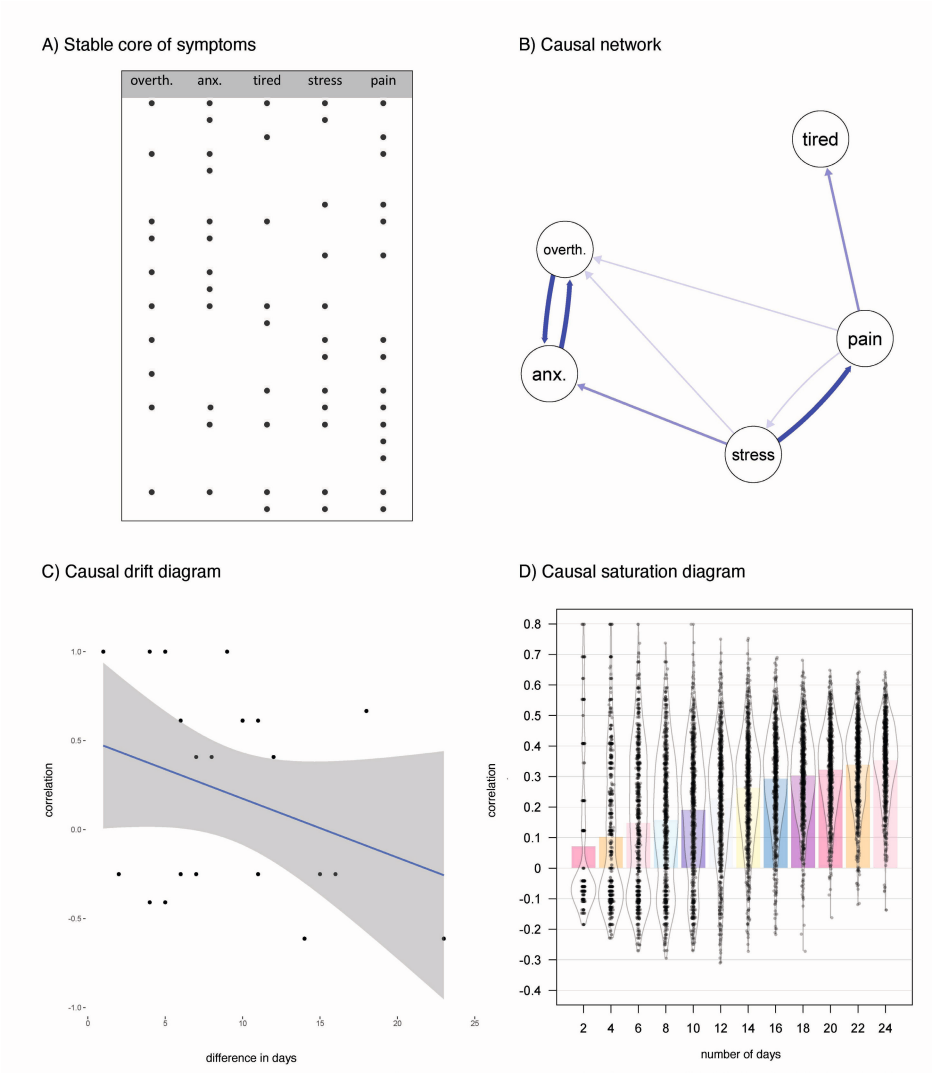


Figure 11.4. Results for example client 2. Top left (A): stable core symptoms experienced on each assessment day, top right (B): causal network based on the combination of cause and effect data (absolute frequencies). For visualization purposes, we only included edges which were reported more frequently than 25% of the weight of the maximum edge in the network. Bottom left (C): causal drift diagram, indicating a slight negative relationship between time distance and similarity of adjacency matrices, bottom right: causal saturation diagram, indicating that saturation as defined above was not achieved.

11.3.3 Free-text answers and evaluation

The "single" group could give free-text answers instead of the other selected symptoms in the cause / effect options, whereas the "multiple" group could give free-text answers to complement the other selected cause / effect options. Also, participants were asked about what was missing in the MAPPIT in an exit-survey at the end of data collection. In these answers, the most commonly-reported miss-

ing variables were actually already in the questionnaire (e.g. the respondent did not select "anxiety" as experienced that day but, when queried about causes for some other symptom that was selected, actually wanted to give anxiety as the cause). Related, participants explicitly wanted to select symptoms that had been experienced several days ago as a cause for current symptoms, e.g. "substance use" causing "anxiety" the subsequent day (i.e. hangover) or "conflicts" causing "rumination" in the following week. Participants also gave contextual factors as causes, where the most frequently recurring was "war in Ukraine", typically as a cause for worrying (the data collection was carried out in March 2022), followed by bullying and financial issues. Sometimes participants wrote a diagnosis as a cause for a symptom, such as "depression". One participant seemed to misunderstand the cause- and effects questions, and gave "war in Ukraine" as another *effect* of worrying (that is, unless the participant in question was in fact suffering from a quite severe delusion).

11.4 Discussion

In this chapter we introduced a new approach to constructing networks using longitudinal assessment, which we termed L-PCR. We developed quality criteria, tested them in a pilot sample of 20 participants, and presented two example networks. Overall, we were able to identify a tendency for higher stability of networks for individuals with more persistent symptom experiences across days. However, less than half of the individuals reached the pre-defined stability criterion. The idiographic analyses showed that this approach may be particularly fruitful for individuals with a large core size of symptoms that occur on most of the assessment days. For one such example, we identified causal saturation after 12 assessment days. Further, causal drift diagrams can give insight into potential changes of the network structure over time. Given that the L-PCR method is new, in this section we discuss strengths and limitations, as well as provide some pointers for future development in this section.

11.4.1 General strengths of L-PCR and PCR networks compared to ESM-based networks

We see two ways in which PCR-based networks (i.e. using either PECAN or the PCR method by Frewen), as well as L-PCR-based networks (i.e. using MAPPIT), can address challenges typically faced in the context of ESM-based statistical networks.

First, (L-)PCR-based networks allow for a more intuitive interpretation of relationships; the directed connections in the network can naturally be interpreted as causal, assuming that the participant understands the questions and is able to rate causal processes reliably. In contrast, ESM-based networks usually rely on many statistical assumptions (e.g. linearity, stationarity, correctly-chosen assessment frequency) that make causal interpretations hazardous. As a consequence, network metrics that are calculated based on relationships are more intuitive in (L-)PCR-based networks compared to ESM-based networks. For example, a prominent interest in the context of network analysis lies in identifying important ("central") targets for intervention. Centrality metrics exist but, as many rely on the information flow of the network (i.e. defining centrality as a function of edge structures), they presume that edges have a natural interpretation. As discussed above, (L-)PCR assessments can be more intuitively interpreted compared to statistical edges and centrality

may therefore be more meaningful in this context as well (e.g. the probability of a symptom causing another symptom a given day). The validity of centrality metrics in the context of (L-)PCR-based networks needs to be investigated in future research.

Second, a common issue in ESM-based networks is that temporal effects can be "missed" if the assessments are administered on an inappropriate timescale. For example, the temporal effect of experiencing heart-racing and developing feelings of panic may not show up in a statistical network because the process unfolds on a different timescale (seconds) than the one administered in the ESM assessment. In L-PCR-based networks, specifically in the MAPPIT, effects are reported that occurred within the same day, irrespective of the exact timeframe. On one hand, this means that edges are not interpretable as referring to the exact same timescale; on the other hand, this may make networks more clinically useful as they represent the relevant dynamics.

11.4.2 Strengths of L-PCR over traditional PCR assessments

Next to general advantages of (L-)PCR-based approaches, we argue that the L-PCR introduced in this paper advances assessment over previous single-session techniques in at least three ways:

First, individuals do not have to accumulate observations retrospectively to arrive at judgments on causal relations, which can lead to bias (Shiffman et al., 2008). Instead, they are asked to focus on very specific events that happened within the same day of assessment. This is in line with the well-established finding that elicited quantities should be based on concrete situations to avoid bias (O'Hagan, 2006).

Second, individuals may find it difficult to rate the causal strength of a relationship on a Likert-scale or by attributing percentages to causes. L-PCR circumvents this problem by operationalizing the strength of a causal relation as the proportion of days that a symptom-symptom interaction is experienced. This way, a causal strength index can be derived without requiring direct reflections from the individual.

Third, longitudinal assessment results in data that is richer in regard to psychometric inspection when compared to single-session PCR. We introduced indices that provide insight into the consistency of assessments (causal saturation) and potential changes in the causal structure over time (causal drift).

11.4.3 Limitations

It is important to distinguish the conceptual idea of L-PCR assessment from the specific MAPPIT implementation. In the current implementation of the MAPPIT, we identified four limitations to the data and results discussed in the paper. The below limitations can be addressed in future iterations of the MAPPIT but do not jeopardize the concept of L-PCR assessments in general. Rather, we discuss the points below to provide pointers for improving the assessment of L-PCR in the future.

First, while the possibility to assess causal associations on different timescales can be a strength, it also requires specific attention because the sequence of symptoms can lead to confusion in the way questions are phrased. For example, in the current implementation, the insomnia symptom was phrased as asking about insomnia experienced the *previous night*. However, the causes available to that insomnia were the other symptoms experienced *today*. In addition, a programming error in

the question about effects of insomnia allowed insomnia to be an effect of itself. This is certainly possible and was indeed selected by some participants, but self-causation was not allowed for the other cause/effect-questions.

Second, and related to the issue of timescales, participants were only able to indicate causes and effects that occurred *within one day*. Free-text answers indicated that participants perceived some causes to extend back several days and such relations were therefore not captured (e.g. a conflict with a spouse might cause rumination over the following week). These relations could not be reported in the current implementation of MAPPIT. To capture these, future iterations of MAPPIT should allow *all* symptoms as potential causes to the symptoms experienced on a particular day. On the other hand, such assessments might quickly become complicated and too overwhelming. Such a set-up might require shorter item lists, for example by combining items into wider categories such as "negative emotions" instead of specific emotions.

Third, the list of items from which individuals could choose may not have been fully representative of the list of causes at play. More specifically, contextual factors were missing that are likely specific to the individual. We have assessed such potential causes in an open text-field, yet there is currently no systematic way of making use of this data in the network itself. Another issue is that it is possible that participants were confused by switching between giving causes and effect to symptoms. If this indeed was the case for even a minority of participants, this would both create false feedback loops and decrease the stability of networks.

Fourth, we observed a high dropout rate that may indicate a systematic missingness problem, therefore making our study sample not representative. As previously discussed, the MAPPIT may be especially fruitful for individuals with persistent symptom experiences but these individuals may also be less likely to complete the necessary number of daily assessments.

11.4.4 Recommendations for future studies using L-PCR

11.4.4.1 Study sample characteristics

Based on our analyses, we can derive some preliminary recommendations for the study sample of future data collection. Although these recommendations are based on a relatively small pilot sample, they can still inform future studies as they are not only based on empirical considerations but also on statistical constraints of the way data is collected (see the discussion on counterfactual explanations below). Irrespective of group-level or idiographic designs, individuals with more persistent symptoms – that is, a large stable core with symptoms frequent on most days – will have more stable networks compared to individuals with infrequent symptoms.

11.4.4.2 Data collection duration

In our sample, individuals who achieved causal saturation ($r = 0.70$) did so after 15 days on average. Adding some buffer onto this timeframe, for example doubling the standard deviation, would result in a recommendation of about 21-22 days. A preliminary indication of achieving causal saturation for suitable individuals would thus be around three weeks. Depending on the research goal, more lenient or strict cut-offs for causal saturation can be set resulting in different recommendations for planning the data collection duration. For example, if the goal is to quickly get a rough idea of the

current causal pattern, one could set lower thresholds for causal saturation and would need less assessment days to achieve them. In such cases, traditional PCR methods may also be a suitable alternative. Conversely, if the interest is to understand the causal system of a single person in depth and observe long-term changes, potentially even looking into effects of interventions, one could set higher cut-offs for causal saturation or specifically focus on the causal drift diagrams. The necessary stability needed to select the optimal treatment target could be investigated, as well as the effects of excluding specific assessments that the respondent reports as unrepresentative.

11.4.4.3 Single versus multiple assessments

We could not find differences between causal saturation efficiency for individuals who indicated single versus multiple causes. This is unsurprising given the small sample size and the results therefore remain inconclusive. We expect that, in larger study samples, individuals indicating multiple causes will achieve causal saturation faster compared to single causes. This is because symptoms are likely multi-causally determined and the single assessment may therefore introduce an artificial all-or-nothing distinction, resulting in different structures on different days. As mentioned above, this hypothesis needs further testing in larger samples, also including assessments of increased burden by asking for multiple causes and effects for each experienced symptom. We did, however, find differences in favor of the multiple condition regarding agreement between cause and effect ratings, which supports a preliminary recommendation for allowing for indicating multiple causes and effects.

11.4.4.4 Choosing a minimum symptom frequency cut-off

Different research interests may guide how conservatively the stable core cut-off should be set: if only the most frequent symptoms are included, one can quickly get a clear picture of consistent causal relations and smaller networks are therefore more likely to reach causal saturation compared to a larger network. However, very small networks (e.g. only three symptoms) may not always be most clinically informative because symptoms can very well be experienced as debilitating, even though they are not present on most days. We recommend the decision on core size be made on an individual basis and in collaboration with the client, depending on the symptom presentation and research interest. The size of the stable core can also be increased by individualizing the symptom list to the specific client, both by phrasing the items in such as way that the participant will endorse them more often, and including person-specific contextual items (e.g. "financial issues").

11.4.4.5 The challenge of counterfactual assessment

One main determinant of how fast causal saturation can be achieved is the extent to which symptoms in the stable core are consistently experienced. Only when symptoms are present on a regular basis is it possible to determine strong causal relationships. This is because, in the current implementation of the MAPPIT, the presence of symptom A is necessary for contributing to a causal effect with symptom B. If symptom A is absent, cause–effect relationships between A and B are not assessed. However, the absence of both symptoms A and B might in fact be quite informative. For example, a client may report that they did not worry on a given day, could also report that they did not consume alcohol that day, which informs us about potential causes for drinking in that individual.

In the field of causality, establishing a cause–effect relationship between two absent symptoms is referred to as a counterfactual explanation (Dablander, 2020; Pearl et al., 2016). Adding such counterfactual assessments to the MAPPIT could make it more stable for individuals with less severe symptom profiles, as their causal networks could also be informed by absent symptoms (e.g. asking "Why do you think you did not drink today?" and give the other symptoms as possible answers, e.g. "because I did not worry."). For this to be feasible, an individualized item list as described above would need to be decided upon for each client, i.e. symptoms that realistically would or would not be experienced on a typical day. However, counterfactual assessment poses additional challenges, for example making the assessment less intuitive by asking for hypotheticals. After all, symptoms have not been experienced and it may therefore be hard for individuals to argue on the basis of absent symptoms. In a broader sense, (L-)PCR assessment can be seen as a form of eliciting prior information from experts (Burger, Epskamp, et al., 2022; O'Hagan, 2006; Stefan et al., 2020); within this field it has been well-established that quantities we assess should be as concrete as possible. Intuitive counterfactual assessment in the context of (L-)PCR remains an important direction for future research.

11.4.4.6 Causality in specific situations

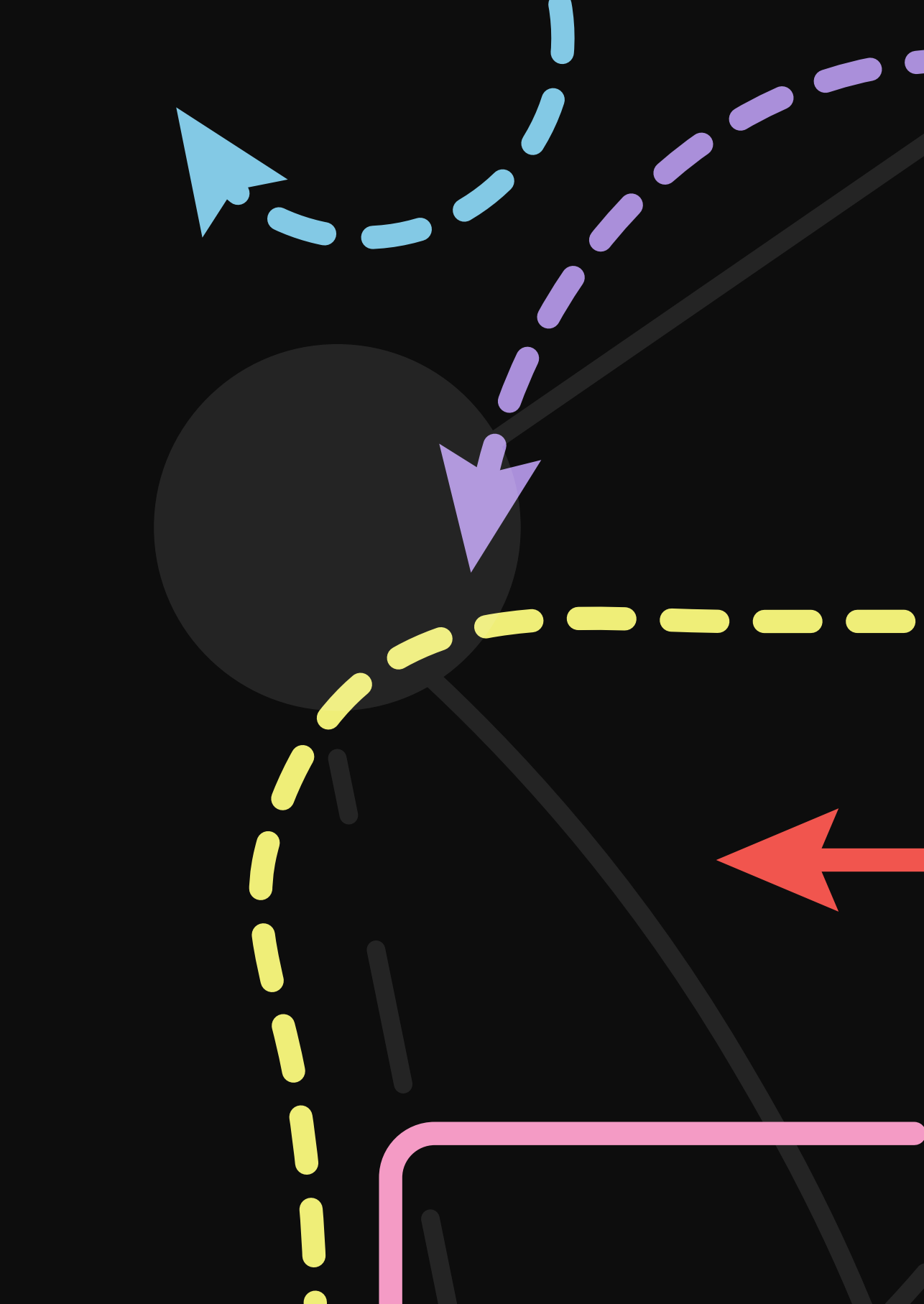
Another possibility that should be mentioned is to create networks not from symptoms experienced throughout a full day, but rather in the context of specific problematic situations. This is typically done in functional analyses in the assessment phase of behavioral therapies, but could be done in a more structured manner. For instance, the respondent could be asked several times a day what situations (or symptoms) they have been experienced. For each situation, both causes leading up to the situation, and (expected) effects could be reported. Different types of situations can then be categorized systematically (e.g., categories like 'situations that lead to feelings of loneliness'), which allows to construct networks where nodes represent categories of situations and related causes and effects.

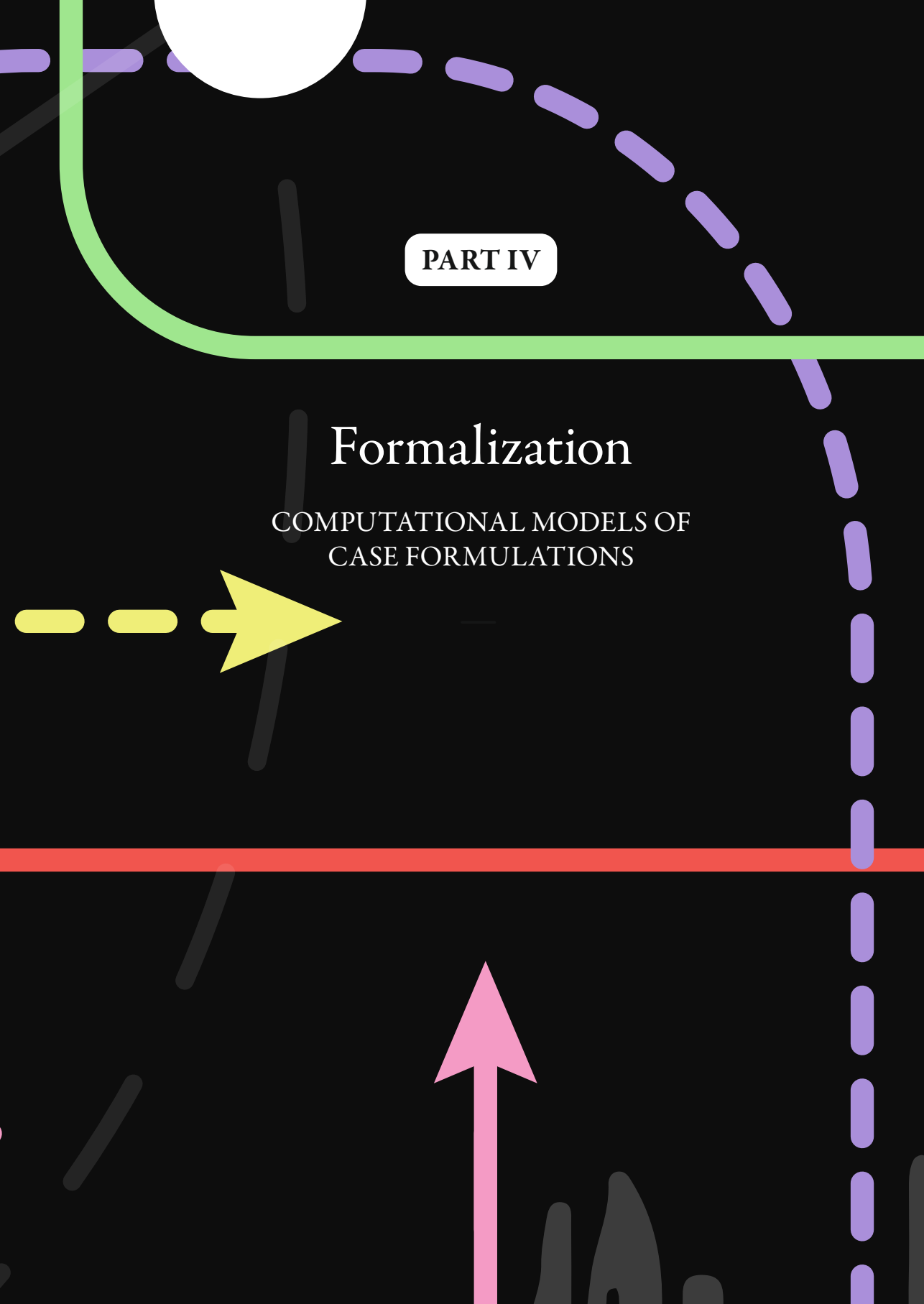
11.4.4.7 Personalizing item content

Finally, depending on the research aim, one needs to decide if the items should be personalized or kept constant across all participants (Bergh et al., 2022). This decision very much depends on the type of inference a researcher is seeking: fully within-person (idiographic) studies do not necessarily need consistency of items, whereas between-person inferences often require that items are aggregated across participants. In the pilot data presented here, we used the same set of 26 items across all participants. A specific advantage of personalizing items is that the item core can already be defined a priori and that more specific contextual factors can be included.

11.5 Conclusions

L-PCR assessments can be used to construct causal symptom networks for clients with persistent symptom experience. This method may provide more ecologically valid assessments compared to traditional PCR methods and can be used for investigating if the network changes over time. Future research should focus on advancing assessment to counterfactual reasoning, as well as clinical trials to examine the clinical utility of networks produced by this approach.




PART IV
Formalization
COMPUTATIONAL MODELS OF
CASE FORMULATIONS

CHAPTER 12

FORMALIZING CASE FORMULATIONS

CHAPTER 12

Abstract

The past decades of research have seen an increase in statistical tools to explore the complex dynamics of mental health from client data, yet the application of these tools in clinical practice remains uncommon. This is surprising, given that clinical reasoning, e.g., case formulations, largely coincides with the dynamical system approach. We argue that the gap between statistical tools and clinical practice can partly be explained by the fact that current estimation techniques disregard theoretical and practical considerations relevant to psychotherapy. To address this issue, we propose that case formulations should be formalized. We illustrate this approach by introducing a computational model of *functional analysis*, a framework commonly used by practitioners to formulate case formulations and design client-tailored treatment. We outline the general approach of formalizing idiographic theories, drawing on the example of a functional analysis for a client suffering from panic disorder. We specified the system using a series of differential equations and simulated different scenarios; first, we simulated data without intervening in the system to examine the effects of avoidant coping on the development of panic symptomatic. Second, we formalized two interventions commonly used in cognitive behavioral therapy (CBT; *exposure* and *cognitive reappraisal*) and subsequently simulated their effects on the system. The first simulation showed that the specified system could recover several aspects of the phenomenon (panic disorder), however, also showed some incongruency with the nature of panic attacks (e.g., rapid decreases were not observed). The second simulation study illustrated differential effects of CBT interventions for this client. All tested interventions could decrease panic levels in the system. Formalizing idiographic theories is promising in bridging the gap between complexity science and clinical practice and can help foster more rigorous scientific practices in psychotherapy, through enhancing theory development. More precise case formulations could potentially improve intervention planning and treatment outcomes. We discuss applications in psychotherapy and future directions, amongst others barriers for systematic theory evaluation and extending the framework to incorporate interactions between individual systems, relevant for modeling social learning processes. With this chapter, we hope to stimulate future efforts in formalizing clinical frameworks.

This chapter has been adapted from: Burger, J., van der Veen, D. C., Robinaugh, D. J., Quax, R., Riese, H., Schoevers, R. A., & Epskamp, S. (2020). Bridging the gap between complexity science and clinical practice by formalizing idiographic theories: a computational model of functional analysis. *BMC medicine*, *18*, 1-18.

12.1 Introduction

Complex system thinking is gaining increasing importance in understanding mental health (Borsboom et al., 2018; Cramer et al., 2016; Fried et al., 2017). In recent years, some clinicians have proposed a move away from the approach of treating mental illness as disorder categories towards a focus on processes and client-specific mechanisms in psychotherapy (Hofmann & Hayes, 2019c). These proposals call for a framework for thinking about mental illness in terms of *systems*, to understand the *processes* underlying psychopathology, and to apply this understanding to *client-specific* contexts. The network perspective to psychopathology (Borsboom, 2017; Borsboom et al., 2011; Borsboom & Cramer, 2013; Wichers, 2014), conceptualizing psychological disorders as complex interactions of symptoms and related mental health factors, provides a framework to address this movement. Statistical procedures that allow for the estimation of psychopathological networks have been developed (Epskamp, Borsboom, et al., 2018; Epskamp et al., 2012; Epskamp & Fried, 2018), and applied across a wide range of mental disorders (David et al., 2018; Dotterer et al., 2020; Fisher, 2015; Lutz et al., 2018).

Furthermore, and arguably most relevant for psychotherapy, tools for *idiographic* network analysis have been developed (Epskamp, Waldorp, et al., 2018; Fisher et al., 2017), allowing us to explore client-specific symptom dynamics from data collected using the experience sampling method (ESM; Stone & Shiffman, 1994). This approach may be especially relevant for psychotherapy, as it has the potential to be embedded within clinical practice through informing the formulation of idiographic theories (i.e., case formulations) and the identification of client-tailored intervention targets (Epskamp, van Borkulo, et al., 2018). Indeed, idiographic network analysis aligns well with the movement towards process-based psychotherapy (Hofmann & Hayes, 2019c). It therefore seems surprising that, despite the availability of supportive statistical tools and efforts to provide primers for conducting idiographic research (Piccirillo, Rodebaugh, et al., 2019), the actual application of personalized network modeling within psychotherapy is to date rare.

12.1.1 From implementation barriers to a clinician's wish list

Implementation gaps between mental health research and clinical practice are a topic of enormous importance (Proctor et al., 2009; Wensing & Grol, 2019). With the emergence of the complex system approach in mental health research, there has been specific interest in implementing statistical tools to explore client-specific symptom dynamics in clinical practice. It is commonly assumed that successful implementation is in part a question of providing technical trainings and accessible guidelines for clinicians (Piccirillo, Rodebaugh, et al., 2019). However, merely training clinicians in adopting tools provided by methodologists does not guarantee that these tools also result in models that map onto the *language* used by practitioners. Indeed, an often-discussed barrier to implementation is the accurate *translation of knowledge* into the relevant practice field (Wensing & Grol, 2019). That is, the language used to discuss promising research findings and techniques does not always match the targeted language of the practitioner.

This issue applies to the estimation of personalized network models. At present, network estimation methods remain technical and do not account for potentially relevant clinical considerations. For example, network estimation methods identify "highly central" symptoms, given some assump-

tions, as promising targets of intervention (Epskamp, van Borkulo, et al., 2018) but these methods generally fail to account for the fact that symptoms differ in their amenability to psychological treatment or that some symptoms may have "low centrality" but remain critical targets for intervention because of their impact on psychosocial functioning (e.g., suicidal thoughts and behavior; Fried & Nesse, 2014; Proctor et al., 2009). Further, currently available techniques to estimating personalized networks are primarily of *exploratory* nature and do not allow clinicians to incorporate relevant a priori knowledge or clinical expertise. By failing to see their ideas reflected in network models, practitioners might consider them as impractical and not in line with their clinical view, likely resulting in hesitancy towards using personalized network models. Indeed, a recent study has shown that case formulations greatly differ from temporal networks estimated from ESM data (Frumkin et al., 2021).

Based on these considerations, we argue that providing trainings and guidelines is necessary, but not sufficient in implementing the complex system approach in clinical practice. For methods to be regarded clinically relevant, it is vital that tools have the flexibility to be guided by clinical needs and allow practitioners to incorporate clinical considerations.

12.1.2 Theories versus data models

In recent literature, special attention has been paid to disentangling conceptual aspects of *data models* and *theories*. According to Haslbeck, Ryan, Robinaugh, Waldorp, and Borsboom (Haslbeck, Ryan, et al., 2021), data models (e.g., a mean, correlation, or idiographic network model) are merely ways of representing or organizing data, often with the aim of establishing a phenomenon: a robust, generalizable feature of the world identified through empirical regularities (Bogen & Woodward, 1988; Haslbeck, Ryan, et al., 2021). In contrast, the aim of a theory is to explain a phenomenon by representing those aspects of the real world that give rise to the phenomenon. Whereas verbal theories are expressed in language, formal theories are expressed in mathematical equations or a computational programming language. This level of specification allows formal theories to simulate theory-implied system behavior, and by observing the effects of simulated interventions, we can draw conclusions about how the real-world system we are targeting would respond to a given treatment (a process referred to as "surrogative reasoning", cf. Swoyer, 1991).

In the following, we will refer to the approach of translating (verbal) case formulations into mathematical systems as the *formalization of idiographic theories*. Although the term "theory" is commonly used to describe phenomena on the nomothetic level, in this chapter, we are focused on the explaining phenomena at the level of the individual client, and will use the term "idiographic theory" in respect to theorized relations within one individual.

12.1.3 Formalizing idiographic theories

To bridge the gap between methodological advances and practical application of the complex system approach, we propose to derive dynamical system models directly from clinical theory, clinicians' expertise and case-specific knowledge. Formalizing client systems tackles the mismatch between technical tools and target language as discussed above at its core; that is, rooting dynamical systems *in the language of practitioners* allows examining the client's system behavior based on clinically relevant considerations.

In other scientific disciplines like biology (Furusawa & Kaneko, 2012), ecology (Scheffer et al., 1993), and political science , it is common to model dynamic processes based on theory and/or knowledge. Unfortunately, the application of formalized theories in mental health research is to date extremely rare. Recently, there have been efforts to propose formal theories in psychiatry, including the relationship between client and clinicians (Körppen et al., 2011; Liebovitch et al., 2011) and models of burnout (Dujmić et al., 2019; Von Kentzinsky et al., 2020), addiction (Grasman et al., 2016), and panic disorder (D. Robinaugh et al., 2019). However, much remains unknown about precisely how such formal theories should be developed and how they should be used in psychotherapy. The main objective of this chapter is to take a step towards addressing this gap in the literature by demonstrating the potential of formalizing idiographic theories in clinical practice and illustrating an approach to formalizing such theories using the framework of functional analysis.

12.1.4 Approaches to constructing idiographic systems

We see two main ways of constructing personalized dynamical systems in psychopathology: First, modeling a *generic* disorder model, and subsequently personalizing the model through estimating control parameters for the equations in the system (*top-down approach*, cf. Robinaugh et al., 2019), and second, modeling relations between *specific* variables directly for and with each client (*bottom-up approach,* cf. Schaub & Schiepek, 1992; Schiepek, 2003; Strunk & Schiepek, 2006). An advantage of the former approach is that it allows modeling individual differences between clients regarding the strength of *shared* relations (e.g., person-specific tendencies to avoid when confronted with fear), which consequently allows for examining for instance tipping points in fear responses following maladaptive coping. An advantage of the latter approach is that it allows to flexibly model any psychological hypotheses, as well as individual problems and resources (Sim et al., 2005).

The method outlined in this chapter is based on the framework of functional analysis, and therefore utilizes elements of both approaches: On the one hand, functional analysis constitutes a generic framework for case formulation (*top-down* elements); on the other hand, it also provides the flexibility to integrate client-specific problems and resources (*bottom-up* elements).

12.1.5 The role of computational models in bridging the scientist-practitioner gap

We argue that formalizing idiographic theories provides advantages for both, clinical *practice* and mental health *research*, schematically displayed in Figure 12.1, and is promising in bridging the gap between the two.

First, computational models of idiographic theories can be used to advance the current practice of a client's case formulation. Sim, Gwee, and Bateman (2005) identified five key advantages associated with formulating thorough case formulations in clinical practice: (a) the integration/relation of multiple problems of a client, (b) the explanatory nature of the resulting model, (c) the prescription of interventions, (d) the prediction of outcomes, and (e) the support for the therapeutic relationship. Schiepek and colleagues (2003; 2017) pioneered the integration of case formulation and idiographic system modeling and argued that these key advantages could be strengthened through computational models. Clinicians are required to make more rigorous decisions in specifying relations in the case formulation, which makes the formalization of idiographic theories a promising avenue to foster more scientific practices in designing client-tailored treatment. This reasoning is in line with

a growing body of literature indicating the need for more rigorous theory development in clinical and social sciences (Borsboom, van der Maas, et al., 2021; Fried, 2020; Guest & Martin, 2021; Haslbeck, Ryan, et al., 2021; D. Robinaugh et al., 2019; van Rooij & Baggio, 2021). The left part of Figure 12.1 illustrates how computational modeling can inform case formulations in clinical practice: Formalizing a case formulations results in a computational model that allows the clinician to subsequently simulate data, given the specified idiographic system. Based on these simulations, it is possible to compare theoretical implications to phenomena observed in clinical practice and to evaluate and adapt theory accordingly (Epstein, 2008; Haslbeck, Ryan, et al., 2021; Smaldino, 2017). Theory formation can thus be adapted by examining *what a theory implies*, and these implications only become fully apparent once a theory is formalized and data can be simulated.

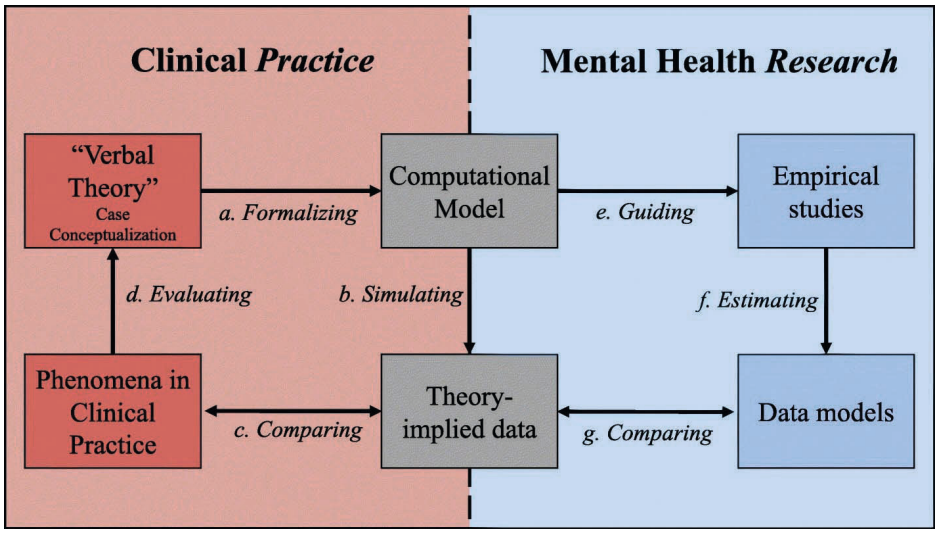


Figure 12.1. The role of computational modeling in bridging the scientist-practitioner gap. Schematic illustration of computational modeling (the product of formalizing a theory), at the intersection of clinical practice and mental health research. Computational models allow us to evaluate case formulations in clinical practice (a-d), and bring clinical theories closer to empirical studies through guiding choices crucial to the estimation of and inferences drawn from data models (b, e-g).

Second, computational models bring clinical theories closer to empirical *research*. For instance, prior to empirically studying a client's systems, the researcher needs to determine variables to include into the analysis. This question is of great importance in network estimation, since parameters in partial correlation networks are heavily dependent on the set-up of variables. The choice of variables has a crucial impact on network estimation and inference, especially if clinically relevant variables are missing, or if included variables stem from theoretically similar constructs, indicating topological overlap (Fried & Cramer, 2017). Formalizing theories can provide useful information regarding the set-up of variables needed to retrieve clinical phenomena. Further, empirical research is often confronted with practical constraints to assessing psychological processes. Many clinically relevant psychological processes are difficult – sometimes even impossible – to assess on their appropriate time scale. For practical reasons, variables are often measured within the same time scale (usually

once a day or every few hours), potentially leading to biased estimates in dynamical models. A recent simulation study suggests that using the most commonly applied ESM time intervals results in data models that are largely unable to recover the micro dynamics of a system (Haslbeck & Ryan, 2022). A stronger focus on theory and the utilization of clinical knowledge could therefore be helpful in informing relationships in the estimated model that cannot reasonably be captured by commonly used ESM data. The right part of Figure 12.1 illustrates how computational modeling can guide mental health research, resulting in data models that are grounded in theory-based considerations. The resulting data models can be compared against theory-implied simulation results and guide further theory development (Haslbeck, Ryan, et al., 2021) as well as future research design planning.

12.1.6 Example of a computational model: Functional analysis of client with panic disorder

In the remainder of this chapter, we will introduce and evaluate an example system based on *functional analysis* (sometimes referred to as *applied behavior analysis* or *SORKC model*; Cheney, 2017), a framework commonly used by clinicians to formulate case formulations in CBT. Functional analysis explains maladaptive behavior in terms of classical and operant conditioning processes: a discriminant stimulus (*Sd*) evokes specific emotional, cognitive and behavioral responses in the client (*Re*, *Rc*, and *Rb*, respectively). Persistent dysfunctional coping is explained through the presence of reinforcing stimuli. In the short term, dysfunctional coping mostly yields positive effects (*perceived benefits*), while on the long term, negative effects (*perceived costs*) are accumulating.

To illustrate, we are modeling the case formulation of a hypothetical client suffering from panic disorder. This example client experiences unusual bodily sensations (arousal) in the cinema and concludes that she will have a heart attack and that there is no chance she can get medical assistance on time. The experience of heart racing in the cinema constitutes her *discriminant stimulus* (*Sd*). Her emotional response is panic (*Re*), due to catastrophic interpretations of the heart racing ("I am having a heart-attack"; cognitive response, *Rc*). In order to cope with the aversiveness of this situation, she leaves the cinema (*Rb*). This behavior yields benefits: The client manages to decrease the intense fear she felt in the cinema (*perceived benefits*). However, constant avoidance also leads to costs: The client withdraws herself socially and experiences problems at work due to her avoidant coping in panic-evoking situations (*perceived costs*). Further, she is faced with a lack of falsification possibilities, increasing the *credibility of her catastrophic thoughts* in confrontation with experiencing heart racing while not being able to get medical assistance.

Figure 12.2 shows a schematic summary of the main factors involved in the client's functional analysis, as typically documented in psychotherapy. Robinaugh and colleagues (2019) proposed a computational model for panic disorder. While Robinaugh et al. focus on the generic approach described above, we also include personal factors as components in the model, in accordance with the principles of functional analysis. As will be discussed later on, client-specific reinforcing factors can be modeled through both, extending equations and altering parameters in the system.

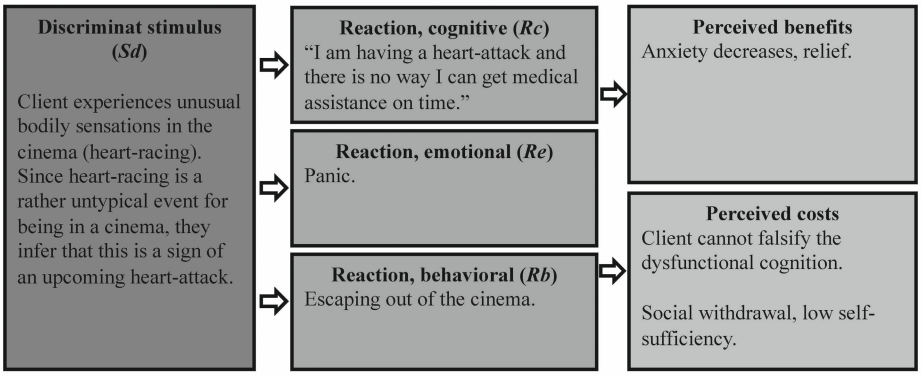


Figure 12.2. Functional analysis of hypothetical client suffering from panic disorder. Case formulation of our example client using the framework of functional analysis, as commonly documented in clinical practice. A discriminant stimulus leads to cognitive, emotional, and behavioral reactions (*Rc*, *Re*, *Rb*, respectively). The behavioral reaction has perceived benefits and costs, reinforcing or inhibiting the behavior

12.2 Methods

In the following, we describe the general approach to formalizing idiographic theories, using the functional analysis of our hypothetical client. To facilitate readability, we focus on introducing the process on a *conceptual* basis. We advise the reader interested in technical detail to consider the supplementary material. Note that the simulation results and the discussion can be followed without having read the mathematical background section.[40]

In many formal theories, including the one that will be presented here, every component of the system is expressed as a *differential equation*, precisely explicating the specific influences of system variables on one another. Intuitively, differential equations can be understood as specifying the rate of change in a given variable (i.e., how a given variable will change over time), as a function of itself and other causally related variables. For instance, in the simplest case of a first-order derivative, the differential equation of the variable *avoidance* captures the extent to which avoidance behavior will increase or decrease moving forward from a given time point. Since our system predicts that avoidance is employed as a consequence of anxiety, the corresponding differential equation would encode that high levels of anxiety increase the first-order derivative (the momentary change) of avoidance.

Note that in this chapter, we primarily focus on modeling *linear* differential equations. Extending the framework to including *non-linear* equations would be a relevant step for future research, given that prior literature found that psychotherapeutic processes are often chaotic, a feature that is characteristic for non-linear dynamics (Schiepek, 2009; Schiepek et al., 2017). For the sake of implementation, however, we decided to focus on linear equations, since many aspects of non-linear dynamics require an extensive mathematical understanding. We will discuss the difference between these approaches and the impact on predictions in systems later on.

40 The R-code to reproduce the simulations can be found in an OSF repository: https://osf.io/spb37/ .

12.2.1 Procedure of formalizing idiographic theories

12.2.1.1 Step 1: Schematic representation

Prior to formulating differential equations, we recommend visualizing the system schematically. This facilitates specifying relations in the equations later on. A graphical depiction of the relations in the client's functional analysis, including the target nodes of the interventions introduced below, is presented in Figure 12.3. This is a crucial step, since it opens the search horizon beyond the given boundaries of functional analysis (i.e., allowing to incorporate person-specific elements into the system, such as competencies and resources), and requires the clinician to explicate relations between the variables.

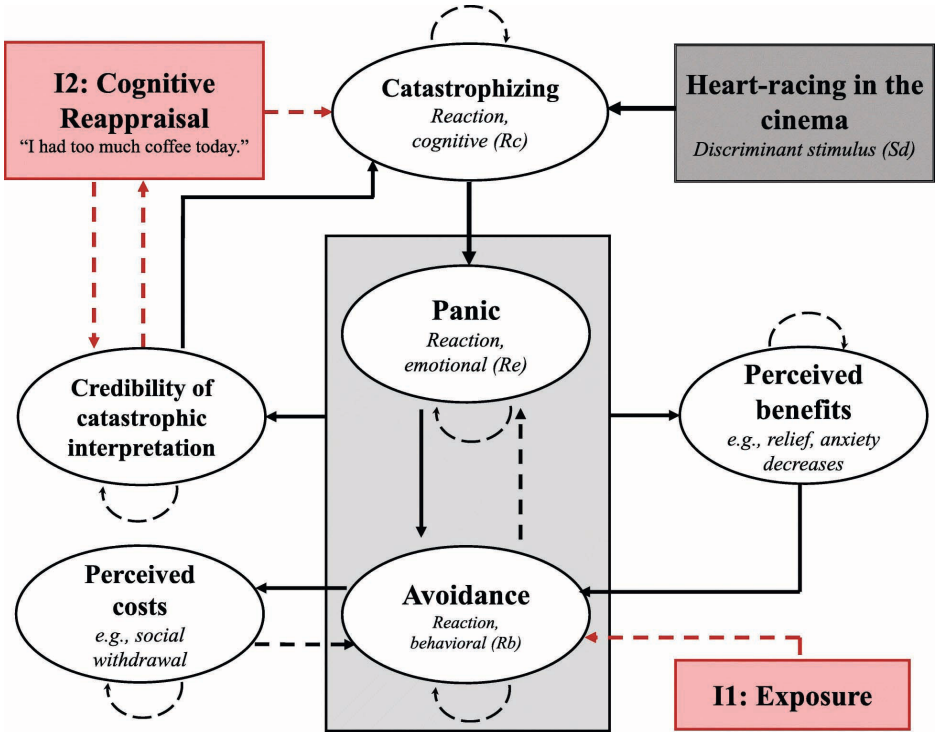


Figure 12.3. Schematic representation of the functional analysis. Theoretical relations adapted from the client's functional analysis as a basis for deriving the system equations. Anxiety (*Re*) is reduced through applying avoidance behavior (*Rb*). In addition, avoidance behavior is reinforced through *perceived benefits* and inhibited through *perceived costs*. Persistent avoidant behavior increases the *credibility of catastrophic interpretations*, in turn leading to more *catastrophizing* during exposure. We formalized and tested three interventions, *exposure*, *cognitive reappraisal*, and their combination, represented through the red boxes.

12.2.1.2 Step 2: Deriving differential equations

Based on the schematic representation of the client's functional analysis, we formulated differential equations for each component in the system. Practical guidelines for defining dynamical systems from both theory and data have been recently described elsewhere (Chow, 2019).

As a starting point, we modeled catastrophic interpretations (*Rc*) of the discriminant stimulus (*Sd*) as input for the occurrence of panic (*Re*); heart racing in the cinema leads to the catastrophic idea that this is a sign of an upcoming heart attack, and the client consequently experiences panic symptoms. In turn, the client copes through avoidance behavior. We modeled coping behavior using equations commonly applied to model the dynamics between prey and predator populations in ecology (Wangersky, 1978). In our model, panic (*Re*) is analogous to "prey" and avoidance (*Rb*) is analogous to "predator". Thus, increases in panic give rise to increases in avoidance behavior, while increases in avoidance behavior lead to lower panic.

Avoidant coping is modulated through the presence of reinforcing/inhibiting factors. First, if the client perceives avoidance to be effective in decreasing panic (i.e., experiencing relief; *perceived benefits*), her tendency to cope through avoidance increases. Second, avoidance behavior comes with detriments for the client, for instance social withdrawal or potential problems at work. These detriments (*perceived costs*) are theorized to have an inhibiting effect on the client's avoidance behavior. Third, persistent application of avoidance behavior comes with a lack of opportunities to falsify the catastrophic interpretation. Therefore, we modeled increasing *credibility of the catastrophic interpretation* as a consequence of avoidant coping. The credibility of the catastrophic interpretation increases the client's tendency to catastrophize in confrontation with the discriminant stimulus.

12.2.1.3 Step 3: Formalizing interventions

One of the main advantages of computational modeling in clinical practice is that interventions on a system can be examined in silico, and their effects evaluated on the basis of a case formulation. Note that the simulated effects are dependent on the accuracy of the model, highlighting the importance of theory evaluation (Haslbeck, Ryan, et al., 2021). We will discuss future avenues for systematic evaluations later on.

Similar to step 2, interventions need to be formalized. We modeled two commonly used interventions in CBT: *exposure therapy* and *cognitive reappraisal*. First, we implemented exposure through setting avoidant coping to 0. Second, cognitive reappraisal was implemented through formalizing another system variable, capturing the credibility of an alternative *functional* interpretation of heart racing. The credibility of the functional interpretation was theorized to "compete" with the credibility of the catastrophic interpretation, and we thus formalized the former as an inverse function of the latter; if the *functional* interpretation of the stimulus increases, the *dysfunctional* interpretation decreases and vice versa. This change in interpretation of the stimulus influences the extent to which the client catastrophizes. We therefore extended the equation for catastrophizing with an inhibitive term; increasing the credibility of the functional interpretation (e.g., "I simply had too much coffee") leads to less catastrophic interpretations of the discriminant stimulus.

12.2.1.4 Step 4: Choosing initial values of system variables and parameters

Prior to conducting simulations, initial values of each system variable and parameters need to be defined. In contrast to many data-driven approaches of estimating networks, these values are difficult to interpret numerically. This is because formalizing idiographic theories does not require the clinician to *operationalize* variables, since these will not (necessarily) be *measured*. The units

of system variables are therefore not meaningful. We will discuss advantages and disadvantages of aligning theory components with the measurement procedure later on.

In contrast to common parameter estimation techniques in data models, the approach outlined in this chapter treats parameters in formalized theories as "tuning-knobs" to tailor the relations towards the client's case until theory-implied behavior resembles phenomena of interest. For instance, one can increase the parameter encoding the extent to which avoidance behavior follows panic, if it is known that the client has a strong tendency to employ avoidance behavior as coping. Further, one can vary values of parameters to examine differential effects of *unknown* relations; for instance, clinician and client can collaboratively examine the effects of different parameter choices for catastrophizing leading to panic. This allows clients to experimentally examine the responses of their system towards alterations.

For our example model, we chose parameters and initial values of the variables according to a qualitative examination of the system behavior, i.e., through adjusting parameters until the system resembled behavior to be expected given the information on the case of our hypothetical client. The choice of parameters and initial values can be found in the mathematical appendix, alongside all differential equations used in the simulations.

12.2.1.5 Step 5: Simulating and visualizing theory-implied data

Following the system specification, we can simulate and visualize data. We provide the code to reproduce our analysis and plots in R. System data is commonly visualized in time series plots and phase portraits. Time series plots indicate the time trajectories of all system variables, with time on the *x*-axis and variable levels on the *y*-axis. Phase portraits are useful to display the relationship between two or three variables over time. Each variable is represented on an axis, and following the trajectory in the phase portrait gives us information regarding the time course of the displayed variables. To illustrate, we used the example of three-dimensional phase portraits, indicating the relationship between *panic*, *avoidant coping*, and *the credibility of the catastrophic interpretation*.

12.2.1.6 Step 6: Evaluating case formulations

In a last step, the simulated ("theory-implied") data can be compared to phenomena observed in clinical practice. Differences between simulated data and observed patterns can be an indication that specific system relations need to be adapted or that important variables are missing in the system (Epstein, 2008; Smaldino, 2017). As illustrated in Figure 12.1, these considerations can be important pointers for setting up empirical investigations of symptom dynamics (e.g., which variables to include in an ESM study). We will address formal aspects of theory evaluation in the "Discussion" section.

12.3 Results

12.3.1 Scenario 1: System behavior without intervention

Figure 12.4 shows time series plots and phase portraits for the simulated system behavior without intervention. Being confronted with the discriminant stimulus led to a rapid increase in catastrophizing, followed by panic. Over time, avoidance behavior gradually built up as a coping mech-

anism. While this was associated with a momentary decrease in panic, persistent avoidance was also accompanied by increasing credibility of the catastrophic interpretation, in turn leading the client to catastrophize even more when confronted with the discriminant stimulus. In the short term, avoidance behavior was mainly associated with benefits, while in the long term, the perceived costs built up. The three-dimensional phase portrait shows that persistent avoidance behavior did not allow the client to decrease panic states in the long term. Instead, panic tendencies manifested as a function of the credibility of the catastrophic interpretation. A clinical interpretation could be that the client was not able to falsify catastrophic interpretations due to the lack of exposure to the discriminant stimulus.

12.3.2 Scenario 2: Behavioral therapy (*exposure*)

Figure 12.5 shows the time series plots and phase portrait when applying exposure. This intervention led to a sudden increase in panic states in the short term. In the long term, panic decayed even under absence of avoidant coping, accompanied by a decrease in catastrophizing and credibility of the catastrophic interpretation, demonstrating the effectiveness of behavioral therapy for our client. With the introduction of exposure therapy, the perceived benefits of avoidance behavior disappeared, e.g., the client could not experience relief through avoidance anymore, and the associated costs decayed over time.

12.3.3 Scenario 3: Cognitive therapy (*cognitive reappraisal*)

Figure 12.6 shows the time series plots and phase portrait when applying cognitive reappraisal. While functional interpretations of the discriminant stimulus could help decreasing panic tendencies, avoidance behavior only decreased after the functional interpretation gained sufficient credibility. Additionally, catastrophizing and the credibility of the dysfunctional cognition decreased, while avoidance behavior gave rise to both, the perceived costs and benefits.

12.3.4 Scenario 4: Cognitive behavioral therapy (*exposure + cognitive reappraisal*)

Figure 12.7 shows the time series plots and phase portrait when applying exposure and cognitive reappraisal simultaneously. Similar to scenario 2, this combination of interventions led to an increase in panic tendencies in the short term. The introduction of the functional interpretation of the discriminant stimulus was accompanied by a decrease in catastrophic interpretation and its credibility, ultimately leading to a decrease in panic tendencies. Similar to scenario 2, confrontation led the associated benefits of the behavior to disappear and the costs to decay over time.

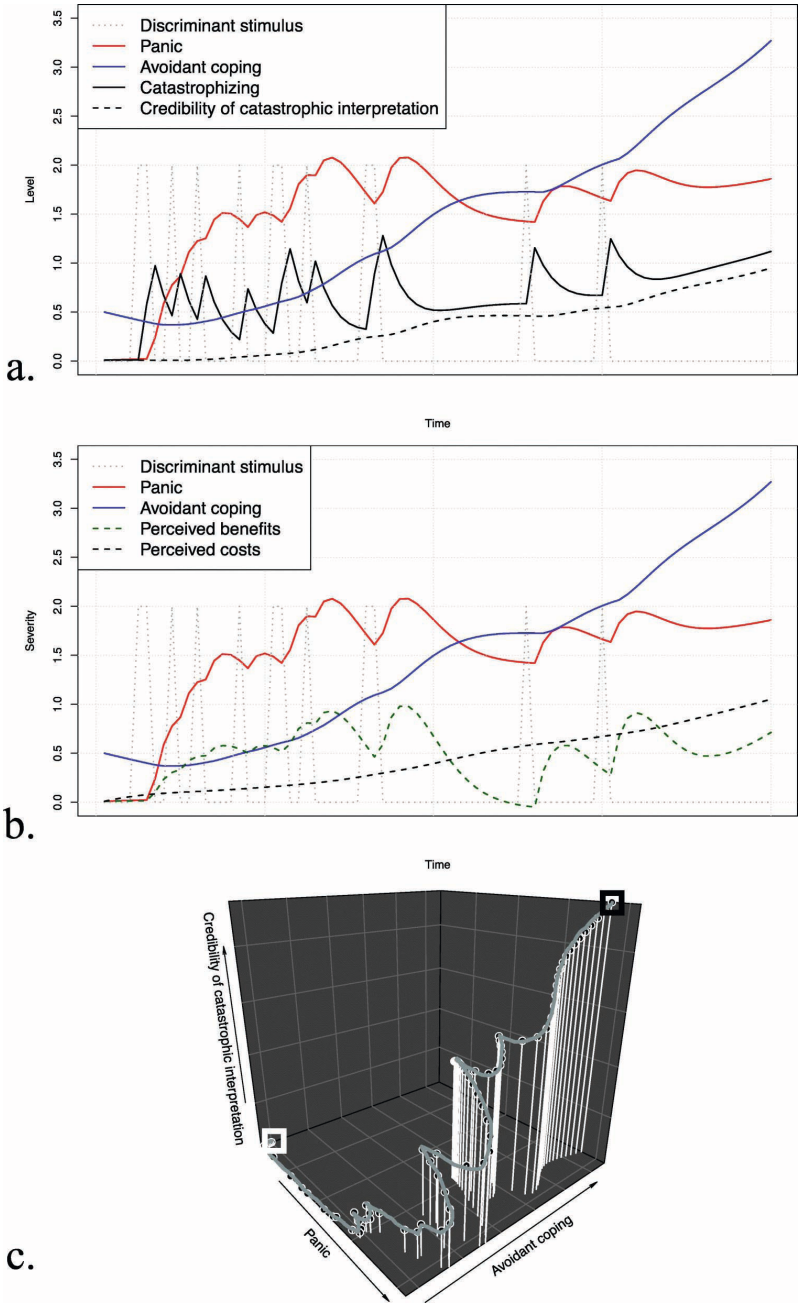


Figure 12.4. Simulation results of scenario 1 (no intervention). The top and middle parts show the simulated time series for the discriminant stimulus, panic, and avoidant coping along with **a** catastrophizing and the credibility of the catastrophic interpretation and **b** perceived benefits and costs. The bottom part of the figure (**c**) shows the three-dimensional phase portrait for panic, avoidant coping, and the credibility of the catastrophic interpretation, where the white box indicates the start and the black box the end of the trajectory.

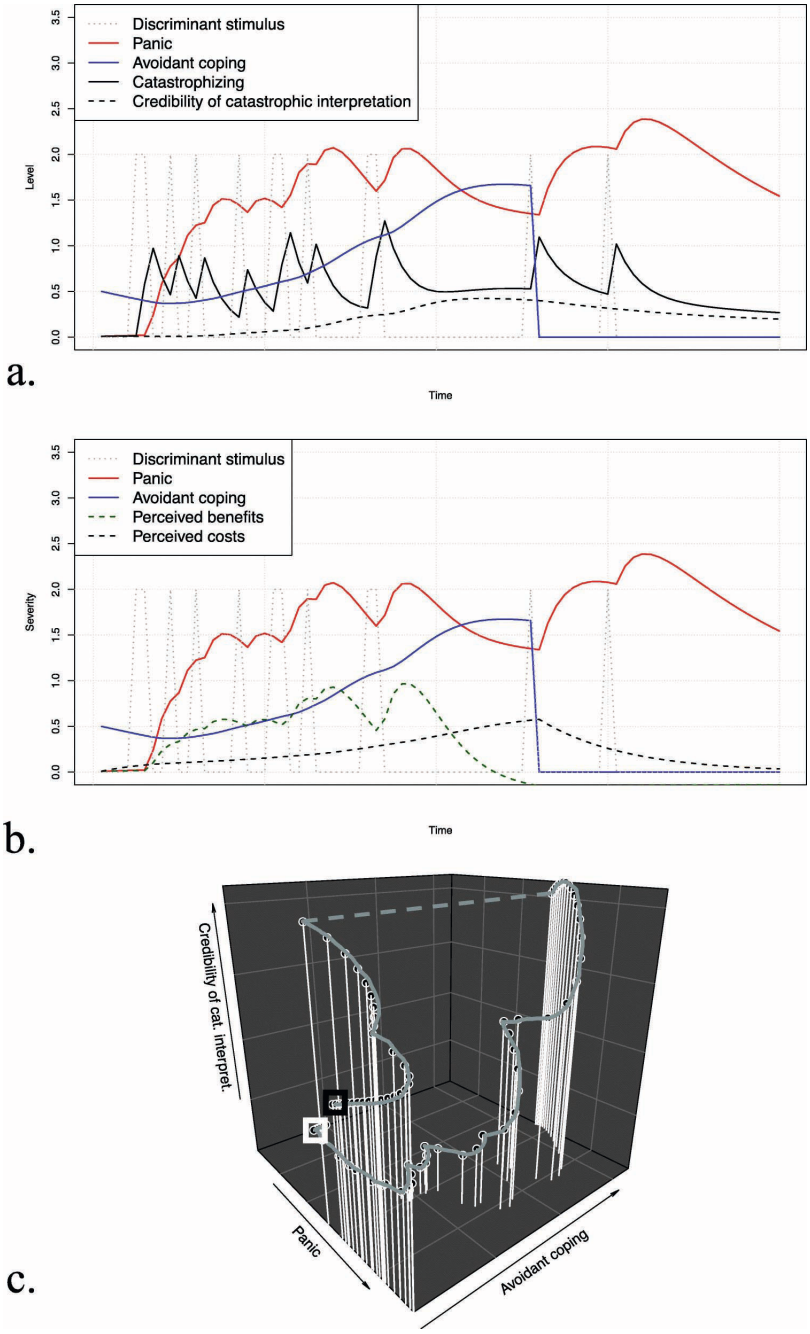


Figure 12.5. Simulation results of scenario 2 (exposure; behavioral therapy). The top and middle part show the simulated time series for the discriminant stimulus, panic, and avoidant coping along with catastrophizing and the credibility of the catastrophic interpretation (**a**) and perceived benefits and costs (**b**). The bottom part of the figure (**c**) shows the three-dimensional phase portrait for panic, avoidant coping, and the credibility of the catastrophic interpretation, where the white box indicates the start and the black box the end of the trajectory.

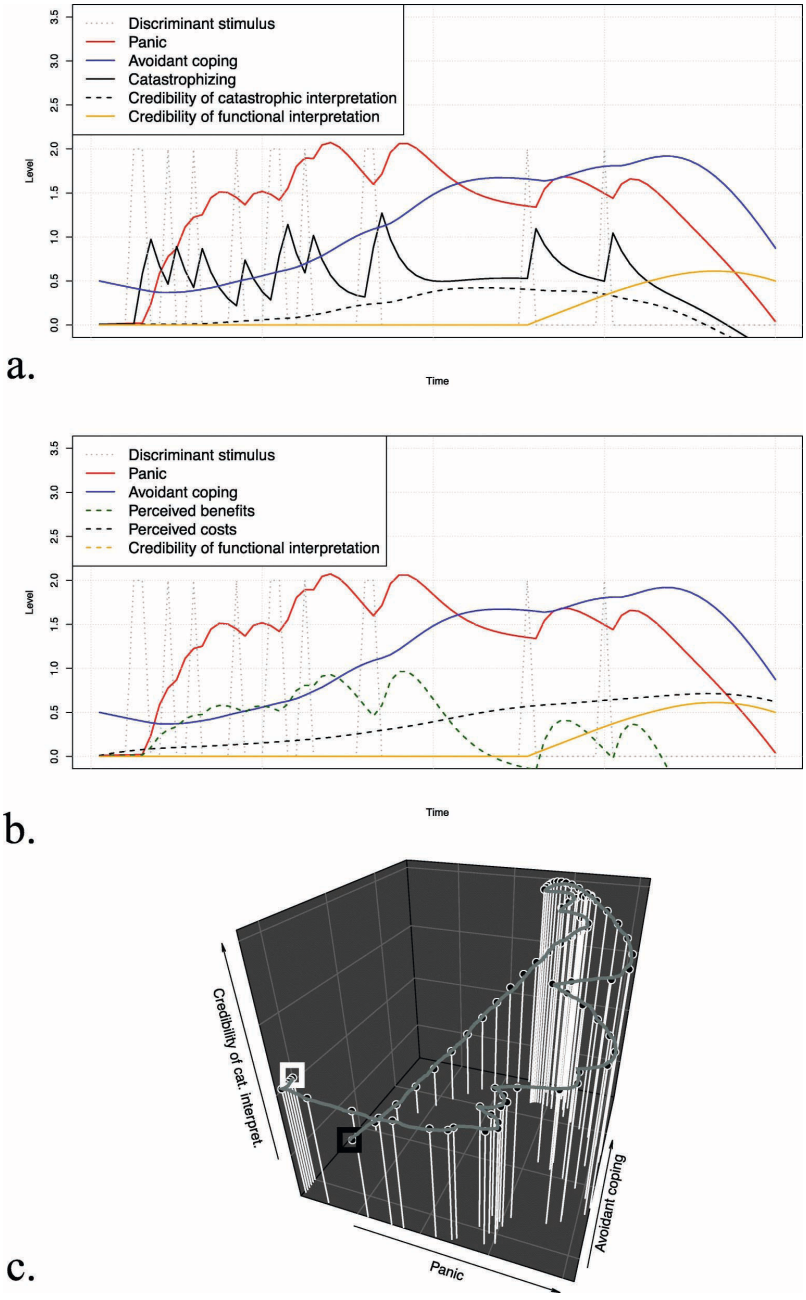


Figure 12.6. Simulation results of scenario 3 (cognitive reappraisal; cognitive therapy). The top and middle part show the simulated time series for the discriminant stimulus, panic, and avoidant coping along with catastrophizing and the credibility of the catastrophic interpretation (**a**) and perceived benefits and costs (**b**). The bottom part of the figure (**c**) shows the three-dimensional phase portrait for panic, avoidant coping, and the credibility of the catastrophic interpretation, where the white box indicates the start and the black box the end of the trajectory.

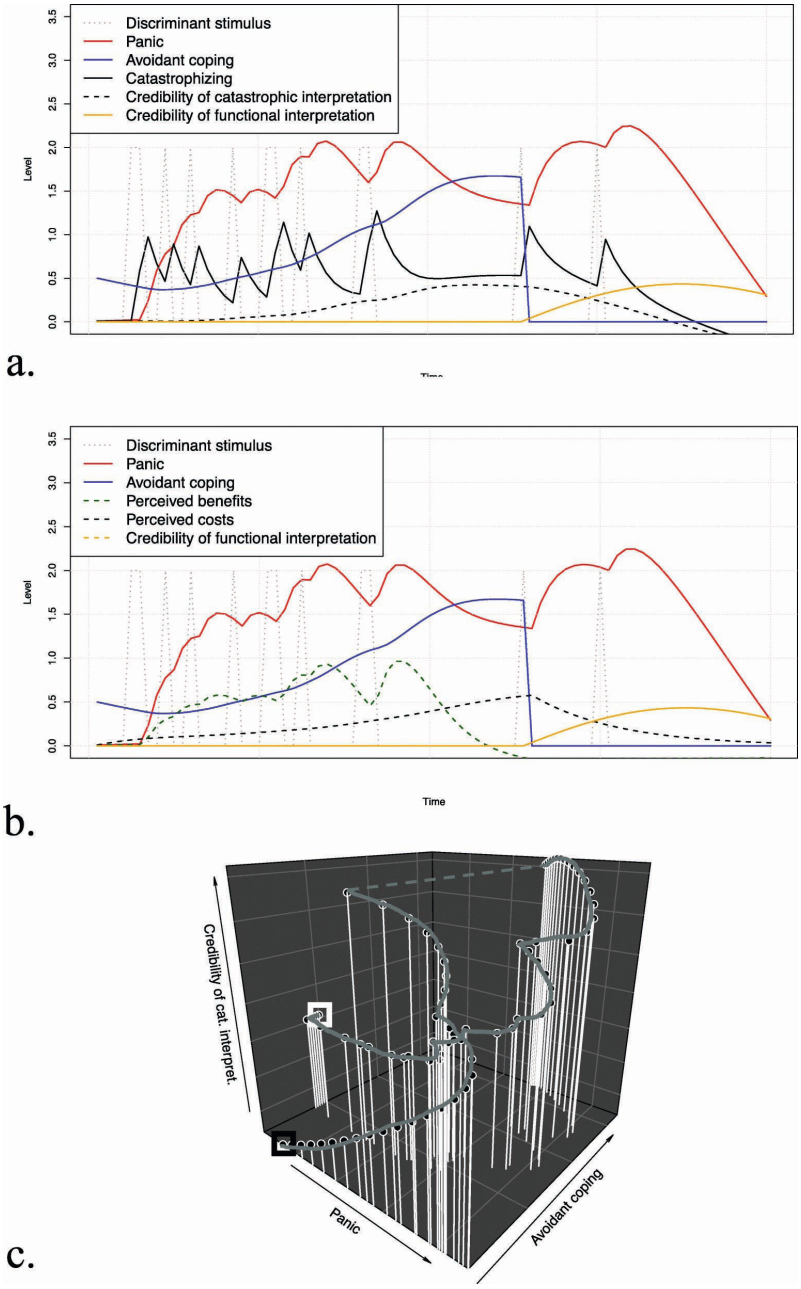


Figure 12.7. Simulation results of scenario 4 (exposure and cognitive reappraisal; CBT). The top and middle parts show the simulated time series for the discriminant stimulus, panic, and avoidant coping along with catastrophizing and the credibility of the catastrophic interpretation (**a**) and perceived benefits and costs (**b**). The bottom part of the figure (**c**) shows the three-dimensional phase portrait for panic, avoidant coping, and the credibility of the catastrophic interpretation, where the white box indicates the start and the black box the end of the trajectory.

12.4 Discussion

Current movements in psychotherapy strongly align with technical advances in dynamical modeling tools—yet their implementation in clinical practice is rather scarce. To bridge this gap, we call for a stronger focus on tools that make use of *frameworks and theories* embedded in clinical practice. In this chapter, we discussed the formalization of idiographic theories, through the use of differential equations, as an alternative to data-driven network modeling approaches. Our main objective for promoting the use of formalized idiographic theories is that data models cannot always account for considerations relevant to clinical practice. In consequence, even though techniques seem to be promising in analyzing client data, their implementation might be hampered due to the lack of options to incorporate theoretical and practical considerations. This barrier can be addressed through grounding dynamical systems in the theories of practitioners. Differential equations are commonly used in a variety of other scientific fields to describe systems, and are a promising avenue for formalizing theories of mental health.

To illustrate this approach, we formulated a computational model based on dynamics of the functional analysis for a client suffering from panic disorder and examined implications for the case formulation and the effects of commonly applied CBT interventions. The results of the simulations are largely congruent with phenomena observed in clinical practice and in line with predictions of other theoretical frameworks. In the following, we discuss further benefits for clinical practice, concrete examples for theory adaptation, and future directions.

12.4.1 Benefits for clinical practice and clinical relevance

We identify at least five benefits from formalizing case formulations in respect to challenges faced in clinical practice.

12.4.1.1 Scientific rigor

One of the main advances in mental health care over the past decades is its increasing focus on scientific practices. The introduction of the scientist-practitioner model (Baker & Benjamin, 2000) was an attempt to strengthen scientific practices in psychotherapy, for instance through theory-guided hypothesis testing. It became vital for designing client-tailored psychotherapy to formulate a testable theory regarding intervention effects. The case formulation is an example of a framework for such scientific theories in clinical practice. However, if a theory is vague, the resulting hypotheses, predictions, and tests become scientifically questionable (Fried, 2020a). Especially in the current landscape of replicability issues (Pashler & Wagenmakers, 2012; Simmons et al., 2011), we see value in enhancing theory development through formalizing idiographic systems in clinical practice. As became evident in this report, especially when comparing the initial *verbal* theory in Figure 12.2 to the system of differential equations, the process of formalizing idiographic theories is mostly a process of increasing specificity, in which clinicians need to thoroughly reflect on and justify all relations between system variables.

12.4.1.2 Idiography

While the model used in this chapter uses concepts that are relevant for a broader range of clients suffering from panic disorder (*generic approach*), there are many individual differences in how exactly these relations should be specified. For instance, client A might have more exposure to their discriminant stimulus in their everyday life compared to client B, or client C has stronger avoidance tendencies than client D. These considerations can be reflected in altering the parameters in the system, aligning this approach with the idea of idiographic modeling. Further, specific components in the system can be added/removed, if applicable for a given individual. The framework of functional analysis is transdiagnostic in nature and can be applied to a broad range of disorders that involve dysfunctional coping, for example, substance abuse, post-traumatic stress disorder, obsessive-compulsive disorder, and depression.

12.4.1.3 Explanation

Functional analysis provides a framework that allows explaining the function of maladaptive behavior and helps understanding symptom maintenance. The explanatory character of these verbal theories can be advanced through formalization, since case formulations can subsequently be evaluated in respect to how well they can reproduce clinical phenomena (Epstein, 2008; Smaldino, 2017). If a case formulation fails to explain relevant phenomena, this will more easily be detected if data is simulated from a formalized case formulation, compared to a verbal theory.

12.4.1.4 Prediction

While computational modeling can foster the development of theoretical relations, it is also a useful tool for predicting theory-implied system behavior under given interventions. Most relevant for clinical practice, this allows the clinician to examine the effects of formalized clinical intervention in silico. Testing interventions in computational models offers efficient insight into intervention effects without having to collect data.

12.4.1.5 Didactics

Simulation outcomes of a formalized idiographic theory can be beneficial for didactics in clinical practice. First, visualizing the simulation results allows the clinician to collaboratively examine symptom dynamics with the client. This can be used in the process of psychoeducation, and communicating a treatment rationale, especially for interventions that might be aversive for the client (e.g., exposure). Second, in the long term, we see potential in implementing formalized idiographic theories to enhance more concise communication between clinicians through more rigorous documentation and visualization.

12.4.2 Theory evaluation of the example model

A main benefit to formalizing idiographic theories is that simulated data can directly be compared against expected/reported behavior in the client. One potential interpretation of discrepancies between simulated data and clinical phenomena is that the case formulation in its current form cannot account for potentially relevant clinical phenomena, for instance, if important relations or variables are missing. If this is the case, the clinician might want to adapt specific theoretical

relations until the simulated data adequately represents clinical phenomena. This is crucial when testing formalized interventions in a client's system.

In some aspects, the computational model presented in this chapter is congruent with clinical phenomena, while in other aspects theory adaptation might be needed. Note that the set-up of the simulation represents panic-symptomatology experienced by one hypothetical individual. Phenomena observed in simulations might differ if parameters are altered, which allows capturing individual differences in experiencing panic symptoms, and differences in treatment response. First, the simulations showed that for this client, persistent avoidance behavior is accompanied by increasing tendencies to catastrophize and increasing credibility of the catastrophic interpretation. This finding highlights the role of falsification in fear disorders; avoidant coping is associated with a lack of opportunities to falsify the catastrophic interpretation, subsequently leading to increasing tendencies to experience panic in confrontation with discriminant stimuli. Second, the simulations indicate that all interventions (exposure, cognitive reappraisal, and combination) are effective in decreasing panic tendencies for this client, which is in line with empirical studies testing the efficacy of CBT interventions for panic disorder (Barlow, 1997). Third, the simulation results showed that panic manifests in the long term, if no intervention is applied. This finding does not seem to adequately represent the experience of panic attacks, since these usually emerge rapidly and decline after a short amount of time. To account for this feature of panic attacks, we propose to model stronger decay of panic. Alternatively, one could conceptualize this variable as a *tendency* to experience panic in the presence of the discriminant stimulus, rather than the actual experience of panic itself.

12.4.3 Future directions

The approach of formalizing idiographic theories is still fairly new to clinical psychology, and there is a lot of research that needs to be conducted to help implementing it in clinical practice. In this section, we aim to give some directions for future research.

12.4.3.1 Systematic theory evaluation and testing

A crucial barrier for implementation is that the explanations and predictions provided by a theory need to be as accurate as possible, especially if the aim is to test formalized clinical interventions; such interventions will depend heavily on the accuracy of the model. We outlined that through comparisons of theory-implied and empirical data, systems can be evaluated to increase accuracy. Notably, any *systematic* comparison between theory-implied and empirical data models would require that variables used in *data collection* either directly map on to components in the *theory*, or that they can be precisely derived from those components. As outlined above, there are many elements in idiographic systems that are difficult to capture in common forms of data collection (e.g., ESM data), suggesting direct mapping of theory components to variables in empirical data may be difficult. Accordingly, it will be necessary for researchers to not only formalize theories, but also the auxiliary hypotheses about measurement that link the theory components to the variables in empirical data. In this chapter, we opted for modeling idiographic systems without restrictions to what can be operationalized and compared how well theory-implied data qualitatively resembles clinical phenomena based on expert discussions, but did not go through the process of formalizing our assumptions about measurement or deriving what should be expected in any given empirical data model.

Second, it needs to be noted that the origin of a potential mismatch between theory-implied and empirical data remains unknown. Such discrepancies can have a multitude of sources and can be ascribed to either shortcomings in the structure of the theory (e.g., missing crucial variables in the theory, mis-specified or missing relations between present elements of the theory), the set-up of the simulation (e.g., exact initial conditions, valid parameter values, input and boundary conditions), or shortcomings in empirical data collection and modeling (e.g., inappropriate modeling assumptions, measurement issues). Further, estimating parameters from non-linear time series data is often difficult and undergoes strong limitations (Gábor & Banga, 2015). We call for future research to investigate systematic ways of identifying the core of such discrepancies.

12.4.3.2 Technical expertise and effort

Another barrier to implementation is that, in the current practice of formalizing idiographic theories, constructing a series of differential equations to formalize a client's system can be immensely challenging and requires technical expertise that is not part of psychotherapy trainings. To address this issue, we propose that methodologists elaborate on a set of functions relevant to relations between clinical variables that can readily be used by clinicians to formalize idiographic theories. To enhance accessibility, this set of functions could be implemented in an interactive tool to visualize variable interactions. Clinicians could then pick from this set and construct formalized systems without the need for understanding the mathematical background in depth. Further, implementation would greatly benefit from a procedure that allows clinicians to formalize idiographic theories using graphical tools. Such tools could incorporate a simple three-step procedure: In a first step, clinician and client collaboratively specify variables and sketch relations between the variables. Second, they select the qualitative nature of these specified relationships from the aforementioned list. This step encompasses the derivation of differential equations adapted to clinical practice. Third, simulations are conducted and client and clinician can interpret and explore symptom dynamics given the case formulation and the differential effects of interventions.

12.4.3.3 Clinicians' skepticism and utility

Recent investigations suggest that clinicians are skeptical regarding the utility of idiographic assessment approaches, specifically regarding ESM data collection and modeling techniques (Frumkin et al., 2021; Zimmermann et al., 2019). While these surveys suggest that clinicians find idiographic data models to be generally intuitive and aligning well with their clinical reasoning, it was also found that clinicians are not always convinced that they can learn something new from idiographic data models. Further, recent studies suggest that there is little incremental information in time series measures beyond mean levels and general variability (Dejonckheere et al., 2019), and that time series effects show largely unacceptable reliability after partialling out redundancies with mean and variability (Wendt et al., 2020). It is important to note that these findings pertain to the utility of idiographic *data models*. As discussed above, these data models face several challenges in the clinical context (e.g., insufficient number of observations, time scaling, measurement artifacts, modeling assumptions), offering a potential explanation for the questionable performance of time series measures.

Formalized idiographic *theories*, on the other hand, aim to *explain* phenomena that can be observed in the client. They do so by representing the system posited to give rise to the phenomenon. We outlined how formalizing such systems can foster theory development and therefore potentially help clinicians gaining insight into the effects of (formalized) clinical interventions. Valid inferences from such intervention simulations require clinicians to thoroughly evaluate their theories, and formalizing theories can help in doing so. We argue that, if proof-of-principle studies can support the hypothesis that formalizing idiographic theories improve treatment planning, this could greatly benefit clinical practice. However, to facilitate implementation, future research should conduct surveys with practitioners to understand potential barriers of implementing formalized idiographic theories.

12.4.3.4 Linear versus non-linear dynamics

We introduced two perspectives in constructing idiographic systems: First, a *top-down approach* in which generic factors are modeled and subsequently personalized through adapting parameters, and second, a *bottom-up approach* in which personalized factors are modeled directly—extending the search horizon to incorporate any factor that can be related to the client's system. In the present chapter, we formalized a case formulation within the generic framework of functional analysis, using (primarily) linear equations. It is important to note that, especially when following the bottom-up approach of constructing idiographic systems for and with each client, system dynamics should encompass not only linear, but also non-linear dynamics. Indeed, prior research examining the quality of system dynamics found that processes in therapy are often non-linear and chaotic (Schiepek et al., 2011; Schiepek et al., 2017). Such dynamics are, by definition, hard to predict and are heavily dependent on the specific set-up of the simulation; slight changes in the set-up of initial conditions and parameters might have dramatic effects on the simulated behavior. In such cases, it may only be possible to make broad predictions about expected behavior, for example, not when a panic attack will occur, but rather whether a system is vulnerable to such attacks. We encourage future research to further investigate how such dynamics should precisely be incorporated in the formalization of theories.

12.4.3.5 Incorporating social and contextual dynamics

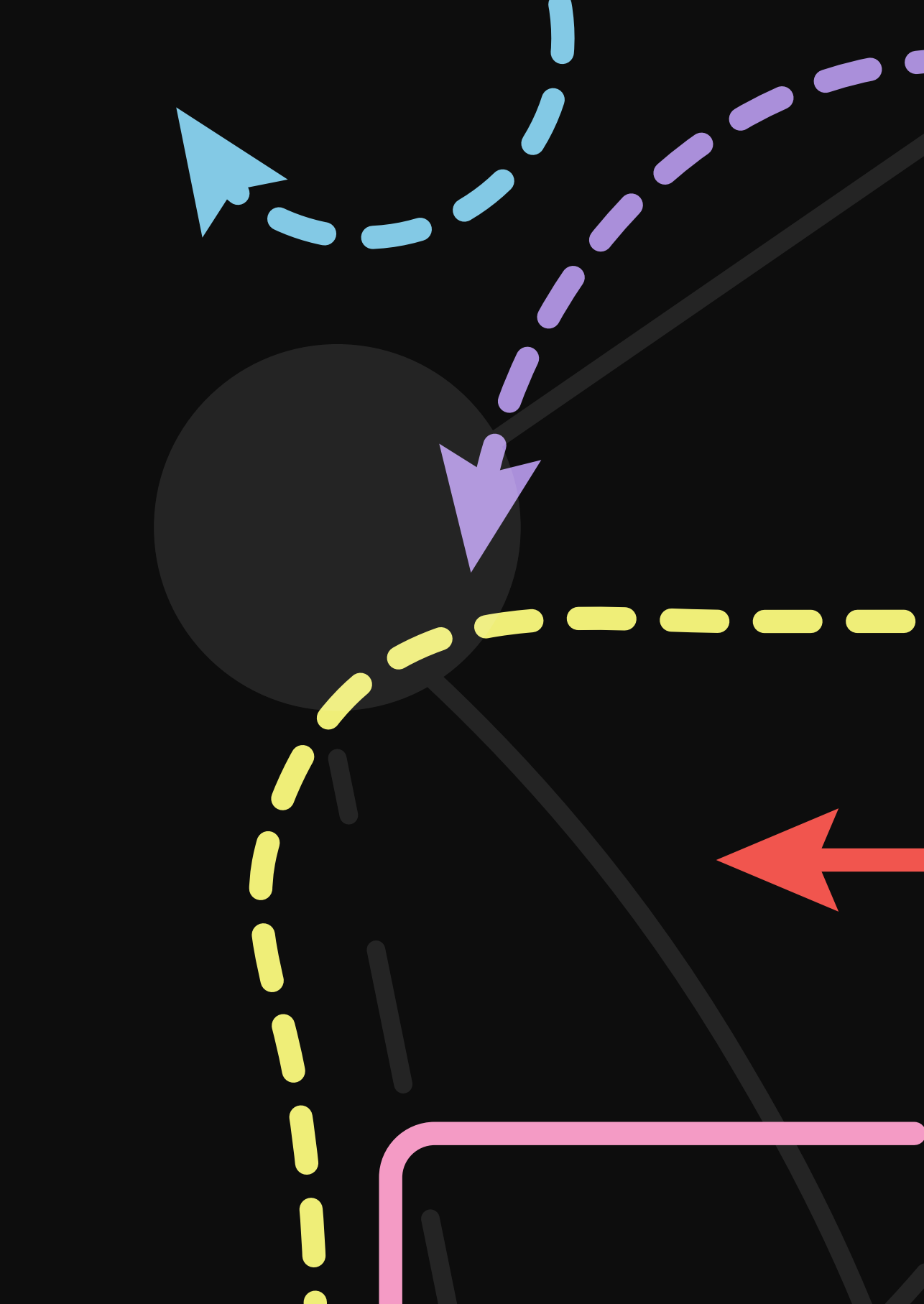
Computational models, as the one presented in this chapter, can account for processes that occur within an individual, and explain psychopathology on the basis of reinforcing factors. However, it seems unrealistic that these processes occur in isolation, independent from a social context. Indeed, clinical reasoning often includes the influence of the social environment on certain psychological processes, for instance, the link between avoidant coping tendencies and a certain attachment style, or the influence of peers in substance use. Incorporating interactions between different systems could open doors to model these clinical phenomena. Future research could use methods from agent-based modeling to simulate social interactions between client-specific computational models and investigate how these interactions can inform parameters or variables in the client's system.

12.4.3.6 Proof-of-principle

In order for new techniques to be considered relevant to clinical practice, they should provide practitioners with a clear incentive, and a main incentive for psychotherapy is to improve treatment outcomes. For many health care systems, case formulations form the starting point for hypothesis-driven intervention planning and execution. We expect that formalizing idiographic theories can improve the precision of intervention predictions, through enhancing explanatory and predictive precision in formulating case formulations; however, this idea needs empirical support. We hope that future research will follow up on this hypothesis and provide us with proof-of-principle studies validating the utility of formal theories in enhancing predictive precision of case formulations.

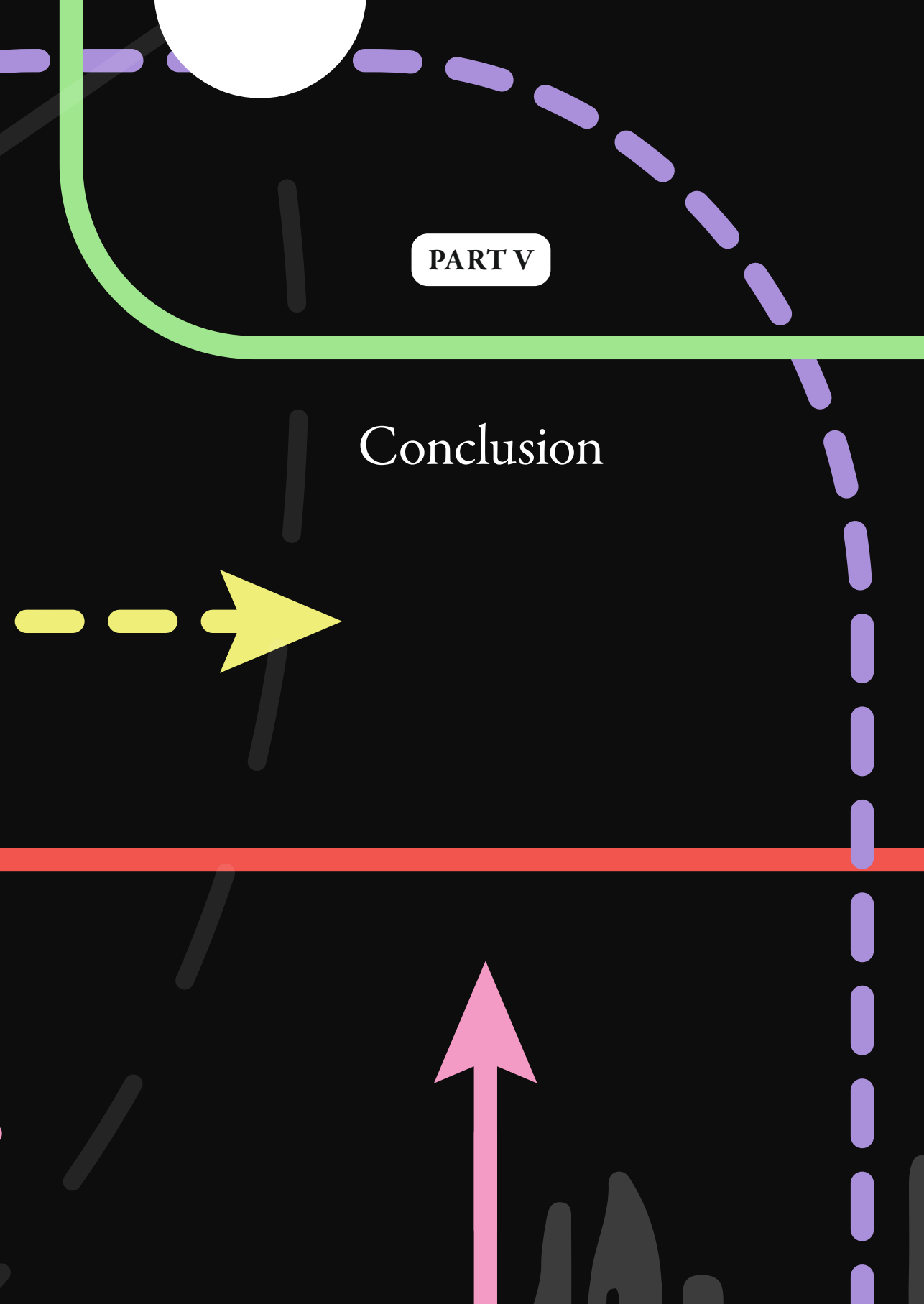
12.5 Conclusion

Complexity models are of great relevance for psychotherapy. Case formulations, even if only incorporating a small set of variables, can produce highly complex behavior. We present the formalization of idiographic theories through differential equations as an approach to align the movement of process-based psychotherapy to dynamical system methodology. Simulation results based on formalized theories can account for considerations that are vital to clinical practice. Furthermore, the process of formalizing a system promotes more scientific rigor in clinical practice and could help in improving explanatory and predictive precision of case formulations, as well as treatment planning.



PART V

Conclusion



CHAPTER 13

GENERAL DISCUSSION

In this final chapter, I will take a step back and discuss the findings of the thesis in light of future developments at the intersection of case formulation, network analysis, and simulation-based science. This thesis investigated current and developed new modeling approaches to advance the practice of case formulation, and applied the approaches to a variety of diagnoses, including generalized anxiety disorder, depression, eating disorders, obsessive-compulsive disorder (OCD), and panic disorder. To evaluate the contribution of the thesis to the main research goal, I will begin by summarizing the findings of each chapter, integrating findings with the current literature on theory formation, and answering the three overarching research questions for the main parts, *exploration – integration – formalization*. Based on this evaluation, I will highlight areas for future research, and draw conclusions for the overall research aim.

13.1 Thesis summary and answers to research questions

13.1.1 Part I: Methodological background

This thesis is strongly influenced by the network theory of mental disorders (Borsboom, 2017), which conceptualizes mental illness as arising from the causal interplay of psychological symptoms and related factors. Over the past decade, the statistical toolbox has been extended with methods to estimate statistical networks from multivariate data (Isvoranu et al., 2022). The first part of the thesis adds further methodological notes on conceptualizing research questions, creating research designs, estimating models, and reporting results in the field of network analysis. **Chapter 2** discussed the importance of relating longitudinal design choices to characteristics of the data. There are different types of longitudinal data (e.g., single measurement data, $N = 1$ time series data, $N > 1$ time series data, panel data) that can be used to estimate statistical networks. The characteristics of the data, along with the choice of the statistical model to analyze them, determine the precise interpretation of the resulting edges. In addition, interpretation of effects is dependent on specific characteristics of the research design, such as the assessed time scales. **Chapter 3** described how longitudinal networks can be estimated from $N = 1$ and $N > 1$ time series data, such as data collected via the Experience Sampling Method (ESM). The graphical vector auto-regressive (GVAR) model can be applied to this type of data to estimate contemporaneous and temporal effects, indicating associations between variables within the same time frame and over time, respectively. The same model can be estimated in a multi-level fashion for multiple individuals, resulting in fixed effects estimates, as well as between subject estimates for averages of person-wise means. There are specific challenges to this type of modeling technique, which include potentially unfeasible power requirements, and strong statistical assumptions such as stationarity. **Chapter 4** shed light on using network analysis for evaluating treatment effects. In the empirical literature, there is a wide variety of design and analysis choices to address this question, ranging from cross-sectional networks estimated from RCT data that include the treatment as a binary node, to personalized time series networks estimated before and after treatment. This systematic review highlighted the need for clear reporting and open science practices for using network analysis to evaluate treatment. **Chapter 5** introduced general reporting standards for cross-sectional network analysis for the most common research aims found in the empirical network analysis literature. The chapter also illustrated that reporting standards are not only important for scientific rigor in writing articles, but also for the

research design and planning phase, because they allow to make decisions based on anticipating the precise analytic challenges that may arise. To this end, this chapter has been used to develop a pre-registration template for cross-sectional network analysis.[41]

13.1.2 Part II: Exploration – Statistical networks based on empirical data

13.1.2.1 Summary of chapters

The second part of the thesis presented empirical network contributions that provide exploratory insights relevant for case formulations. **Chapter 6** presented a multi-level longitudinal network analysis of 1,368 individuals who completed 30 daily assessments on anxiety symptoms during the COVID-19 pandemic. The results showed that anxiety symptoms were especially well predicted by uncontrollability of worry, generalized worry, fear of being infected, fear of significant others being infected, and threat monitoring on the previous day. **Chapter 7** presented a multi-level longitudinal network analysis of the same population in the context of depressive symptoms. The main findings of this study were that depressive symptoms were mostly predicted by experiences of helplessness during the previous day, while within the same day, anhedonia, emotion regulation deficits, and lethargy, were most predictive. **Chapter 8** presented a cross-sectional network analysis of 724 older individuals. The study investigated between-subject relationships of depressive symptoms following two severe adverse life events, spousal loss and separation, and found that separated compared to bereaved individuals were more likely to experience an unfriendly environment and oneself as a failure. For both life events, the network showed strong relations with loneliness, which was in turn connected with a host of other depressive symptoms.

13.1.2.2 Answering research question and integration with literature

The three empirical chapters of this section generated exploratory insights that can inform case formulations of individuals who experienced symptoms of depression and anxiety during loss experiences or during the COVID-19 pandemic. Based on these examples, I will now address the research question "*How can statistical networks be used for exploration of symptom relationships, providing supporting insights for case formulations?*". First off all, there are general considerations for relating data models to case formulations, irrespective of the specific findings of these contributions. Recent literature brought attention to the fact that data models, in this context statistical networks, generally cannot represent important properties of theories, here case formulations (Haslbeck, Ryan, et al., 2021b). This is because data models are based on statistical assumptions and characteristics of the collected data that constrain the type of system properties they can represent. For example the frequency at which ESM beeps are administered determines the temporal effects that can be detected. Although caution is warranted for any type of statistical inference, it is important to be explicit about the issue specifically with networks, because the way these are usually visualized – as network plots – often resembles illustrations of diagrams constructed in many case formulation approaches, which in turn may suggest that statistical networks are indeed direct representations of theories, here case formulations (von Klipstein et al., 2020). Haslbeck and colleagues (2021)

41 The pre-registration form can be found in an OSF repository: https://osf.io/p9wn2/ .

suggest that an alternative route, one that treats data models as comparison points of models implied from data simulated from formal theories, may be more promising in appropriately using statistical models to develop theories, here case formulations.

I support this notion, as discussed in part IV, **chapter 12** (see Figure 12.1), however, in idiographic research, there may also be phenomena that are unique to the specific context a given individual is experiencing, and nomothetic theories may not cover these person-specific processes sufficiently (Zuidersma et al., 2020). In the clinical practice of case formulation, specifically in the approach spearheaded by Persons (Persons, 2012; Kuyken et al., 2009), a nomothetic theory is extended by an extensive assessment of idiographic processes. Establishing initial case formulations could therefore still benefit from exploratory findings in data models, if these can highlight the person-specific contextual factors that are otherwise not well represented in general clinical theories. As discussed in **chapter 2**, different data and study designs will result in different types of inferences that can be relevant to shed light on different aspects of a case formulation. For example, if a client presents with depressive symptoms after separating from their long-term partner, findings from **chapter 8** could inform an initial case formulation by highlighting differences in between-person symptom relationships to another client who experienced depressive symptoms following spousal loss. In contrast to these between-person findings, the results of **chapters 6** and **7** can deliver insights for clients who experience depression or anxiety symptoms in response to emergency situations, such as the COVID-19 pandemic.

In line with the cautionary notes mentioned above, it is important that findings of the empirical chapters are not simply taken as representations of case formulations if a client shares the specific contextual factors included in the studies (i.e., here, loss experiences or the COVID-19 pandemic). This is because data may still be sensitive to the unique situation in which they were collected, and statistical assumptions inherent with the estimated models do not map onto the kind of causal relationships that are commonly established in case formulations. In line with suggestions of von Klipstein and colleagues (2020), these findings could instead be used to start a conversation with a client in a more abstract manner, and to then focus on designing small thought or behavioral experiments to identify which of the relationships in networks are truly relevant for them, an approach I discussed in detail also in **chapter 9**.

13.1.3 Part III: Integration – Combining clinical prior information with statistical networks

13.1.3.1 Summary of chapters

The third part of the thesis introduced a new framework to systematically integrate case formulations with personalized networks. **Chapter 9** presented the Prior Elicitation Module for Idiographic System Estimation (PREMISE). PREMISE is a novel approach that formally integrates case formulations with personalized network estimation via prior elicitation and Bayesian inference. In doing so, it addresses some of the most pressing issues in implementing personalized network modeling in clinical practice: Estimation is more efficient because readily available clinical information are used, and clinicians can include theories and person-specific information in the model. The chapter showcased the clinical utility of PREMISE using the case formulation and ESM data from a client

diagnosed with OCD. **Chapter 10** presented a PREMISE investigation of two clients diagnosed with Anorexia Nervosa, specifically focusing on treatment implications. The main finding was that PREMISE networks for both clients had different implications for centrality-based treatment targets depending on the type of model (PREMISE versus traditional versus case formulation network). In particular, for one of the clients, the PREMISE network could be matched to CBT treatment modules to reduce excessive exercising and exposure approaches for fear of weight gain, whereas cognitive symptoms were more prominent in the data-driven network, calling for cognitive modules within the CBT-E protocol. **Chapter 11** introduced an alternative approach to combining longitudinal assessments with constructing networks, the Longitudinal Perceived Causal Relations (L-PCR) approach. The chapter illustrated how L-PCR networks can circumvent several of the potentially unfeasible assumptions in statistical estimation of personalized networks, such as the issue of restrictive time scales. Using data from 20 participants who completed between 20 to 28 daily assessments of depressive symptom relations, this chapter showed that L-PCR is generally feasible, well accepted, and may lead to clinically relevant insights on the structure and stability of perceived causal networks.

13.1.3.2 Answering research question and integration with literature

The chapters in this part of the thesis aimed to shed light on the second research question: "*How can personalized networks systematically be combined and integrated with the case formulation approach?*". I highlighted two avenues in which this question can be answered, either by using an initial case formulation as informative prior in the statistical estimation of time series networks, or by constructing personalized networks from longitudinal perceived causal relation (L-PCR) assessments. The two approaches have their respective benefits and limitations.

An advantage of the PREMISE approach over the L-PCR networks is that they do not only rely on the ability of the client to actively recall their symptom dynamics, but that they are also updated based on statistical inference. **Chapter 10** highlighted that ('prior') networks that are based on case formulations may indeed be very different from the statistical ('posterior') networks estimated from ESM data. As discussed in **chapter 9**, this may be for different reasons, and it is not possible to determine if differences between the networks indicate learning something about the actual case formulation, or if they are introduced due to errors resulting from bias in statistical estimation. Irrespective of the interpretation of these discrepancies, they can inspire concrete thought and behavior experiments, as discussed in the previous section, which is not possible in the L-PCR approach.

Statistical models vary in the number and strength of assumptions. These assumptions are helpful to arrive at conclusions that simplify complex structures in data, but they can give a biased picture of patterns if they are notoriously hard to meet. This is the case for the VAR model, because it seems unreasonable that relevant dynamics are only linear and unfold at only one specific time scale (in the case of the commonly used lag-1 model) (Olthof et al., 2020; Schiepek et al., 2017). An advantage of L-PCR networks over the PREMISE networks is that they remain agnostic regarding several of the problematic assumptions of statistical network estimation. They do not assume that dynamics only unfold on one specific time scale, and they do not per se limit relationships to linear interpretations (although it seems plausible that this will be how most clients and clinicians interpret them; Bos, 2021).

A currently unanswered question pertains to the clinical utility of both, the PREMISE and LPCR networks. In an upcoming study, we assess the perceived utility of clinicians and clients for both types of networks (the study uses PECAN networks instead of L-PCR networks, which follows the same logic of assessing perceived causal relations, but assessed in a single session as opposed to the longitudinal approach, see Klintwall et al., 2023). Preliminary data of this study suggest that the perceived clinical utility for both models is high. Specifically, clinicians indicate both types of networks as helpful in their communication with clients, which aligns with the perspective on using these tools as means of communication, rather than actual representations of case formulations, see also the discussion section of **chapter 9**. Recently, Wagner et al. (2023) have found similar responses in a qualitative study with clinicians, indicating that case formulation and psychoeducation are seen as useful applications of networks by clinicians.

13.1.4 Part IV: Formalization – Computational models of case formulations

13.1.4.1 Summary of chapter

The fourth part of this thesis discussed the potential of formalizing case formulations, and using a simulation-based approach to evaluate case formulations and clinical treatments for a given client. **Chapter 12** illustrated the benefits of this simulation-based approach using an example of a client diagnosed with panic disorder. The resulting computational model showed how case formulations can be evaluated by means of simulation results. In this specific example, the simulations aligned with general dynamics of panic disorder, but specific observations, such as rapid onset and decline of panic attacks were not depicted. The chapter discussed how these inconsistencies can give rise to refining the case formulation, overall contributing to stronger theories.

13.1.4.2 Answering research question and integration with literature

This chapter addressed the research question "*How can case formulations be advanced as idiographic theories and computational models using simulations?*". In this chapter, I described how formulating case formulations can be seen as an act of theory construction. Seeing case formulations as *theories* brings the benefit of drawing on a wealth of tools that have been developed to make theories more accurate and precise. One branch of theory development that has gained attention in the recent literature is formalization (Borsboom, van der Maas, et al., 2021; Guest & Martin, 2021; Haslbeck, Ryan, et al., 2021; Robinaugh et al., 2019, 2021; van Rooij & Baggio, 2021). In a world where case formulations can be formalized, we would have a much stronger understanding of the psychopathological mechanisms, and could precisely predict which clinical intervention would lead to the most desirable effects for a given individual. When starting the work on the corresponding paper to **chapter 12** in 2017 at the Institute of Advanced Study (IAS) in Amsterdam, I was a lot more enthusiastic about the feasibility of this approach, and I have since mostly focused on advancing a framework of formalizing case formulations that makes it more feasible for the future. Below, I will not only discuss *how*, but also *if* case formulations can be advanced using computational models.

The main obstacle in formalizing case formulations is the complexity of its process. First, formalization is a time consuming endeavor. For the relatively simple system presented in **chapter 12**, experts from different disciplines (psychiatry, clinical psychology, statistical methods, computational

science) were involved, and it took about two years as well as many iterations to produce meaningful simulation results. Obviously, doing this routinely for each client would not add clinical utility, irrespective of how strong the benefits of formalization are. Second, the process of constructing case formulations is rather diffuse, and there are many different ideas of how this can or should be done. For example, historically, case formulations have often been employed within the context of diagnostic interviewing, and the case formulation of a client was pre-specified depending on the specific diagnosis they receive. In Jacqueline Persons' approach (Persons, 2012), case formulations are seen as more flexible. While a nomothetic theory usually is the starting point, there are also personalized factors and relationships built in based on an extensive assessment of problems and resources. This push for personalization has found even stronger proponents in the recent literature. For example, in process-based therapy (Hayes & Hofmann, 2018; Hofmann & Hayes, 2019) and synergetic/idiographic system modeling (Schiepek, Stoger-Schmidinger, et al., 2016), case formulations are built from the ground up, without (explicit) connection to nomothetic principles. Differences in the specific approach to case formulation impact their formalization, because slight changes in the way differential equations are set up, let alone new variables that are introduced and conceptualized, can change simulation outcomes drastically, and the feasibility of formalization becomes unfeasible.

In 2022, during my research visit with Dr. Donald Robinaugh in Boston[42], and after our visit with Prof. Dr. Günter Schiepek in Salzburg[43], we developed a framework of formalizing case formulations that aims to address these problems. The main idea of this framework is that case formulations should *explain as much as possible* with *as little as possible*, as is also the case with theories in general. Therefore, the goal is not to construct a theory of the actual 'true system', one that includes all processes at play, from processes on the micro-level such as biological mechanisms all the way up to macro-level processes such as larger societal influences. Instead, we propose to identify what we refer to as a 'practical case formulation', one that aims for sparsity in processes based on the specific interest and research question of clinician and client. Formalizing practical case formulations aims to simplify the theory construction process by using core modules that can be formalized a priori, and only introducing personalized processes where relevant. We have devised three dimensions along which complexity can be reduced.

First, clinician and client need to identify the relevant *time scales*. In many cases, explanations can exist with two time scales. For example, avoidant coping can be explained by focusing on a short time scale (e.g., experiencing relief that feared situation can be avoided), and a long time scale (e.g., problems piling up due to avoidant behavior). Other time scales can be at play (e.g., trauma), but they do not necessarily impact the treatment rationale or are relevant to explain the avoidant behavior.

Second, clinician and client need to identify the relevant level of *abstraction*. Abstraction refers to the granularity of the included processes. It is possible to look at each component of a case formulation with a magnifying glass, which shows the finer processes at play. For example, a client might

42 Dr. Donald J. Robinaugh is a pioneer in the formalization of panic disorder (Robinaugh et al., 2019), one of the currently most developed formal theories in the field of clinical psychology.

43 Prof. Dr. Günter Schiepek is an expert in the field of synergetics and non-linear dynamics of case formulations (Schiepek et al., 2017; Schiepek, Stoger-Schmidinger, et al., 2016)

present the commonly co-occurring symptoms of an anxiety disorder and depression. Depending on the goals of clinician and client, a focus can be defined within the case formulation, and the relevant processes can be modeled at a finer versus coarser level.

Third, clinician and client need to identify the relevant level of *personalization*. Although a fully idiographic approach brings the advantage of case formulations entirely tailored toward the individual, it may also be inefficient to formalize them from scratch if there are patterns that can be observed across clients. Clinical theories provide explanations for these general patterns, but lack the flexibility of personalization. The goal of the personalization dimension is to identify the extent to which a deviation from the clinical theories is necessary to account for the personal experiences of the individual. From a formalization perspective, it is desirable if case formulations can rely as much as possible on established clinical theories, as these can be formalized and validated a priori by clinical researchers, lifting a big portion of the computational burden inherent to formalization.

Leaving the formalization work to clinical researchers as opposed to practitioners is one part of making formalization more feasible. A separate point is the extent to which we are really able to formalize psychopathological processes. For some disorders and processes that are well defined, such as panic disorder, and processes relating to the presence and persistence of panic attacks, there is hope that formalization could indeed be a future avenue (Robinaugh et al., 2019). However, these disorder categories are also examples where protocols already show promising results (e.g., exposure in panic disorders; Barlow, 2021), and the gains from personalization may not outweigh the resources that go into the construction of formal case formulations. More complex problems and disorders for which existing protocols have much less consistent and sustainable response rates, such as eating disorders (Walsh et al., 2021), would potentially benefit more from formalized case formulations. On the other hand, formalization entails specifying processes on a level that exceeds our current understanding of these less well-defined phenotypes. The extent to which formalization can really add clinical utility to the current practice of case formulations therefore remains unclear at this point, and largely depends on the above considerations.

Finally, there are philosophical considerations for constructing case formulations that impact the feasibility of formalization. This thesis generally takes a *realist* perspective, meaning that we assume there is some "real" or "true" case formulation that we aim to better understand through means of ESM assessment, network modeling, and formalization. In the most objective way, we aim to measure the client's experiences, and understand their individual dynamics, or we try to understand the reason they experience psychological symptoms through simulations. As realists, we hope that the results of these approaches are *not* dependent on the clinician or on other context-specific aspects of estimation. In contrast, there are researchers who put forward a *social-constructivist* perspective on case formulation in clinical practice. The social constructivist perspective entails that a case formulation is the result of the collaborative efforts of one specific clinician-client dyad in a specific setting (i.e., time, place, etc.). Indeed, the idea of collaboration between clinician and client is a core element of several clinical reflections on case formulation (Kuyken, Persons, Riese, Schiepek), and which is thought to result in benefits for therapeutic alliance, and adherence to treatment (Persons). The social constructivist perspective implies that changing any of these variables will result in a different case formulation. For example, clinician A may construct a different formulation with the same client compared to clinician B, or the same clinician-client dyad would

arrive at different formulations in different sessions (even assuming that no intervention has taken place yet). Crucially, the fact that the social-constructivist perspective assumes that there is indeed no one "true" case formulation implies that different formulations can be equally useful for clinical practice. The distinction between the realist and social constructivist perspective has important implications for the feasibility of formalization: If we assume that case formulations are indeed fully dependent on the specific context (social constructivism), formalization quickly becomes unfeasible because the formalization work needs to be completed by the clinician alone. In contrast, if we assume that at least some of the processes of interest are "real", and can be measured and modeled objectively (realism), at least parts of the formalization work can be completed by researchers in the ways I discussed above. This friction between feasibility requirements on the one hand and clinical practice and experience on the other hand needs to be acknowledged and addressed in future implementation efforts.

13.2 Future research

In this section, I will outline an agenda for future research that addresses unanswered questions of the thesis, with a specific focus on clinical implementation. For the clinical implementation of PREMISE I see two main challenges: First, demonstrating its clinical utility by investigating if outcomes such as client communication, improving treatment outcomes, creating insights for clinician and client can be improved above and beyond current standards. Second, identifying ways in which the estimation of PREMISE networks can seamlessly incorporated in clinical practice. In the next two paragraphs, I will discuss ongoing and future projects that aim to answer these questions.

13.2.1 Evaluating the perceived and clinical utility of PREMISE and L-PCR

Before new tools are added to the clinician's toolbelt, it is important to demonstrate that these can indeed advance the current practice, and that these advances outweigh potential costs connected to implementation of the new tool. Utility includes the *perceived utility* (i.e., do clinicians and clients see utility in using the tool to overcome challenges in therapy?), and the *clinical utility* (i.e., can treatment outcomes be improved over currently available approaches?). Finally, beyond demonstrating improvements, it is crucial to understand where they may come from, that is, what are the core ingredients that drive potential benefits?

13.2.1.1 Perceived utility

We are currently collecting data on clinicians' and clients' perceived utility of different types of networks, amongst others networks resulting from the PREMISE approach (**chapter 9** and **10**), traditional statistical estimation techniques (**chapter 3**), and networks resulting from perceived causal relations assessments (**chapter 11**). We are assessing potential gains regarding outcomes such as problem insight, ease of communication, expectation management, and motivation for change. Our preliminary data suggest that perceived utility of PREMISE networks, L-PCR networks, and traditionally estimated networks is likely high. Specifically, clinicians see value in using these tools for treatment communication and providing insight into the relationships of complaints. Due to the small sample size of this preliminary dataset, we are currently not able to speak to the significance of

differences between the different types of networks, but overall, the mean ratings of the perceived utility outcomes are moderate to high, ranging from 3.36 to 4.00 on a 5-point Likert scale. When being asked to indicate which networks clinicians generally favor out of the three, most clinicians indicated that the PREMISE networks best represent the mental health problems of their clients.[44]

13.2.1.2 Clinical utility

In addition to testing the perceived utility, investigating if treatment outcomes can be improved by using ESM-derived networks is needed. The current gold standard for testing this is via randomized controlled trials (RCTs), although other designs may also be useful, such as repeated single-subject designs (Kravitz & Duan, 2014). There are recent examples of trials that demonstrate the clinical utility of personalized networks estimated in the traditional sense (**chapter 3**), for example in eating disorders (Levinson et al., 2023), however, no formal investigations of PREMISE- or L-PCR-derived treatments exist to date. As shown in **chapter 10**, treatments derived from PREMISE networks differ substantially from the ones tested in the cited trial. Therefore, adding a PREMISE treatment arm to such trials would be relevant to speak to the clinical utility of PREMISE.

13.2.1.3 Core ingredients of treatment improvement

Trials can only investigate if outcomes do or do not improve but they usually cannot explain the underlying mechanisms of such improvements. I expect that a core ingredient that drives the improvements of treatment outcomes in the implementation of personalized networks may be effects driven from a stronger involvement of the client. Persons (2012) discussed these benefits of personalization, i.e., a deep involvement of the client, leading to benefits such as reduced nonadherence and improved therapeutic relationship. These benefits may also – to some extent – drive improved outcomes in treatments derived from personalized networks, but may not per se result from the specific treatment recommendation being superior to other approaches. Our current preliminary data on the perceived utility hints at a potential involvement of these general benefits of personalization in the improvement of treatment outcomes. Investigating the underlying mechanisms that drive treatment gains will be an additional question for future research, should benefits be observed in PREMISE-derived treatments.

13.2.2 Challenges for clinical implementation of PREMISE

If and when the clinical utility of PREMISE can be demonstrated, the next important step toward implementation is a concrete proposal for an integration into clinical practice. Recent and ongoing work has provided guidance in setting up ESM studies in practice (Riese et al., 2021), and there are tools that allow for a straightforward ESM data collection with clients (Bos et al., 2022; Mestdagh

---

44 Note that data collection is currently ongoing and the exact (numeric) findings may still change. For the most up to date results, please consider the manuscript that is currently unpublished at the time of writing this discussion: Scholten, Burger, Papp, Winter, Glombiewski, & Klintwall (2023). How do therapists respond to the agony of choice between competing case formulations? Balancing patients' perceived causalities with intensive longitudinal data in constructing personalized networks. *Manuscript in preparation*.

& Dejonckheere, 2021). In the context of PREMISE specifically, there are two additional questions that are relevant for integration with therapy which I will address in this section.

13.2.2.1 Updating networks and treatments

PREMISE networks are established on the basis of an initial case formulation, and are subsequently updated with incoming ESM data. Updates to the network may result in different treatment implications, so it is important to guide clinicians in determining the right time for updating networks and treatment courses. At this point, it is unclear how best to approach this problem. One possible avenue is to combine case adaptive designs with the updating of PREMISE networks (Blackwell et al., 2019), such that a network is updated with ESM data when no further improvements can be observed from the current treatment. However, it is important to note that interventions may, and ideally will, alter the structure of the network, which stands in direct contrast to the stationarity assumption of the VAR model. Updating networks that are based on priors established *before* treatment will therefore not be an appropriate choice, and may indeed be harmful for estimation (i.e., more data may be needed to override outdated network connections compared to not using the prior). An alternative is time-varying VAR networks, which need further implementation work for the PREMISE approach (Haslbeck, Bringmann, et al., 2020). As discussed in chapter 9, choosing a statistical model for PREMISE implementations will be a crucial question for future research. A connected issue is that there is a wide range of data pre-processing steps that can be taken, different approaches to variable inclusion, and different analysis options across and even within one modeling framework that result in different treatment recommendations (Bastiaansen et al., 2020). Future research should investigate the effects these specific choices have on the outcomes, and derive concrete recommendations for given research contexts.

13.2.2.2 Deriving treatments from personalized networks

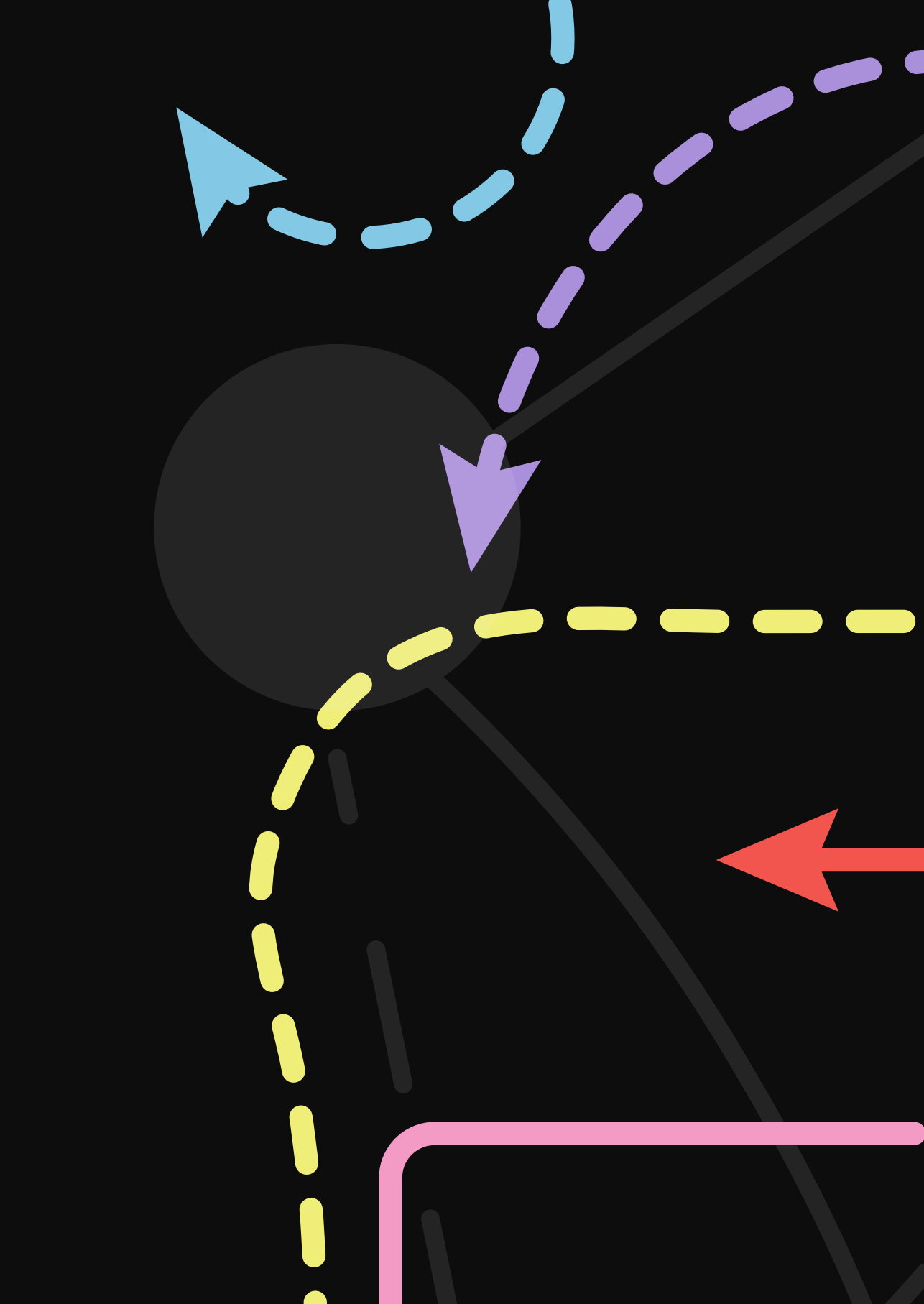
Chapters 9 and **10** have focused on constructing networks from clinical prior information and ESM data. A main interest for deriving personalized treatments is how these network structures can be mapped onto clinical interventions (Rubel et al., 2018). In the literature, centrality indices are often used to conceptualize "influential nodes" as potentially relevant treatment targets. Although this approach is not without problems (Dablander & Hinne, 2019; Rodebaugh et al., 2018; Wichers et al., 2021), there is evidence that it can indeed help identifying promising targets (Levinson et al., 2023). There are new sophisticated simulation-based approaches (Lunansky et al., 2022) that identify targets based on their overall influence on other nodes. Ultimately, any algorithm that identifies single nodes also works under the assumption that a single node can represent a treatment target. The feasibility of this assumption heavily depends on many factors: The selection of nodes influences the network structure, and specific nodes that are central in a given set of nodes may be in the periphery when including other nodes. Further, nodes differ in their general amenability to existing treatments. It may, for example, be easier to target concrete behaviors compared to complex emotions, irrespective of the centrality of these nodes. Finally, as discussed above, case formulations can be formulated on many different levels of abstraction, and not all of these levels are mapping onto the specificity of clinical treatments. This directly relates to the problem that single nodes are often not compatible with interventions that usually target a

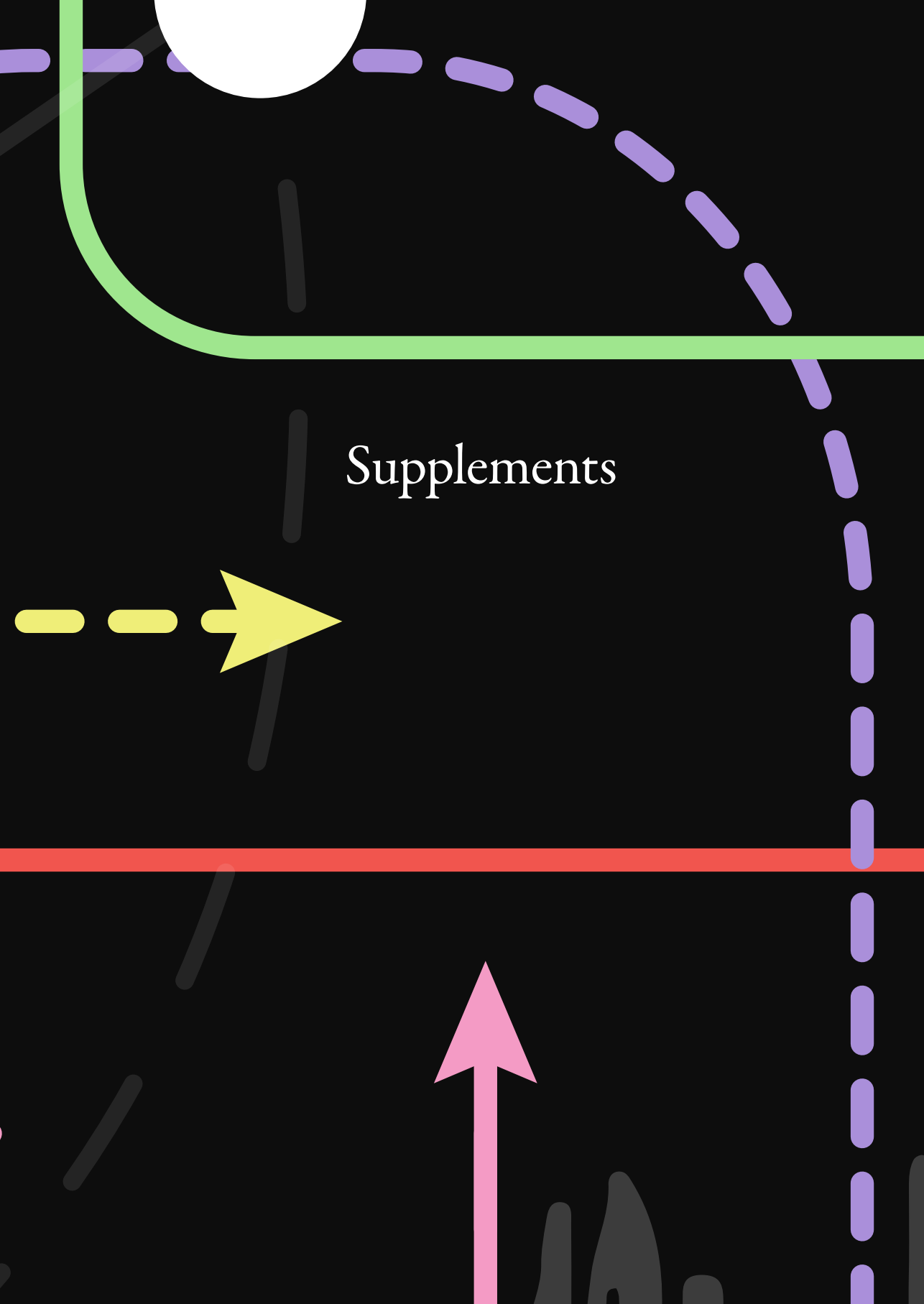
combination of nodes. Algorithms to identify relevant targets in networks should thus focus on specific clusters of nodes that are connected in existing clinical interventions. To summarize, the selection of intervention targets remains a challenging topic in the field of (personalized) network modeling, and future research in algorithms should draw on clinical considerations (amenability, specificity of interventions, etc.) to ensure clinical utility.

13.3 Conclusion

In this thesis, I have outlined current and new statistical and mathematical approaches to advance case formulation in clinical psychology. I have demonstrated how an integration with case formulation approaches is essential for clinical implementation. The thesis also demonstrated how nomothetic principles and theories can be used to inform idiographic models, and introduced considerations for balancing personalization with evidence from group-level studies. In this final chapter, I have outlined future avenues for advancing network analysis and formalization in idiographic research and clinical practice.

Based on the current evidence, I see promise in PREMISE and L-PCR-based networks to aid client communication and psychoeducation. The validity, as well as the clinical utility of these approaches is yet to be established in trials. Regarding the feasibility and clinical utility of formalized case formulations I remain sceptic, at least in the current landscape of possibilities. In this concluding chapter, I have outlined steps toward making the formalization of case formulations more feasible. The success of this approach will largely depend on if these steps can be realized. Should this indeed be possible at some point in the future, we can look forward to a new era of case formulations.




Supplements

SUPPLEMENT TO CHAPTER 5

Example 1: Relationships in later life

Data for the first example stem from the Swiss longitudinal study "Relationships in later life", which followed widowed and separated individuals after their loss experience and collected information on their psychosocial functioning, including depressive symptomatology. The data and project description can be found online (https://www.kpp.psy.unibe.ch/forschung/projekte/nccrlives/index_ger.html), and the results have been discussed in chapter 8. The main research interest here lies in comparing depressive symptom networks between the widowed and separated individuals, specifically comparing how strongly they are connected and the overall structure of the two networks. Next to the *general analysis routine*, we therefore focus on *group comparison* (methods and results), *network visualization* (results), and *network density* (results).

Methods – General Analysis Routine

Sample collection, Variable selection procedure	For this analysis, we included data collected on the German version of the Center for Epidemiologic Studies Depression scale (CES-D; Radloff, 1977; German: Allgemeine Depressions-Skala, ADS-K; Hautzinger & Geue, 2016). The dataset consists of 1,276 married, 566 widowed, and 971 separated individuals. Participants were contacted via post mail and filled in a pen-and-paper questionnaire. To circumvent the issue that participants might be at different stages of adaptation to the adverse life event, we only included participants with a maximum distance of two years to the event (widowhood/separation). This resulted in 145 widowed and 217 separated individuals. To be able to include widowhood/separation as a node in the network, we added 145 married controls to the widowed sample, and 217 married controls to the separated sample[45]. This way, widowhood/separation is included as a binary node, indicating the presence versus absence of the respective life event. In order to investigate conceptual overlap between variables, we examined bivariate correlations between all variables, and combined items if their content suggested strong conceptual similarity, and their bivariate correlation was $r \geq .50$. Accordingly, we combined the original items *mood*, *upset*, and *depressed* (new item "mood"), as well as the items *happy* and *enjoy* (new item "happy"). This resulted in 12 variables, each rated on an ordinal scale with four answer categories [1 = "rarely or none of the time (less than 1 day)", 2 = "some or a little of the time (1–2 days)", 3 = "occasionally or a moderate amount of time (3–4 days)", 4 = "most or all of the time (5–7 days)"].
Estimation method	We estimated partial correlation networks for both, the widowed and separated sample, using the *glasso* regularization and a tuning-parameter gamma set to 0.5 (Foygel & Drton, 2010). Due to the ordinal, non-normal nature of the data, we used Spearman's rank-correlation and pairwise complete observations to handle missing data. In total, of all variables included in the network analysis, 6.6% of the ratings were missing in the widowed/married sample and 5.1% in the separated/married sample. Here, we assume that these ratings are missing at random (Rubin, 1976).

45 Note that adding control participants and including group membership in the network is only but one way to approach group comparisons. Many other techniques have been discussed recently, such as moderated network analysis (Haslbeck, Borsboom, et al., 2019), or Bayesian approaches (Williams et al., 2020). For more detailed information on the approach used here, we advise to consider the original publication.

Accuracy and stability of edge-estimates	To assess accuracy of the edge estimates, we conducted the routine implemented in the *bootnet* package (Epskamp, Borsboom, et al., 2018), using nonparametric bootstrapping with 1,000 bootstrap samples.
Statistical packages	The analyses have been conducted using *R*-version 3.5.2 on October 8th, 2020. For network estimation, we used the *estimateNetwork* function in the *bootnet* package (Epskamp, Borsboom, et al., 2018). Networks have been visualized using the *qgraph* package (Epskamp et al., 2012).

Methods – Analysis-specific Routine

Group comparisons	Groups were compared by obtaining the difference in global strength within the *Network Comparison Test* (van Borkulo, Boschloo, et al., 2017), using 2,000 iterations, and with seed set to '123'. This test assesses if the two networks differ in their overall level of connectivity. Since we are primarily interested in global differences in network connectivity, other tests available within the *NCT* were disregarded in the present analyses. Additionally, we correlated the weighted adjacency matrices of the two networks as an additional measure of similarity between the networks.

Results – General Analysis Routine

Final sample size	The widowed network included 290 individuals (145 widowed and 145 married controls), and the separated network included 434 individuals (217 separated and 217 married controls).
Results of accuracy and stability checks	Results of the nonparametric bootstrap analysis can be found in Figure S5.1. In general, the confidence intervals were rather broad and overlapping. The order of edge estimates should therefore be interpreted with caution.

Results – Analysis-specific Routine

Network visualization	The networks of widowed and separated individuals are visualized in Figure S5.2. Here, edges represent regularized partial correlations between symptoms. Edge weights in the widowed network ranged from 0.002 (*sad – getgo*) to 0.300 (*lonely – widowed*). Edge weights in the separated network ranged from 0.001 (*mood – unfriendly*) to 0.320 (*lonely – separation*). To facilitate interpretability, we used the colorblind-theme in *qgraph* (Epskamp et al., 2012), fixed the average layout between the two network plots using the *averageLayout* function, curved edges that would otherwise cross nodes, and made negative edges dashed.[46] No specific minimum/maximum/cut values have been used for network visualization.[47]

46 This is useful if printed without colors.

47 Note: Any exploratory reporting of findings, such as relevant edges, will be specific to the given research context. The figures presented below are based on an adapted version of the publicly available code from chapter 8, see also Burger et al., (2020).

Network density and average absolute edge weights	Since we are interested in comparing the two networks with regard to their connectivity, we computed the density of the two networks by determining the ratio of detected edges to the total number of edges in a fully connected network. The network of widowed/married individuals had a density of .615 (48/78 edges), with a mean weight of 0.044, and the separated network had a density of .744 (58/78 edges), with a mean weight of 0.053.
Group comparisons	While the global invariance test within the *Network Comparison Test* procedure indicated that there were some differences in the overall level of connectivity between the widowed and separated network (p = .003), the weighted adjacency matrices showed a rather large correlation (r = .750), indicating that the overall structure between the networks was similar. This shows that the networks differed in how strongly connected they are (sum of absolute edge weights, connectivity), while edges that were detected showed a similar pattern across the two networks (correlation of edges), i.e., edges that were large (small) in the separated network were generally also large (small) in the widowed network.

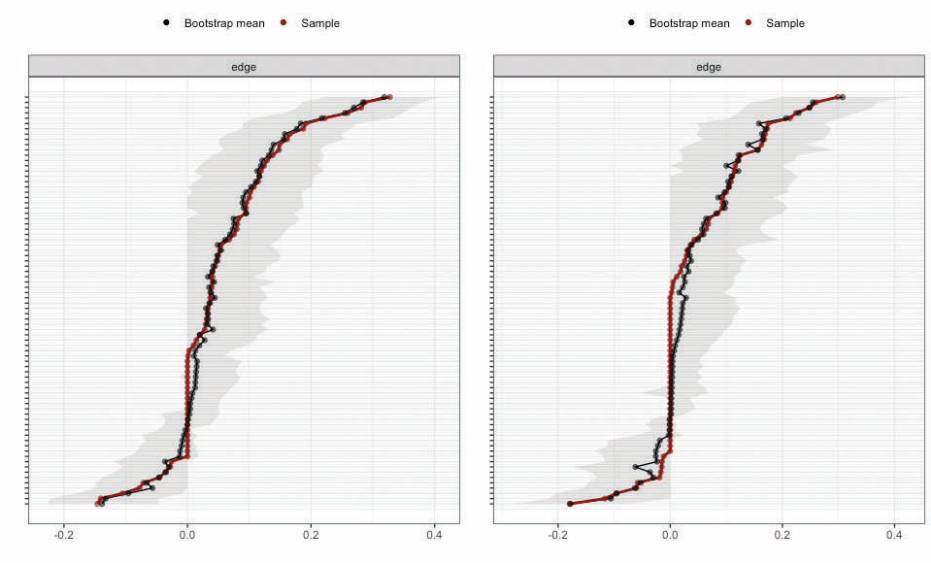


Figure S5.1. Nonparametric bootstrapping results with 1,000 samples for the separated (left) and the widowed network (right).

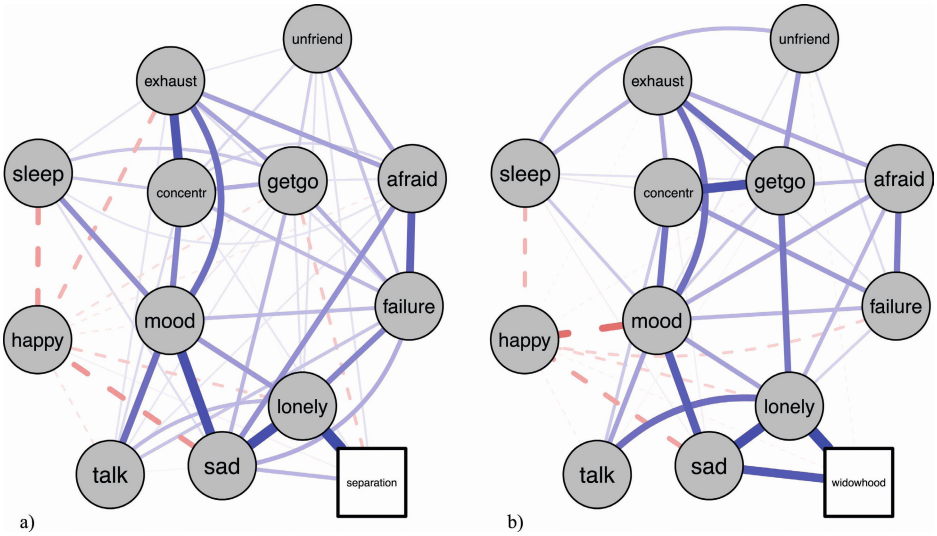


Figure S5.2. Regularized partial-correlation networks (tuning-parameter gamma = 0.5) for the separated (left) and the widowed sample (right). Solid-blue edges represent positive, regularized partial-correlations, dashed-red edges represent negative, regularized partial-correlations. No specific minimum/maximum/cut values have been used. Edge weights in the separated network ranged from 0.001 (*mood – unfriendly*) to 0.320 (*lonely – separation*). Edge weights in the widowed network ranged from 0.002 (*sad – getgo*) to 0.300 (*lonely – widowed*).

Table S.1. Overview of routines covered in the two examples.

	Example 1	Example 2
Data description	**Relationships in later life** (data and results from *Bereavement or breakup: Differences in networks of depression*; Burger et al., 2020).	**Taylor Manifest Anxiety Scale** (data and results from an online offering of the *Taylor Manifest Anxiety Scale*; Taylor, 1953).
Methods: General analysis routine		
Sample collection	✓	✓
Variable selection procedure	✓	✓
Deterministic relations between variables and skip-structures	✓ (not applicable)	✓ (not applicable)
Estimation method	✓	✓
Accuracy and stability of edge-estimates	✓	✓
Statistical packages	✓	✓
Methods: Analysis-specific routine		
Group comparison	✓	
Centrality indices		✓
Differences between edges		✓
Clustering		
Results: General analysis routine		
Final sample size	✓	✓
Results of the accuracy and stability checks	✓	✓
Results: Analysis-specific routine		
Network visualization	✓	✓
Network density and average absolute edge weights	✓	
Centrality indices		✓
Predictability		
Specific nodes and edges		✓
Group comparisons	✓	

Example 2: Taylor Manifest Anxiety Scale

Data for the second example data stem from the *openpsychometrics.org* project, using the *Taylor Manifest Anxiety Scale* (Taylor, 1953). The data and project description can be found online (https://openpsychometrics.org/tests/TMAS). Let us assume the main research interests here lie in the general network structure of anxiety, edge differences in the network structure, as well as in which items play a more central role in the network. Next to the *general analysis routine*, we therefore focus on *centrality results* (methods and results), *edge differences* (methods and results), *network visualization* (results), and *local network properties* (results).

Methods – General Analysis Routine

Sample collection, Variable selection procedure	For this analysis, we included data collected on the *Taylor Manifest Anxiety Scale* (Taylor, 1953). This data was collected online; at the end of the test users were asked if their answers were accurate and could be used for research. 76% said yes and data have been published on the *openpsychometrics.org* project. The dataset consisted of 5410 individuals. The network model included all questions from the *Taylor Manifest Anxiety Scale* (Taylor, 1953), thus resulting in 50 nodes. Each item was rated on a binary scale with two answer categories [0 = FALSE, 1 = TRUE]. In addition, missing data was encoded as NA and we used listwise deletion for missing data points, as the chosen estimation algorithm does not allow for missing data. Data were assumed to be missing completely at random.
Estimation method	We estimated the network structure using an Ising model (Van Borkulo et al., 2014). An Ising model represents associations between dichotomous variables using pairwise log linear relationships, similar to partial correlation coefficients in a Gaussian Graphical Model (GGM; Epskamp, Waldorp, et al., 2018). To control for potential spurious associations, the estimation procedure here uses a penalized nodewise regression approach, specifically the eLasso penalty based on the Extended Bayesian Information Criterion (Ravikumar et al., 2010). Default values as set in the package were used, with the EBIC hypertuning parameter set to 0.25.
Accuracy and stability of edge-estimates	To assess the accuracy of the edge weight estimates, we conducted the routine implemented in the *bootnet* package (Epskamp, Borsboom, et al., 2018), using nonparametric bootstrapping based on 1,000 bootstrap samples.
Statistical packages	The analyses have been conducted using *R*-version 3.5.2 on October 12th, 2020. For network estimation, we used the *estimateNetwork* function in the *bootnet* package (Epskamp, Borsboom, et al., 2018), using the *IsingFit* package (van Borkulo et al., 2014). The accuracy of estimates has been assessed using the *bootnet* function. Networks have been visualized using the *qgraph* package (Epskamp et al., 2012).

Methods – Analysis-specific Routine

Centrality Indices	To further quantify how well a node is directly connected to other nodes in the network structure, we investigated *strength* as a centrality measure (Costantini et al., 2015; Opsahl et al., 2010). To assess accuracy of the *strength* centrality estimates, we conducted the routine implemented in the *bootnet* package (Epskamp, Borsboom, et al., 2018), using case-drop bootstrapping based on 1,000 bootstrap samples. Further, to ensure interpretable differences in centrality, we used the bootstrapped difference-test in the *bootnet package*.
Differences between edges within one network	Finally, as we were interested in an exploratory fashion whether certain edges were stronger and stood out in the network structure, we carried out a bootstrapped difference-test using the *R* package *bootnet* (Epskamp, Borsboom, et al., 2018).

Results – General Analysis Routine

Final sample size	Following removal of missing data, 4,474 subjects were included in the current analyses.
Results of accuracy and stability checks	In general, the confidence intervals were very narrow, indicating stable results. In addition, *strength* centrality estimates were stable, with a centrality stability coefficient of 0.75, indicating that 75% of the data could be dropped to retain with 95% certainty a correlation of 0.7 with the original dataset. Of note, while the most central items were more central than most other items in the network, they were not more central than each other (see Figure S5.7).

Results – Analysis-specific Routine

Network visualization	The network visualization is presented in Figure S5.3. To facilitate interpretability, here we used the colorblind-theme in *qgraph* (Epskamp et al., 2012), included a legend with the description of each item, and used a *cut* value of 0. Edge weights ranged from –1.82 (Q47–Q50) to 2.28 (Q6–Q41). The layout used was the automatically generated layout based on the Fruchterman-Reingold algorithm (Fruchterman & Reingold, 1991a). Any exploratory reporting of findings, such as relevant edges, will be specific to the given research context.
Centrality indices	Figure S5.4 presents the results of the centrality analyses. In addition, Supplementary Table S5.2 presents the standardized and raw centrality indices. The three most central items were: Q27, Q31, and Q48. Of note, while these were more central than many other items in the network, differences between the items themselves were not robust (see Figure S5.5).
Specific nodes and edges	Figure S5.8 presents the results of the edge difference test. The labels are omitted for clarity. In general, the bootstrapped difference test identified several edges as significantly different from most other edges in the network. Of note, the two strongest edges in the current network structure were significantly different from each other and all other edges in the network. These are the edge between *Q6* and *Q41* and between *Q40* and *Q46*.

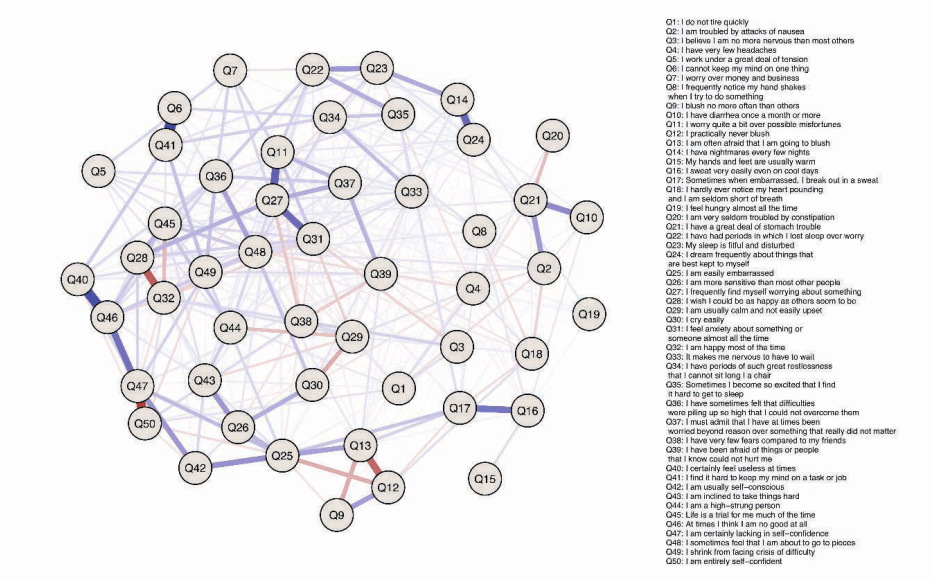


Figure S5.3. Example 2: Regularized log-linear relations. Blue edges represent positive relations, red edges represent negative relations. The cut argument has been set to 0. Edge weights ranged from –1.82 (Q47–Q50) to 2.28 (Q6–Q41).

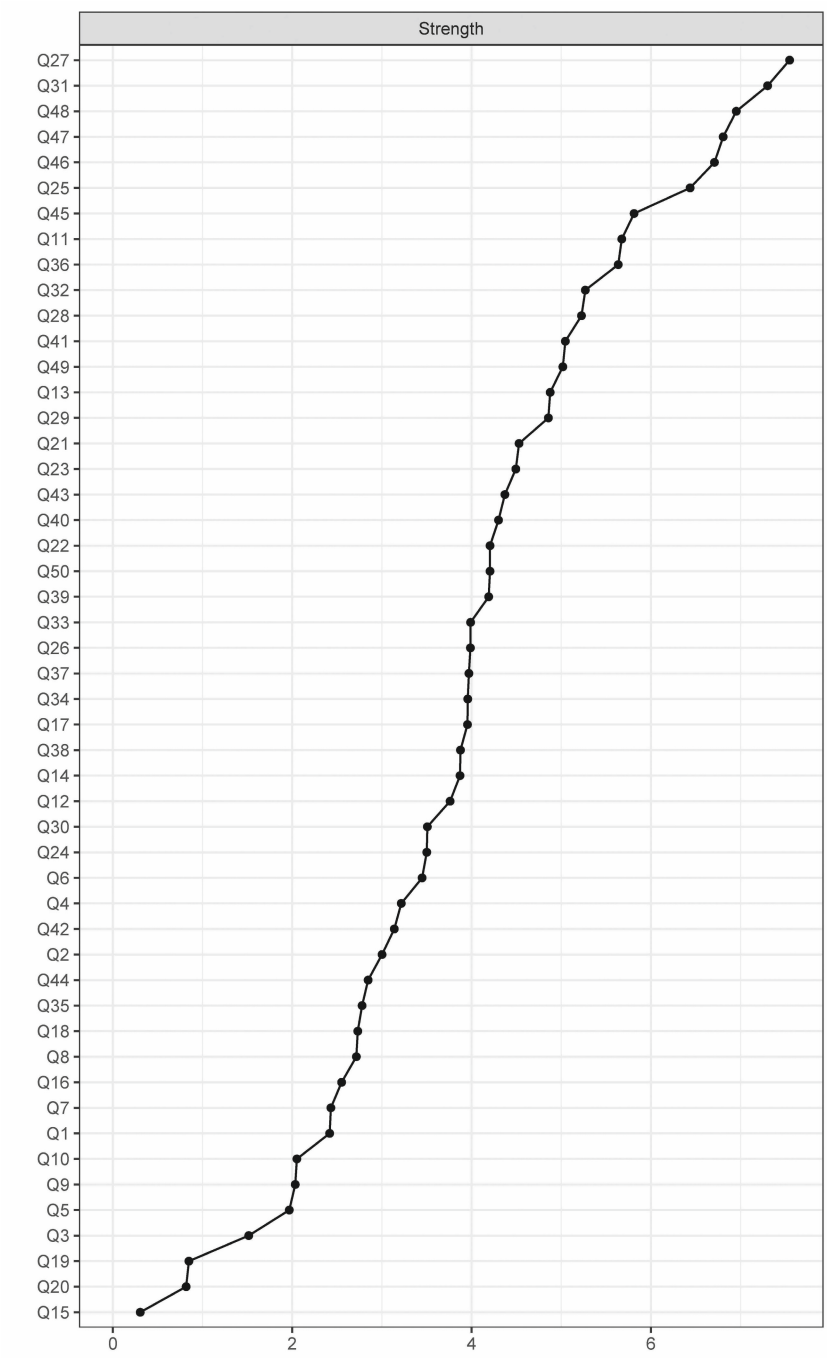


Figure S5.4. Centrality plot denoting *strength* centrality results. The order is set from the strongest to the weakest item. The x-axis represents raw centrality scores.

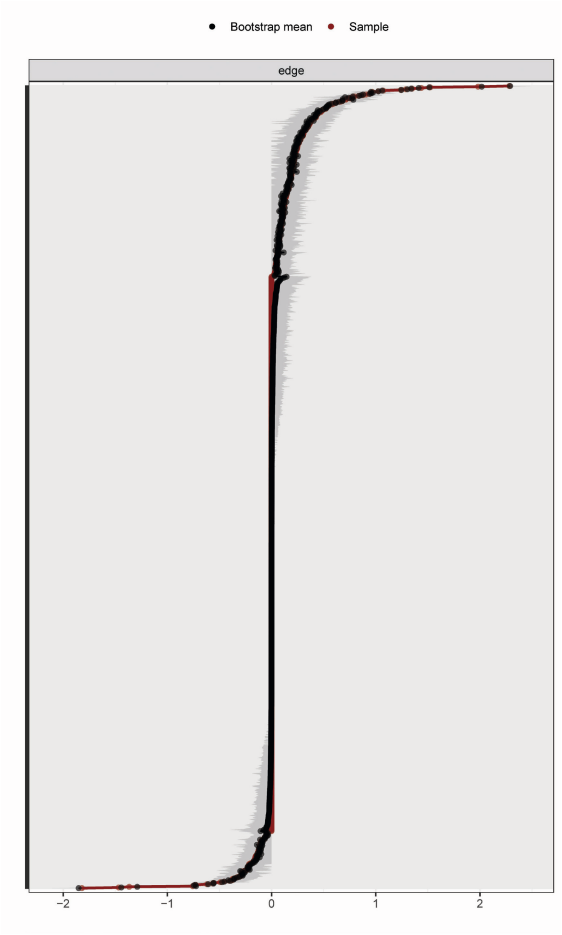


Figure S5.5. Accuracy of the edge-weights for the estimated network model. The horizonal area within the plot represents the 95% quantile range of the parameter values across 1,000 bootstraps. The red dots indicate the sample values, while the black dots indicate the bootstrap mean values.

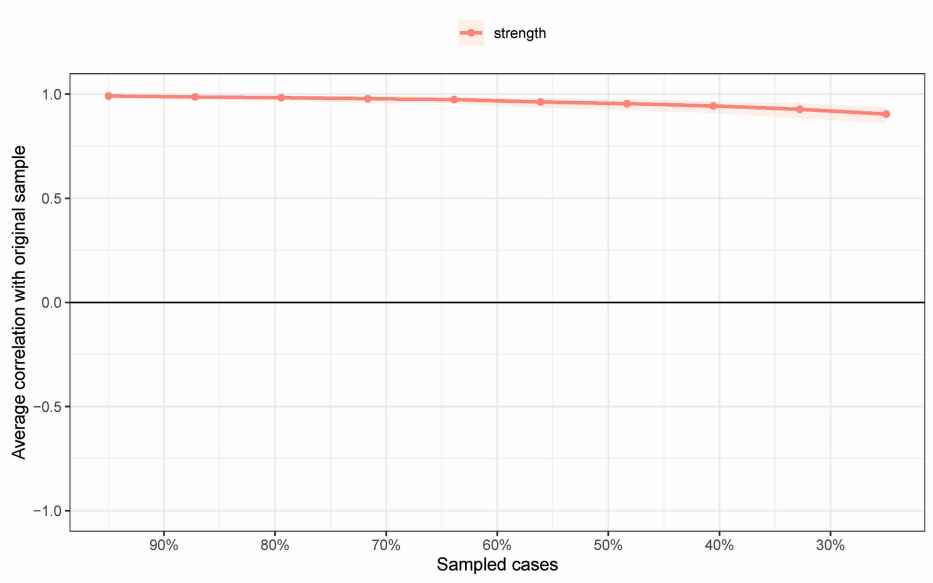


Figure S5.6. Stability of *strength* centrality estimates.

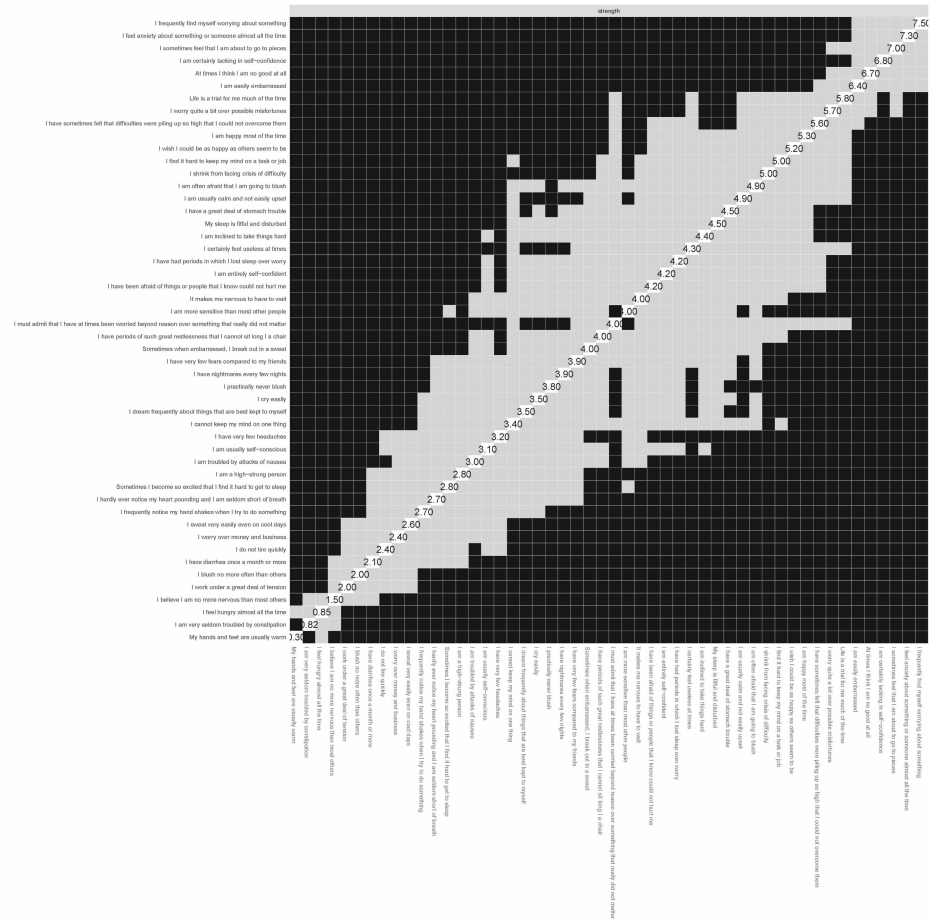


Figure S5.7. *Strength b*ootstrapped difference test (α=0.05). Grey boxes reflect no significant differences and black boxes reflect significant differences.

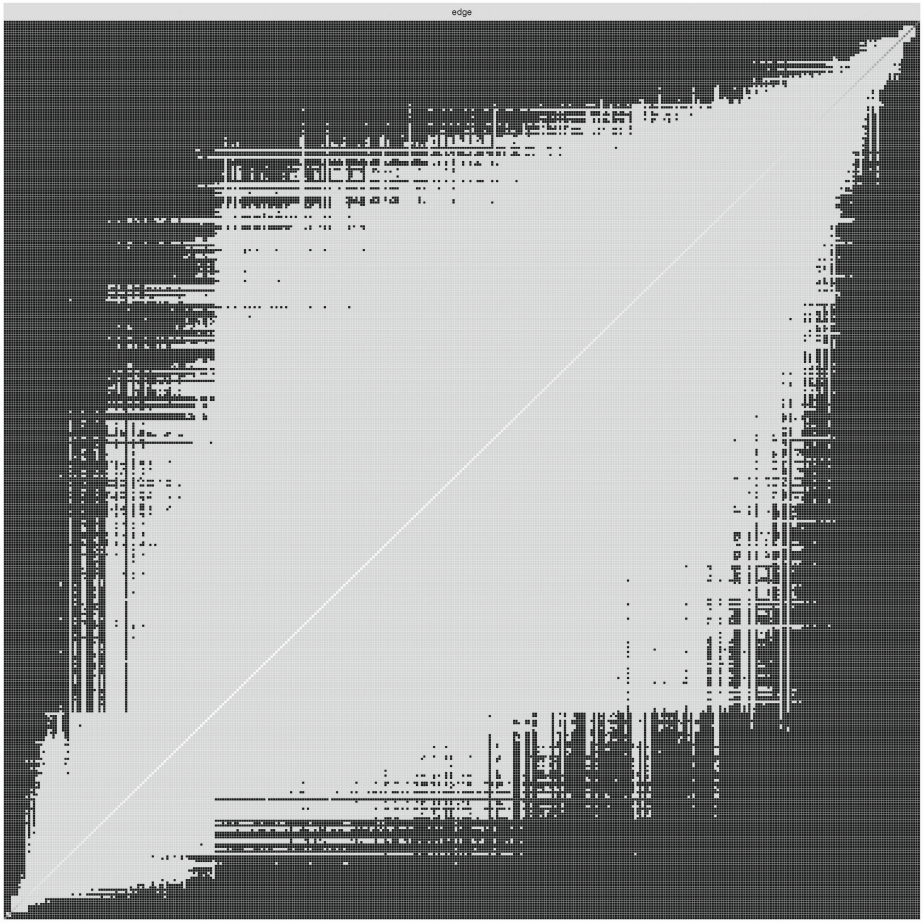


Figure S5.8. *Edge* bootstrapped difference test for all non-zero edges in the network structure (α=0.05). Grey boxes reflect no significant differences and black boxes reflect significant differences.

Table S5.2. Standardized and unstandardized *strength* centrality values

	Standardized *strength* values	Unstandardized *strength* values
Q1	-0.93	2.42
Q2	-0.58	3
Q3	-1.47	1.51
Q4	-0.45	3.22
Q5	-1.2	1.97
Q6	-0.31	3.45
Q7	-0.92	2.43
Q8	-0.75	2.71
Q9	-1.16	2.03
Q10	-1.15	2.05
Q11	1.02	5.67
Q12	-0.13	3.76
Q13	0.54	4.88
Q14	-0.06	3.87
Q15	-2.19	0.3
Q16	-0.85	2.55
Q17	-0.01	3.95
Q18	-0.74	2.73
Q19	-1.86	0.85
Q20	-1.88	0.82
Q21	0.33	4.53
Q22	0.14	4.21
Q23	0.31	4.49
Q24	-0.28	3.5
Q25	1.47	6.44
Q26	0.01	3.99
Q27	2.13	7.55
Q28	0.75	5.22
Q29	0.53	4.86
Q30	-0.28	3.51
Q31	1.98	7.3
Q32	0.77	5.27
Q33	0.01	3.99
Q34	-0.01	3.96
Q35	-0.71	2.78
Q36	0.99	5.63

Table S5.2. Continued

	Standardized *strength* values	Unstandardized *strength* values
Q37	0	3.97
Q38	-0.06	3.88
Q39	0.13	4.19
Q40	0.2	4.3
Q41	0.64	5.04
Q42	-0.5	3.14
Q43	0.24	4.37
Q44	-0.67	2.85
Q45	1.1	5.81
Q46	1.63	6.71
Q47	1.69	6.8
Q48	1.78	6.95
Q49	0.62	5.02
Q50	0.14	4.2

B

SUPPLEMENT TO CHAPTER 6

Network replicability

The replicability of the edges of the networks was assessed through correlating the estimated edges weights in random split half subsamples and in the first and the second half of the time series. The Pearson correlations between edge weights in the two random subsamples were .933 for the temporal network and .999 for the contemporaneous network. For the first and second half of the time series, the corresponding correlations were .906 and .999. To examine the stability of centrality values, estimated centrality indices were compared through correlations across each respective pair of subsamples. Between the random split-half samples, the correlations were .700 for outstrength and .879 for instrength in the temporal network and .995 for strength in the contemporaneous network. Between the split-half time series, the corresponding correlations were .717, .790. and .998. Finally, rate of consistency among the nodes with the highest centrality was assessed through comparing the total number of times the most central nodes identified by the main analyses were replicated across the four subsamples described above, used as a proxy to estimations approximating the rank-order stability of the centrality indices. The nodes revealing the highest centrality were consistent across subsamples, with 90% of the edges with the highest centrality re-obtained in the subsample analyses across all networks.

Table S6.1 The results of mixed models for the change patterns of the network variables

Variable	B	SE	Df	t
Anxiety				
Intercept	1.508	0.020	1851.0	75.56*
Weekend	-0.125	0.008	1384.9	-15.67*
Day	0.001	0.000	1575.4	2.83
Uncontrollable worry				
Intercept	1.448	0.019	1801.2	78.03**
Weekend	-0.062	0.007	1372.4	-9.32**
Day	0.000	0.000	1549.8	0.92
Generalized worry				
Intercept	1.753	0.022	1769.7	81.58*
Weekend	-0.122	0.008	1372.6	-15.81*
Day	0.002	0.001	1539.4	3.46*
Irritability				
Intercept	1.495	0.015	2494.2	99.00*
Weekend	-0.076	0.009	1399.6	-8.85*
Day	0.000	0.000	1809.2	0.46
Fear of being infected				
Intercept	1.208	0.015	2084.2	96.12*
Weekend	-0.025	0.005	48314.2	-5.04*
Day	0.004	0.001	1673.9	8.52*
Panic				
Intercept	1.174	0.014	1673.1	87.08*
Weekend	-0.025	0.004	1392.6	-5.74*
Day	0.000	0.000	1471.1	1.15
Fear of others being infected				
Intercept	1.238	0.014	1691.0	85.92*
Weekend	-0.019	0.005	47984.1	-3.98*
Day	0.006	0.001	1482.6	12.64*
Threat monitoring				
Intercept	1.626	0.021	1720.7	78.60*
Weekend	-0.049	0.007	1394.9	-7.17*
Day	0.003	0.001	1512.4	5.51*
Avoidance				
Intercept	1.368	0.017	1743.5	78.03*
Weekend	-0.051	0.006	1385.1	-8.02*
Day	0.001	0.000	1497.4	2.66

Table S6.1 The results of mixed models for the change patterns of the network variables

Variable	B	SE	Df	t
Thoughts lead to losing mind				
Intercept	1.220	0.015	1639.2	83.33*
Weekend	-0.021	0.004	45890.6	-4.75*
Day	0.001	0.001	1472.2	1.87
Intolerance of uncertainty				
Intercept	1.428	0.175	2101.0	81.75
Week	-0.173	0.009	1385.4	-19.99
Day	0.001	0.000	1646.5	3.40*
Physical activity				
Intercept	2.450	0.028	2338.7	87.72*
Weekend	0.146	0.019	1386.3	7.79*
Day	0.001	0.001	1650.1	2.08
Alcohol use				
Intercept	1.550	0.016	3103.6	96.54*
Weekend	0.255	0.011	1377.0	22.45*
Day	-0.000	0.000	14489.9	-0.13
Interpersonal conflict				
Intercept	1.250	0.010	3001.3	122.61*
Weekend	0.002	0.007	1418.1	0.30
Day	0.001	0.000	1994.5	1.68
Sleep satisfaction				
Intercept	3.153	0.025	1861.4	126.83*
Weekend	0.172	0.011	1391.6	15.11*
Day	0.002	0.001	1537.9	3.51*
Sufficient information				
Intercept	3.493	0.029	1487.0	121.77*
Weekend	-0.042	0.006	46309.7	-6.72*
Day	-0.001	0.001	1400.9	-2.22
Social media use				
Intercept	2.950	0.027	1638.4	110.33*
Weekend	0.073	0.010	1382.4	7.28*
Day	0.001	0.001	1467.8	1.69

Note. For most variables, an uncorrelated random intercept, weekend effect and slope and a one-lag autoregressive (AR(1)) covariance structure for the residuals turned out to have the best fit for the model without fixed predictors except the intercept. In a few cases, the random weekend effect or slope had to be removed for the model to converge. This is reflected in the *df*s.

*$p < .001$ (two-tailed).

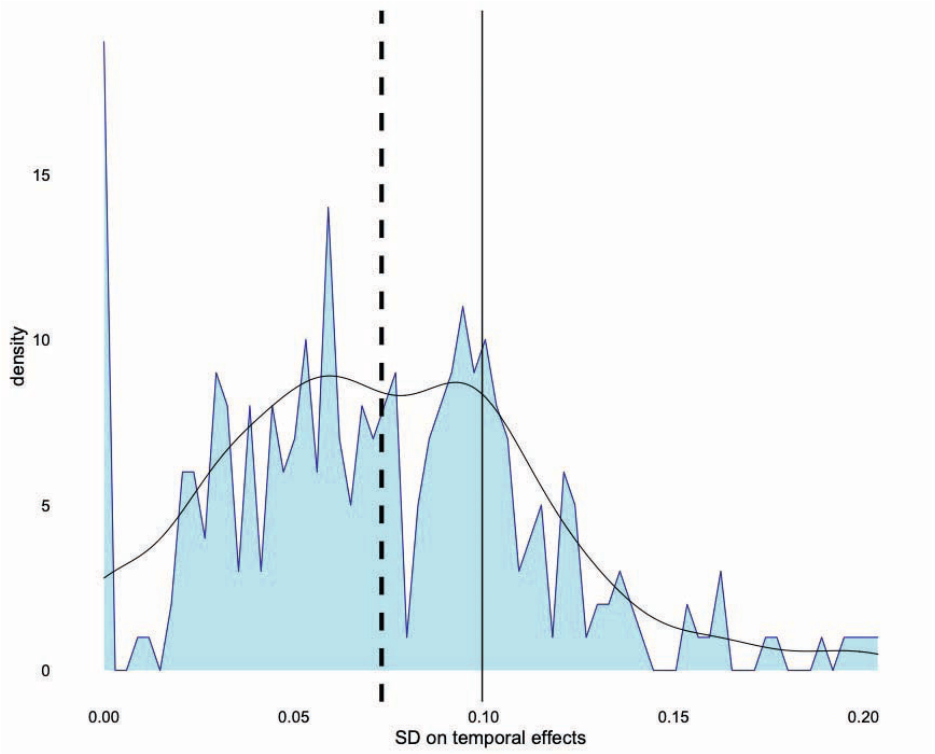


Figure S6.1. Distribution of standard deviations of random effects in the temporal network. *Note.* The mean of the standard deviations was 0.073 (dashed line). The solid line represents the cut-off for the edges with the 25% largest standard deviations in the network, resulting in a cut-off of SD = 0.100. The largest SD was observed for the auto-regressive effect of *Thoughts lead to losing mind* (SD = 0.204).

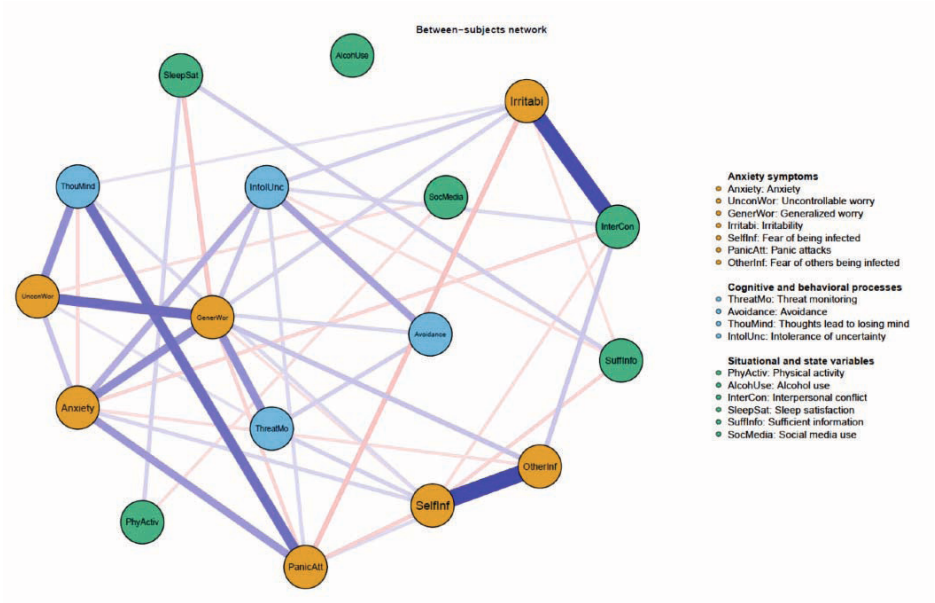


Figure S6.2. Between-person network. *Note.* The between-person network represents the correlations between the person-means on the variables, given the person-means on the other variables. Edges at a significance level of $p < .001$ are shown.

SUPPLEMENT TO CHAPTER 8

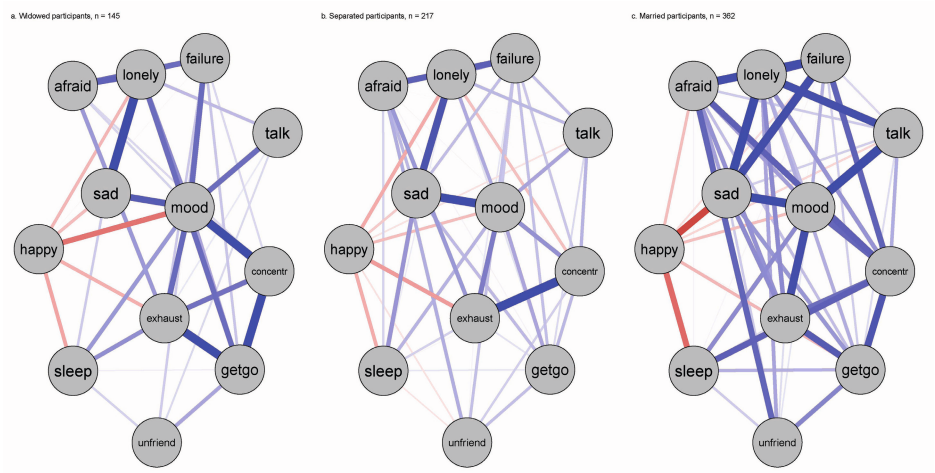


Figure S8.1. Regularized partial-correlation networks estimated according to estimation method b (separate networks for each of the three samples, as opposed to two networks for widowed and separated participants with married controls).

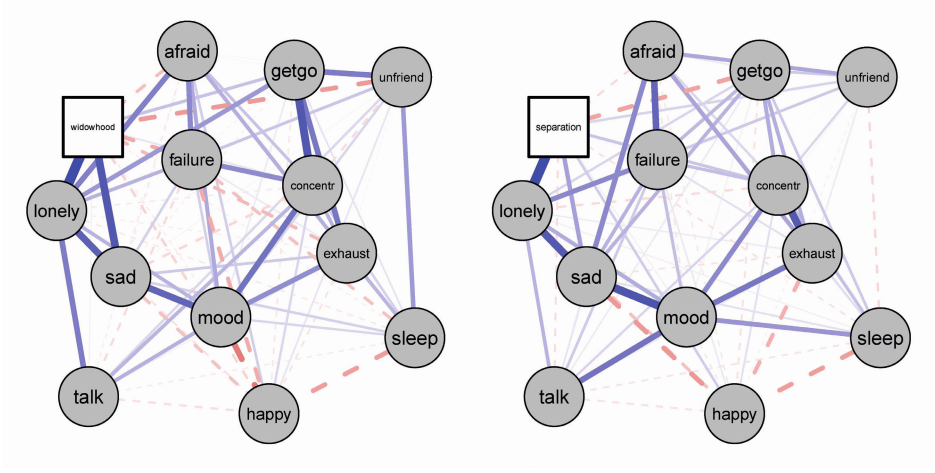


Figure S8.2. Non-regularized partial-correlation networks estimated according to estimation method a.

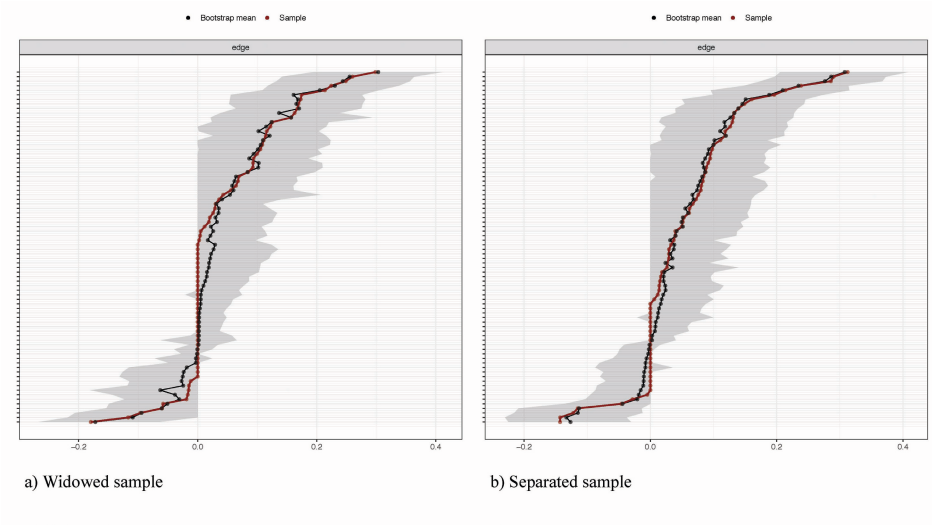


Figure S8.3. Edge weight estimates and confidence intervals for both samples in the bootstrap procedure.

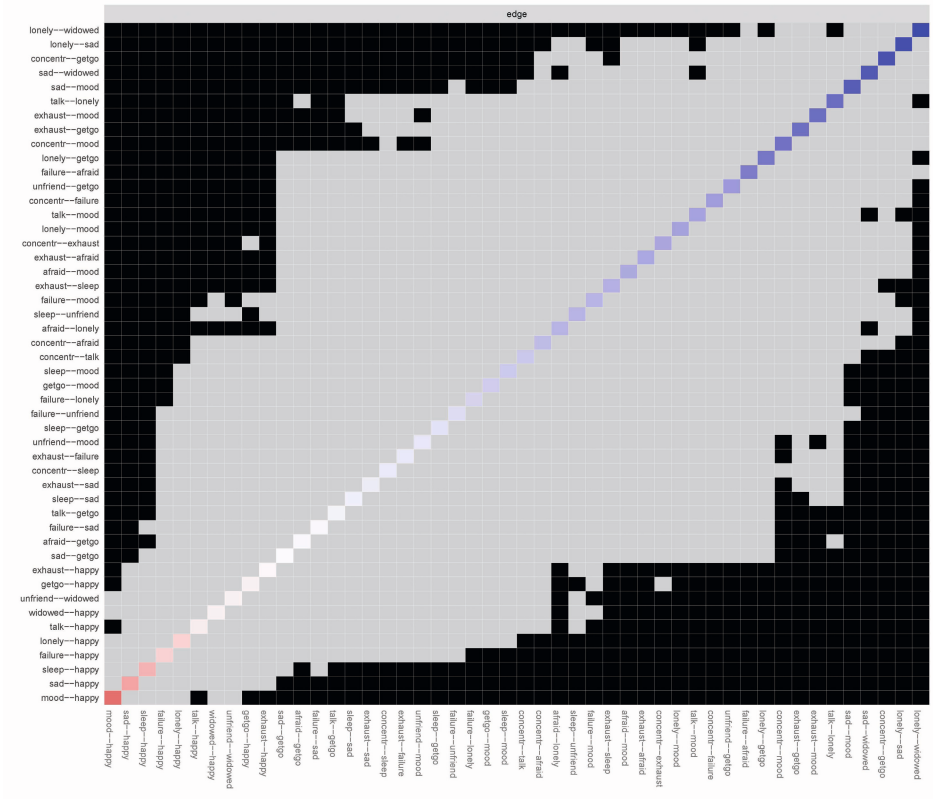


Figure S8.4. Edge difference test for the widowed sample in the bootstrap procedure. Only nonzero edges are included and edges are ordered by weight. A black box indicates a significant difference in edge weight, grey boxes indicate non-significant differences in edge weight.

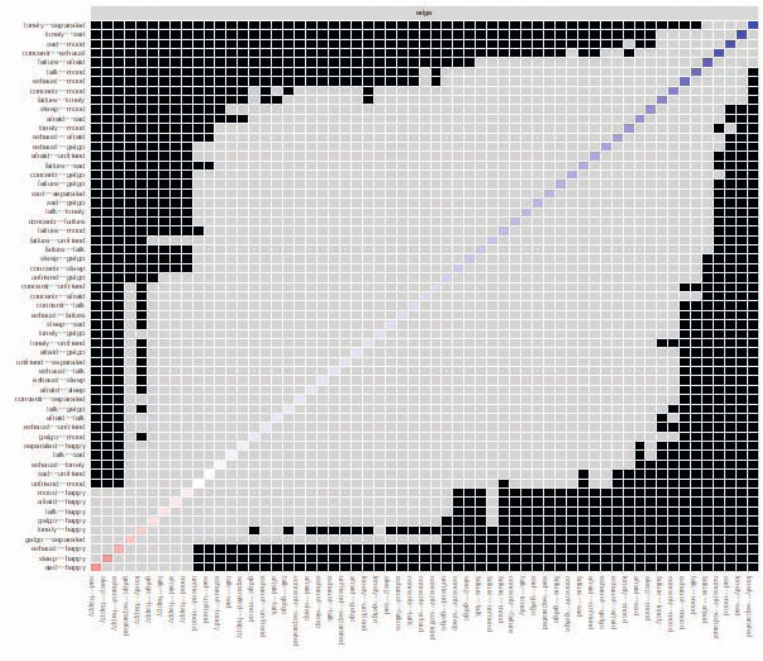


Figure S8.5. Edge difference test for the separated sample in the bootstrap procedure. Only non-zero edges are included and edges are ordered by weight. A black box indicates a significant difference in edge weight, grey boxes indicate non-significant differences in edge weight.

Table S8.1
Original items, labels and item combinations

Original CES-D item (translated from German ADS-K short scale). *During the past week…*	Labels of original items	Labels of combined items
1)…*I felt that I could not shake off the blues even with help from my family or friends.* […konnte ich meine trübsinnige Laune nicht loswerden, obwohl mich meine Freunde / Familie versuchten aufzumuntern.]	Mood	Mood (combined)
2)…*I was bothered by things that usually don't bother me.* […haben mich Dinge beunruhigt, die mir sonst nichts ausmachen.]	Upset	Mood (combined)
3)… *I felt depressed.* […war ich deprimiert / niedergeschlagen.]	Depressed	Mood (combined)
4)…*I was happy.* […war ich fröhlich gestimmt.]	Happy	Happy (combined)
5)…*I enjoyed life.* […habe ich das Leben genossen.]	Enjoy	Happy (combined)
6)…*I had trouble keeping my mind on what I was doing.* […hatte ich Mühe, mich zu konzentrieren.]	Concentr	Concentr
7)…*I felt that everything I did was an effort.* […war alles anstrengend für mich.]	Exhaust	Exhaust
8)…*I thought my life had been a failure.* […dachte ich, mein Leben ist ein einziger Fehlschlag.]	Failure	Failure
9)…*I was afraid.* […hatte ich Angst.]	Afraid	Afraid
10)…*my sleep was restless.* […habe ich schlecht geschlafen.]	Sleep	Sleep
11)…*I talked less than usual.* […habe ich weniger als sonst geredet.]	Talk	Talk
12)…*I felt lonely.* […fühlte ich mich einsam.]	Lonely	Lonely
13)…*I felt sad.* […war ich traurig.]	Sad	Sad
14)…*people were unfriendly.* […hatte ich das Gefühl, dass mich die Leute nicht leiden können.]	Unfriendly	Unfriendly
15)…*I could not get "going."* […konnte ich mich zu nichts aufraffen.]	Getgo	Getgo

D

SUPPLEMENT TO CHAPTER 10

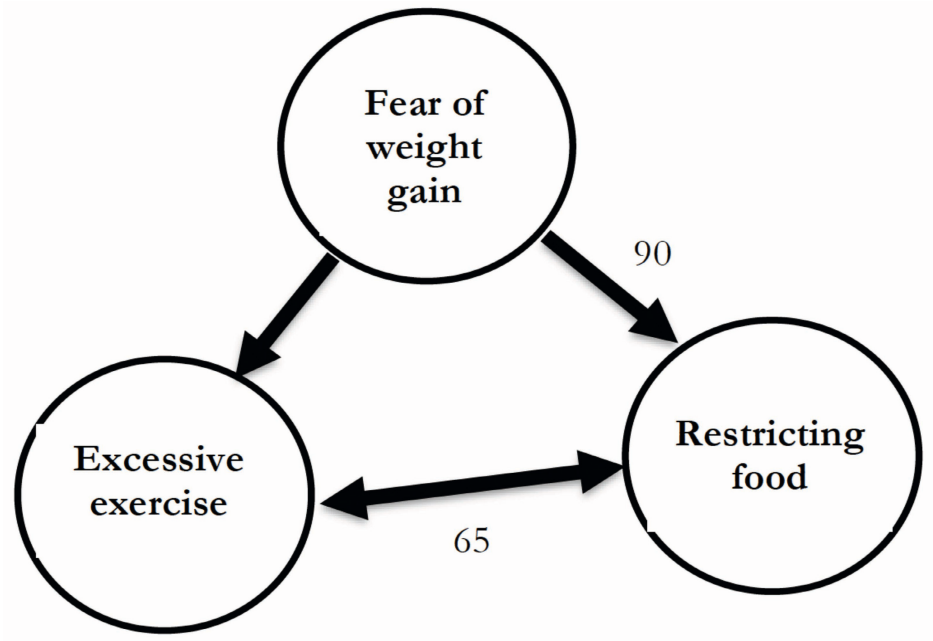


Figure S10.1. Example of case formulation rating.

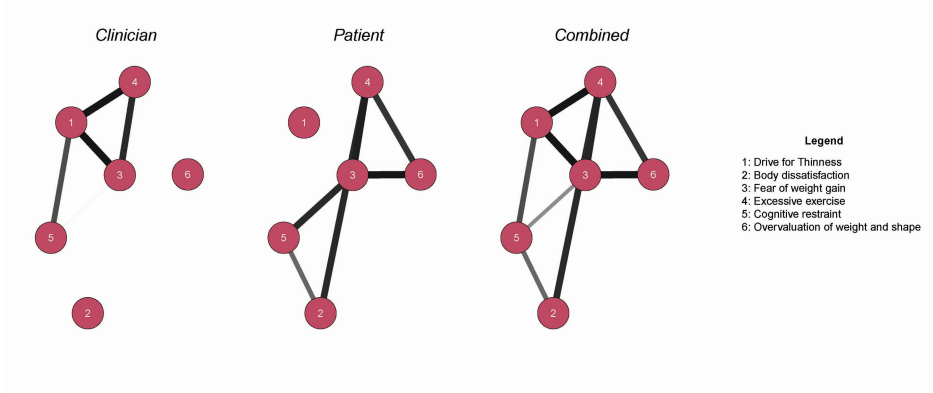


Figure S10.2. Clinician, client, and combined case formulation networks for client A.

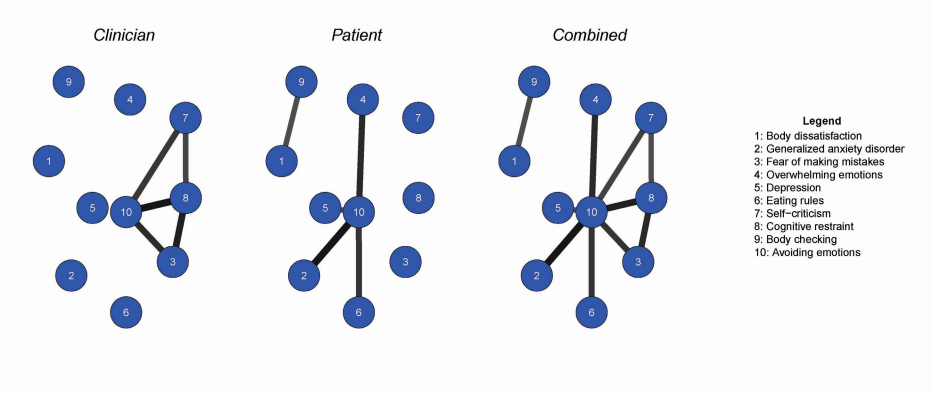


Figure S10.3. Clinician, client, and combined case formulation networks for client B.

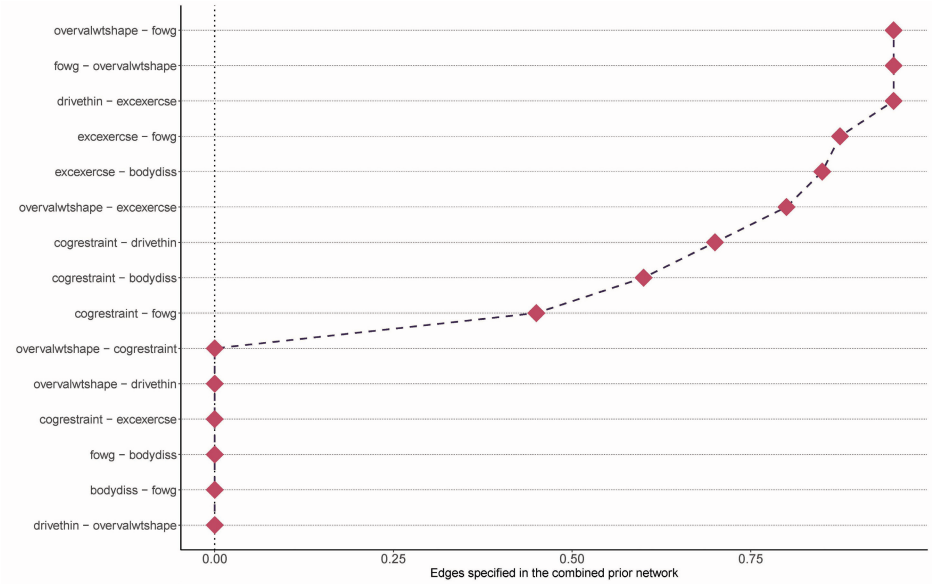


Figure S10.4. Edge specifications for the case formulation network of client A.

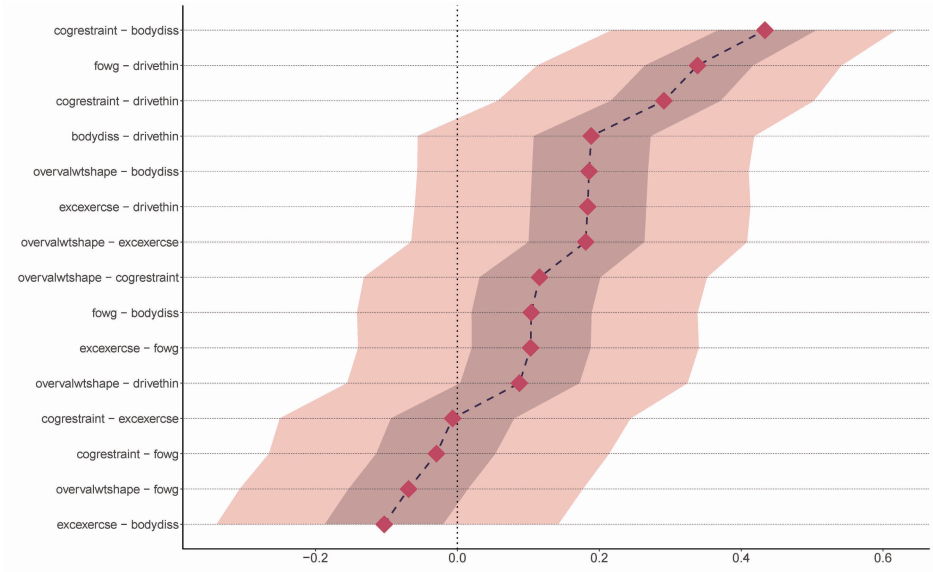


Figure S10.5. Edge estimates and respective 95% and 50% credibility intervals for the EMA network of client A.

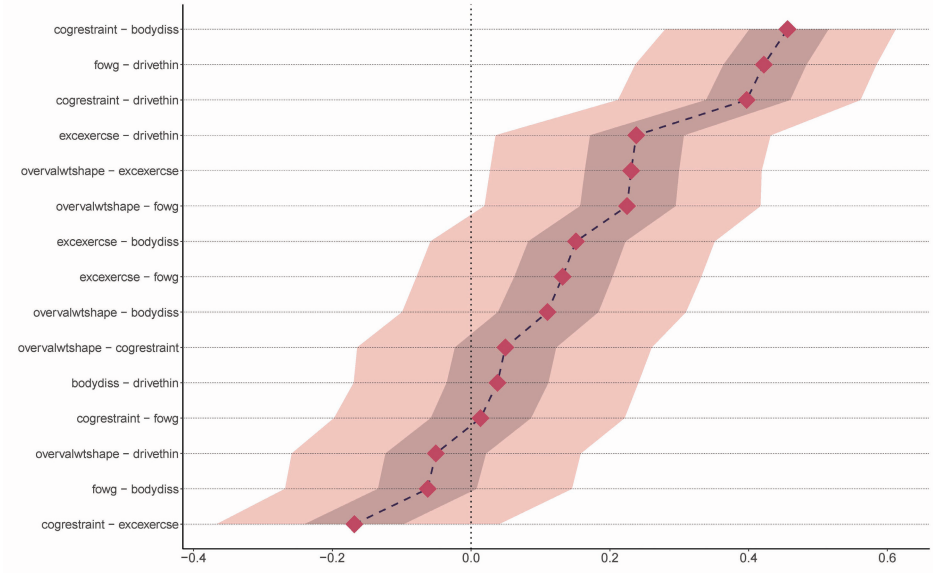


Figure S10.6. Edge estimates and respective 95% and 50% credibility intervals for the PREMISE network of client A.

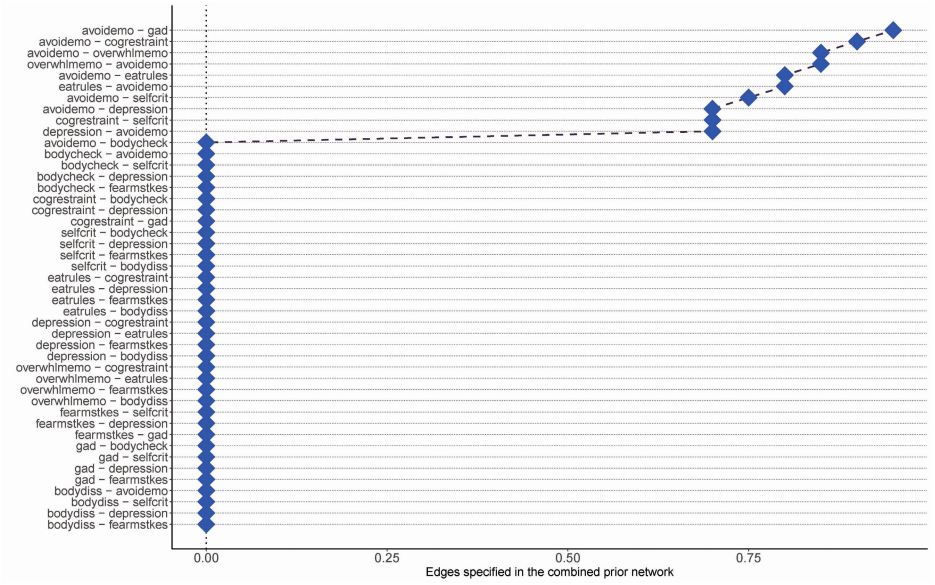


Figure S10.7. Edge specifications for the case formulation network of client B.

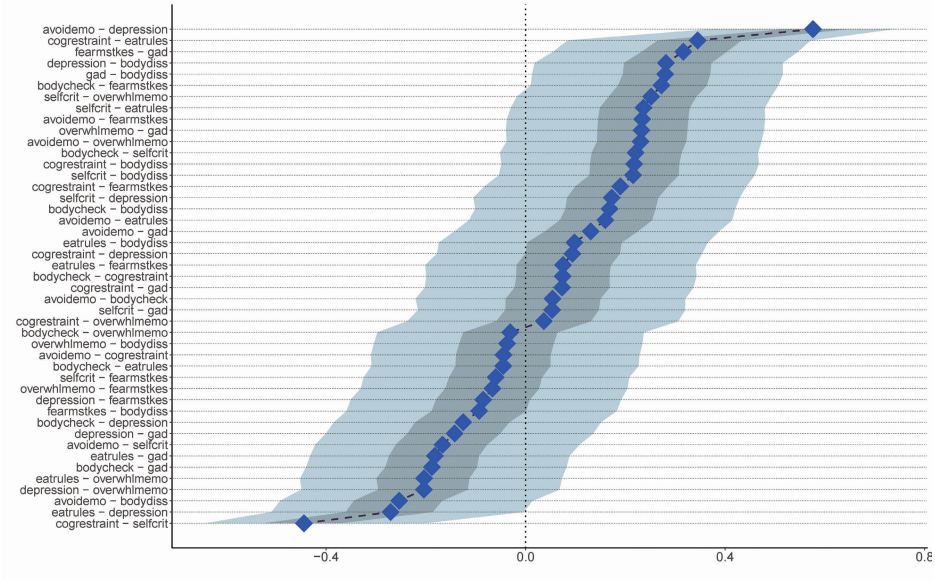


Figure S10.8. Edge estimates and respective 95% and 50% credibility intervals for the EMA network of client B.

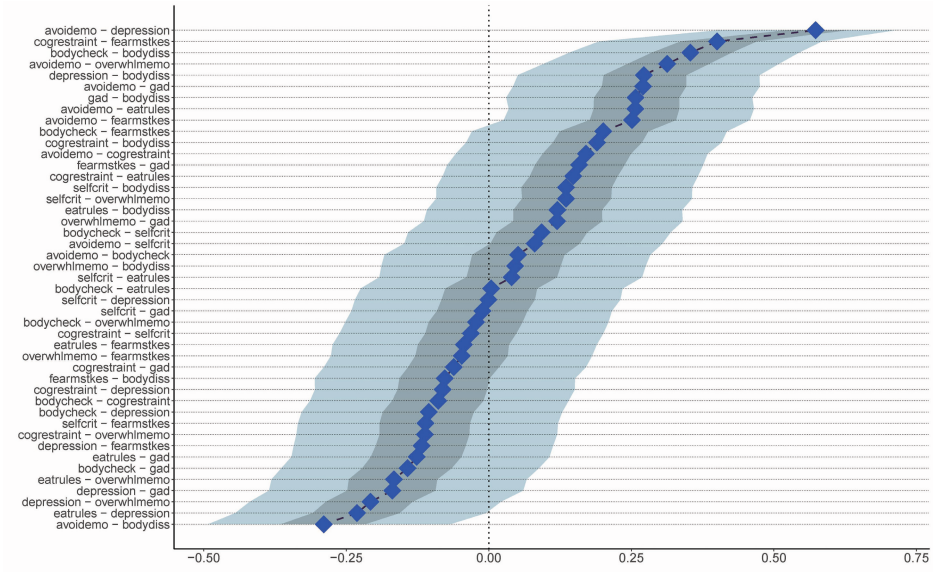


Figure S10.9. Edge estimates and respective 95% and 50% credibility intervals for the PREMISE network of client B.

E

SUPPLEMENT TO CHAPTER 12

This section is meant to give an overview over the mathematical aspects of the model used in the paper. Note that the paper can be followed without having read this section, it is meant for the reader interested in some more mathematical detail.

Modeling external input: discriminant stimulus

First, we modelled input of the discriminant stimulus as a basis of panic and coping dynamics. This variable is considered *external*, meaning that it is not caused by any other variables within the individual's system. The extent to which the individual is exposed to the discriminant stimulus was defined for each time-point by a chance of 12 (not exposed) to 2 (exposed), which results in roughly a 14% chance of exposure at every time-point.

System specification: differential equations

Second, we derived first-order differential equations for all simulation studies from the schematic depiction of the system as shown in Figure 3. Generally, we modelled the momentary rate of change as a function of input variables and a decay term of the variable. The basis for the interaction between panic and avoidant coping were non-linear Lotka-Volterra dynamics, which we extended with other variables related to coping. Since we considered the perceived benefits and the credibility of the catastrophic interpretation as variables that *evaluate* the current coping outcomes, rather than being directly influenced by current levels of panic/avoidance, we included the full term of panic and avoidant coping as components of the respective equations, resulting in the following system equations:

Catastrophizing.

$$\frac{dCat}{dt} = a \cdot Sd + b \cdot Cred - c \cdot Cat \qquad (1)$$

Panic.

$$\frac{dPan}{dt} = d \cdot Cat - e \cdot Pan \cdot Av - f \cdot Pan \qquad (2)$$

Avoidance behavior.

$$\frac{dAv}{dt} = g \cdot Pan \cdot Av - h \cdot Cost + i \cdot Ben - j \cdot Av \qquad (3)$$

Perceived benefits.

$$\frac{dBen}{dt} = k \cdot [(d \cdot Cat - e \cdot Pan \cdot Av - f \cdot Pan) \cdot Av] - l \cdot Ben \qquad (4)$$

Credibility of catastrophic interpretation.

$$\frac{!Cred}{dt} = m \cdot [(g \cdot Pan \cdot Av - h \cdot Cost + i \cdot Ben - j \cdot Av) \cdot Pan] - n \cdot Cred \quad (5)$$

Perceived costs.

$$\frac{dCost}{dt} = o \cdot Av - p \cdot Cost \quad (6)$$

Interventions

We formalized two commonly used CBT interventions: First, exposure was implemented through setting avoidant coping to 0. Second, cognitive reappraisal was modelled through introducing a new system variable, that captures the credibility of an alternative, functional interpretation of the discriminant stimulus. This results in the differential equation

$$\frac{dFunCog}{dt} = q \cdot Cred + r \cdot FunCog \quad (7)$$

According to our theory, the introduction of cognitive reappraisal has an impact on catastrophizing and the credibility of the catastrophic interpretation. Therefore, we extended the differential equations of these variables with the introduction of cognitive reappraisal:

$$\frac{dCat2}{dt} = a \cdot Sd + b \cdot Cred - c \cdot Cat - s \cdot FunCog \quad (8)$$

$$\frac{dCred2}{dt} = m \cdot [(g \cdot Pan \cdot Av - h \cdot Cost + i \cdot Ben - j \cdot Av) \cdot Pan] - n \cdot Cred - t \cdot FunCog \quad (9)$$

Parameter choices and initial values

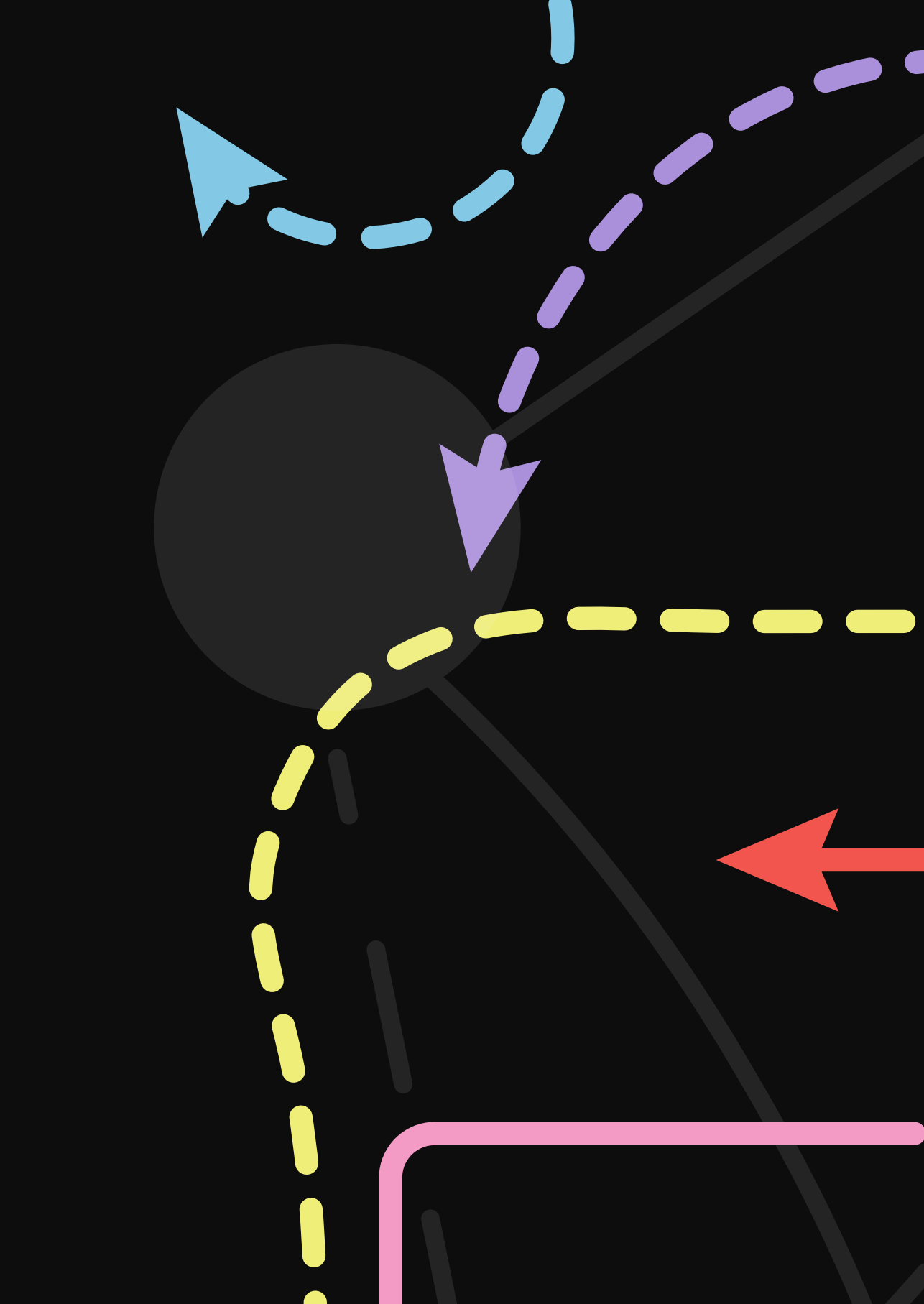
As discussed in the main text, estimating parameters in differential equations from ESM data requires that all system variables can be measured on their appropriate time-scale. Another application in computational modeling is that client and clinician can collaboratively examine the impact of varying parameters on the client's system. Table S12.1 and S12.2 show the parameters used to conduct the simulations. These values are based on varying system parameters until sensible behavior was resembled, given the information on our hypothetical client.

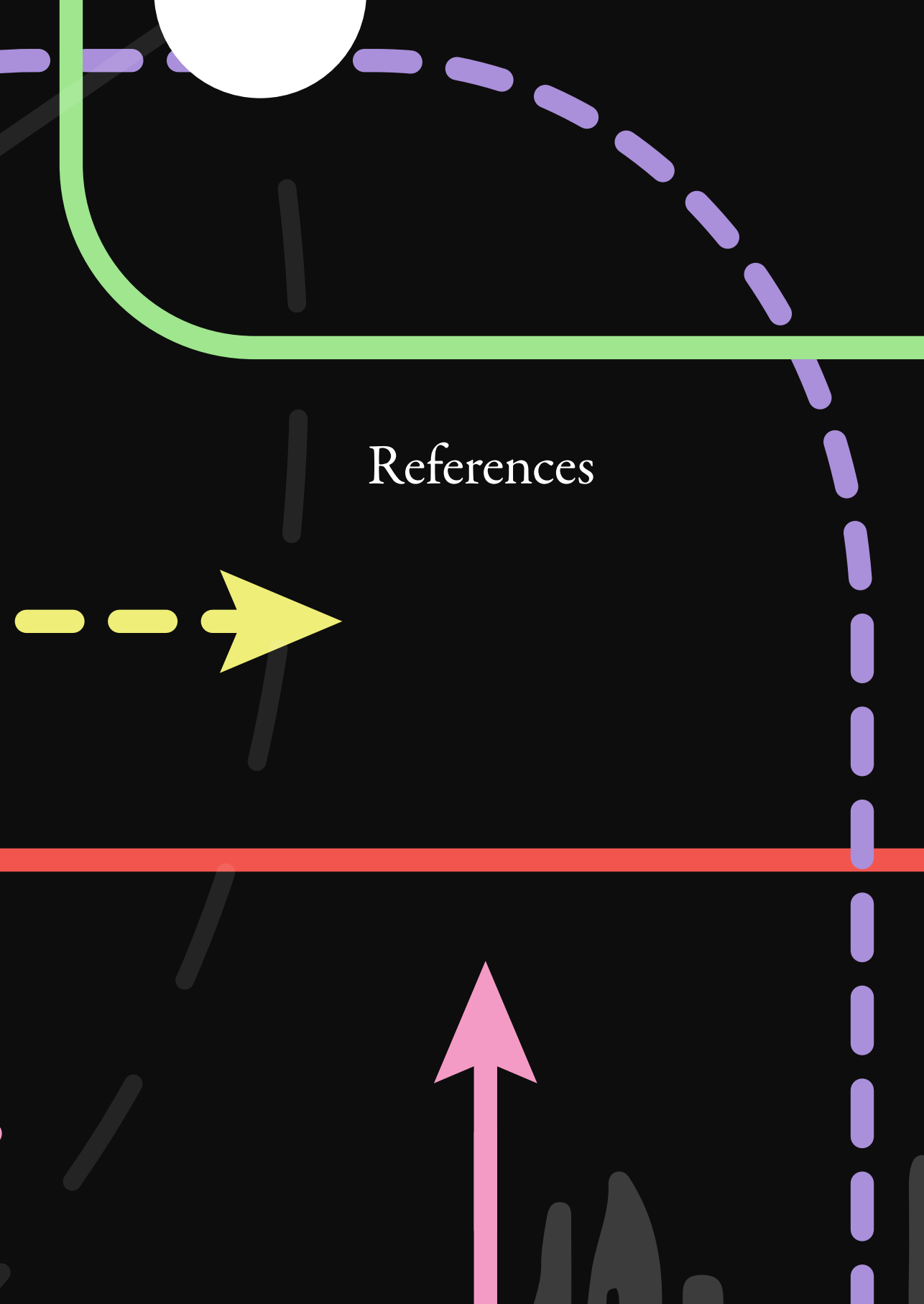
Table S12.1
Parameter choices for differential equations

	Value	Explanation of parameter: *Impact of...on... (sign)*
a	1.5	Discriminant stimulus on catastrophizing (+)
b	2.1	Credibility of catastrophic interpretation on catastrophizing (+)
c	1.65	Decay of catastrophizing (-)
d	2	Catastrophizing on panic (+)
e	0.2	Panic-avoidance interaction on panic (-)
f	0.5	Decay of panic (-)
g	0.2	Panic-avoidance interaction on avoidance (+)
h	0.19	Perceived costs on avoidance (-)
i	0.2	Perceived benefits on avoidance (+)
j	0.22	Decay of avoidance (-)
k	1	Evaluation of avoidance on perceived benefits (+)
l	0.02	Decay of perceived benefits (-)
m	0.2	Evaluation of avoidance on credibility of catastrophic interpretation (+)
n	0.04	Decay of credibility of catastrophic interpretation (-)
o	0.2	Avoidance on perceived costs (+)
p	0.5	Decay of perceived costs (-)
q	0.5	Credibility of catastrophic interpretation on functional interpretation (+)
r	0.02	Growth of functional interpretation (+)
s	0.1	Functional interpretation on catastrophizing (-)
t	0.3	Functional interpretation on credibility of catastrophic interpretation (-)

Table S12.2
Initial values for all system variables

Variable	Initial Value
Discriminant Stimulus *(Sd)*	0
Catastrophizing *(Cat)*	0.01
Panic *(Pan)*	0.01
Avoidance *(Av)*	0.5
Perceived benefits *(Ben)*	0.01
Perceived costs *(Cost)*	0.01
Credibility of catastrophic interpretation *(Cred)*	0.01
Credibility of functional interpretation *(FunCog)*	0




References

References

Aalbers, G., McNally, R. J., Heeren, A., de Wit, S., & Fried, E. I. (2019). Social media and depression symptoms: A network perspective. *Journal of Experimental Psychology: General*, *148*, 1454–1462. https://doi.org/10.1037/xge0000528

Abegaz, F., & Wit, E. (2013). Sparse time series chain graphical models for reconstructing genetic networks. *Biostatistics*, *14*(3), 586–599. https://doi.org/10.1093/biostatistics/kxt005

Abegaz, F., & Wit, E. (2021). *SparseTSCGM: Sparse time series chain graphical models [R package version 4.0]. Https://CRAN.R-project.org/package=SparseTSCGM* [Computer software].

Allen, M. S., Iliescu, D., & Greiff, S. (2022). Single item measures in psychological science: A call to action. *European Journal of Psychological Assessment*, *38*, 1–5. https://doi.org/10.1027/1015-5759/a000699

Alvaro, P. K., Roberts, R. M., & Harris, J. K. (2013). A Systematic Review Assessing Bidirectionality between Sleep Disturbances, Anxiety, and Depression. *Sleep*, *36*(7), 1059–1068. https://doi.org/10.5665/sleep.2810

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (Vol. 5).

American Psychological Association. (2020). *Publication Manual of the American Psychological Association* (7th ed.).

Aminikhanghahi, S., & Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowledge and Information Systems*, *51*(2), 339–367. https://doi.org/10.1007/s10115-016-0987-z

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, *567*(7748), 305–307. https://doi.org/10.1038/d41586-019-00857-9

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA publications and Communications Board task force report. *American Psychologist*. https://doi.org/10.1037/amp0000191

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 359–388. https://doi.org/10.1080/10705511.2017.1406803

Baker, D. B., & Benjamin, L. T. (2000). The affirmation of the scientist-practitioner. A look back at Boulder. *The American Psychologist*, *55*(2), 241–247. https://doi.org/10.1037//0003-066x.55.2.241

Bandelow, B., & Michaelis, S. (2015). Epidemiology of anxiety disorders in the 21st century. *Dialogues in Clinical Neuroscience*, *17*(3), 327–335.

Bansal, P., Bertels, S., Ewart, T., MacConnachie, P., & O'Brien, J. (2012). Bridging the Research–Practice Gap. *Academy of Management Perspectives*, *26*(1), 73–92. https://doi.org/10.5465/amp.2011.0140

Barabási, A.-L. (2012). The network takeover. *Nature Physics*, *8*(1), Article 1. https://doi.org/10.1038/nphys2188

Barlow, D. H. (1997). Cognitive-behavioral therapy for panic disorder: Current status. *The Journal of Clinical Psychiatry*, *58 Suppl 2*, 32–36; discussion 36-37.

Barlow, D. H. (2021). *Clinical Handbook of Psychological Disorders: A Step-by-Step Treatment Manual*. Guilford Publications.

Barlow, D. H., & Nock, M. K. (2009). Why Can't We Be More Idiographic in Our Research? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *4*(1), 19–21. https://doi.org/10.1111/j.1745-6924.2009.01088.x

Bastiaansen, J. A., Kunkels, Y. K., Blaauw, F. J., Boker, S. M., Ceulemans, E., Chen, M., Chow,

S.-M., de Jonge, P., Emerencia, A. C., Epskamp, S., Fisher, A. J., Hamaker, E. L.,

Kuppens, P., Lutz, W., Meyer, M. J., Moulder, R., Oravecz, Z., Riese, H., Rubel, J., Ryan,

O., Servaas, M. N., Sjobeck, G., Snippe, E., Trull, T. J., Tschacher, W., van der Veen, D.

C., Wichers, M., Wood, P. K., Woods, W. C., Wright, A. G. C., Aalbers, C. J. &

Bringmann, L. F. (2020). Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. *Journal of psychosomatic research*, *137*, 110211.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beard, C., Millner, A. J., Forgeard, M. J. C., Fried, E. I., Hsu, K. J., Treadway, M. T., Leonard, C. V., Kertz, S. J., & Björgvinsson, T. (2016). Network analysis of depression and anxiety symptom relationships in a psychiatric sample. *Psychological Medicine*, *46*(16), 3359–3369. https://doi.org/10.1017/S0033291716002300

Beck, A. T. (1979). *Cognitive therapy of depression*. Guilford Press.

Bekhuis, E., Schoevers, R., de Boer, M., Peen, J., Dekker, J., Van, H., & Boschloo, L. (2018). Symptom-Specific Effects of Psychotherapy versus Combined Therapy in the Treatment of Mild to Moderate Depression: A Network Approach. *Psychotherapy and Psychosomatics*, *87*(2), 121–123. https://doi.org/10.1159/000486793

Beltz, A. M., & Molenaar, P. C. M. (2016). Dealing with Multiple Solutions in Structural Vector Autoregressive Models. *Multivariate Behavioral Research*, *51*(2–3), 357–373. https://doi.org/10.1080/00273171.2016.1151333

Benke, C., Autenrieth, L. K., Asselmann, E., & Pané-Farré, C. A. (2020). Lockdown, quarantine measures, and social distancing: Associations with depression, anxiety and distress at the beginning of the COVID-19 pandemic among adults from Germany. *Psychiatry Research*, *293*, 113462. https://doi.org/10.1016/j.psychres.2020.113462

Bergh, R. van den, Olthof, M., Goldbeck, F., Hegewald, K., Pommerien-Becht, F., Daniels-Wredenhagen, N., Weggemans, R. J., Daalmans, S., Otten, R., Aas, B., & Lichtwarck-Aschoff, A. (2022). *The Content of Personalised Network-Based Case Formulations*. PsyArXiv. https://doi.org/10.31234/osf.io/yan4k

Berlim, M. T., Richard-Devantoy, S., Dos Santos, N. R., & Turecki, G. (2021). The network structure of core depressive symptom-domains in major depressive disorder following antidepressant treatment: A randomized clinical trial. *Psychological Medicine*, *51*(14), 2399–2413. https://doi.org/10.1017/S0033291720001002

Bernstein, E. E., Heeren, A., & McNally, R. J. (2019). Reexamining trait rumination as a system of repetitive negative thoughts: A network analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, *63*, 21–27. https://doi.org/10.1016/j.jbtep.2018.12.005

Blackwell, S. E., Woud, M. L., Margraf, J., & Schönbrodt, F. D. (2019). Introducing the leapfrog design: A simple Bayesian adaptive rolling trial design for accelerated treatment development and optimization. *Clinical Psychological Science*, *7*(6), 1222–1243. https://doi.org/10.1177/2167702619858071

Blanchard, A., & Heeren, A. (2022). *Ongoing and future challenges of the network approach to psychopathology: From theoretical conjectures to clinical translations*. https://dial.uclouvain.be/pr/boreal/object/boreal:237881

Blanco, I., Contreras, A., Chaves, C., Lopez-Gomez, I., Hervas, G., & Vazquez, C. (2020). Positive interventions in depression change the structure of well-being and psychological symptoms: A network analysis. *The Journal of Positive Psychology*, *15*, 1–6. https://doi.org/10.1080/17439760.2020.1789696

Blanken, T. F., Borsboom, D., Penninx, B. W. J. H., & Van Someren, E. J. W. (2020). Network outcome analysis identifies difficulty initiating sleep as a primary target for prevention of depression: A 6-year prospective study. *Sleep*, *43*(5), zsz288.

Blanken, T. F., Isvoranu, A.-M., & Epskamp, S. (2022). Estimating Network Structures using Model Selection. In A.-M. Isvoranu, S. Epskamp, L. J. Waldorp, & D. Borsboom (Eds.), *Network Psychometrics with R: A Guide for Behavioral and Social Scientists*. Routledge, Taylor & Francis Group.

REFERENCES

Blanken, T. F., Jansson-Fröjmark, M., Sunnhed, R., & Lancee, J. (2021). Symptom-specific effects of cognitive therapy and behavior therapy for insomnia: A network intervention analysis. *Journal of Consulting and Clinical Psychology*, *89*(4), 364–370. https://doi.org/10.1037/ccp0000625

Blanken, T. F., Van Der Zweerde, T., Van Straten, A., Van Someren, E. J. W., Borsboom, D., & Lancee, J. (2019). Introducing Network Intervention Analysis to Investigate Sequential, Symptom-Specific Treatment Effects: A Demonstration in Co-Occurring Insomnia and Depression. *Psychotherapy and Psychosomatics*, *88*(1), 52–54. https://doi.org/10.1159/000495045

Bogen, J., & Woodward, J. (1988). Saving the Phenomena. *The Philosophical Review*, *97*(3), 303. https://doi.org/10.2307/2185445

Booij, M. M., van Noorden, M. S., van Vliet, I. M., Ottenheim, N. R., van der Wee, N. J. A., Van Hemert, A. M., & Giltay, E. J. (2021). Dynamic time warp analysis of individual symptom trajectories in depressed patients treated with electroconvulsive therapy. *Journal of Affective Disorders*, *293*, 435–443. https://doi.org/10.1016/j.jad.2021.06.068

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *16*(1), 5–13. https://doi.org/10.1002/wps.20375

Borsboom, D., & Cramer, A. O. J. (2013). Network Analysis: An Integrative Approach to the Structure of Psychopathology. *Annual Review of Clinical Psychology*, *9*(1), 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Borsboom, D., Cramer, A. O. J., & Kalis, A. (2018). Brain disorders? Not really: Why network structures block reductionism in psychopathology research. *The Behavioral and Brain Sciences*, *42*, e2. https://doi.org/10.1017/S0140525X17002266

Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The Small World of Psychopathology. *PLOS ONE*, *6*(11), e27407. https://doi.org/10.1371/journal.pone.0027407

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G., Isvoranu, A.-M., Wysocki, A. C., van Borkulo, C. D., van Bork, R., & Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, *1*(1), 1–18. https://doi.org/10.1038/ s43586-021-00055-w

Borsboom, D., Fried, E. I., Epskamp, S., Waldorp, L. J., van Borkulo, C. D., van der Maas, H. L. J., & Cramer, A. O. J. (2017). False alarm? A comprehensive reanalysis of "Evidence that psychopathology symptom networks have limited replicability" by Forbes, Wright, Markon, and Krueger. *Journal of Abnormal Psychology*, *26*(7), 989-999. https://doi.org/10.1037/abn0000306

Borsboom, D., van der Maas, H. L. J., Dalege, J., Kievit, R. A., & Haig, B. D. (2021a). Theory Construction Methodology: A Practical Framework for Building Theories in Psychology. *Perspectives on Psychological Science*, 1745691620969647. https://doi.org/10.1177/1745691620969647

Borsboom, D., van der Maas, H. L. J., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, 1–11. https://doi.org/10.1177/1745691620969647

Bos, F. (2021). *Ecological momentary assessment as a clinical tool in psychiatry: Promises, pitfalls, and possibilities* [University of Groningen]. https://doi.org/10.33612/diss.177817937

Bos, F. M., Fried, E. I., Hollon, S. D., Bringmann, L. F., Dimidjian, S., DeRubeis, R. J., & Bockting, C. L. H. (2018). Cross-sectional networks of depressive symptoms before and after antidepressant medication treatment. *Social Psychiatry and Psychiatric Epidemiology*, *53*(6), 617–627. https://doi.org/10.1007/s00127-018-1506-1

Bos, F. M., Snippe, E., de Vos, S., Hartmann, J. A., Simons, C. J. P., van der Krieke, L., de Jonge, P., & Wichers, M. (2017). Can We Jump from Cross-Sectional to Dynamic Interpretations of Networks Implications for the Network Perspective in Psychiatry. *Psychotherapy and Psychosomatics*, *86*(3), 175–177. https://doi.org/10.1159/000453583

Bos, F. M., von Klipstein, L., Emerencia, A. C., Veermans, E., Verhage, T., Snippe, E., Doornbos, B., Hadders-Prins, G., Wichers, M., & Riese, H. (2022). A Web-Based Application for Personalized Ecological Momentary Assessment in Psychiatric Care: User-Centered Development of the PETRA Application. *JMIR Mental Health*, *9*(8), e36430. https://doi.org/10.2196/36430

Boschloo, L., Bekhuis, E., Weitz, E. S., Reijnders, M., DeRubeis, R. J., Dimidjian, S., Dunner, D. L., Dunlop, B. W., Hegerl, U., Hollon, S. D., Jarrett, R. B., Kennedy, S. H., Miranda, J., Mohr, D. C., Simons, A. D., Parker, G., Petrak, F., Herpertz, S., Quilty, L. C., … Cuijpers, P. (2019). The symptom-specific efficacy of antidepressant medication vs. cognitive behavioral therapy in the treatment of depression: Results from an individual patient data meta-analysis. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *18*(2), 183–191. https://doi.org/10.1002/wps.20630

Boschloo, L., Cuijpers, P., Karyotaki, E., Berger, T., Moritz, S., Meyer, B., & Klein, J. P. (2019). Symptom-specific effectiveness of an internet-based intervention in the treatment of mild to moderate depressive symptomatology: The potential of network estimation techniques. *Behaviour Research and Therapy*, *122*, 103440. https://doi.org/10.1016/j.brat.2019.103440

Briganti, G., Kornreich, C., & Linkowski, P. (2021). A network structure of manic symptoms. *Brain and Behavior*, *11*(3), e02010. https://doi.org/10.1002/brb3.2010

Bringmann, L. F. (2021). Person-specific networks in psychopathology: Past, present, and future. *Current Opinion in Psychology*, *41*, 59–64. https://doi.org/10.1016/j.copsyc.2021.03.004

Bringmann, L. F., Albers, C., Bockting, C., Borsboom, D., Ceulemans, E., Cramer, A., Epskamp, S., Eronen, M. I., Hamaker, E., Kuppens, P., Lutz, W., McNally, R. J., Molenaar, P., Tio, P., Voelkle, M. C., & Wichers, M. (2022). Psychopathological networks: Theory, methods and practice. *Behaviour Research and Therapy*, *149*. https://doi.org/10.1016/j.brat.2021.104011

Bringmann, L. F., Elmer, T., Epskamp, S., Krause, W., Schoch, D., Wichers, M., & Snippe, E. (2019). What do centrality measures measure in psychological networks. *Journal of Abnormal Psychology*, *128*(8), 892–90. https://doi.org/10.1037/abn0000446

Bringmann, L. F., Lemmens, L. H. J. M., Huibers, M. J. H., Borsboom, D., & Tuerlinckx, F. (2015). Revealing the dynamic network structure of the Beck Depression Inventory-II. *Psychological Medicine*, *45*(4), 747–757. https://doi.org/10.1017/S0033291714001809

Bringmann, L. F., van der Veen, D. C., Wichers, M., Riese, H., & Stulp, G. (2020). ESMvis: A tool for visualizing individual Experience Sampling Method (ESM) data. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*. https://doi.org/10.1007/s11136-020-02701-4

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). A Network Approach to Psychopathology: New Insights into Clinical Longitudinal Data. *PLOS ONE*, *8*(4), e60188. https://doi.org/10.1371/journal.pone.0060188

Brose, A., Lindenberger, U., & Schmiedek, F. (2013). Affective states contribute to trait reports of affective well-being. *Emotion*, *13*(5), 940–948. https://doi.org/10.1037/a0032401

Brose, A., & Ram, N. (2012). Within-person factor analysis: Modeling how the individual fluctuates and changes across time. In *Handbook of research methods for studying daily life* (pp. 459–478). The Guilford Press.

REFERENCES

Bulteel, K., Tuerlinckx, F., Brose, A., & Ceulemans, E. (2016). Using Raw VAR Regression Coefficients to Build Networks can be Misleading. *Multivariate Behavioral Research*, *51*(2–3), 330–344. https://doi.org/10.1080/00273171.2016.1150151

Burger, J., Epskamp, S., van der Veen, D. C., Dablander, F., Schoevers, R. A., Fried, E. I., & Riese, H. (2022). A clinical PREMISE for personalized models: Towards a formal integration of case formulations and statistical networks. *Journal of Psychopathology and Clinical Science 131*(8). 906-916. https://doi.org/10.1037/abn0000779

Burger, J., Hoekstra, R. H., Mansueto, A. C., & Epskamp, S. (2022). Network estimation from time-series and panel data. In A.-M. Isvoranu, S. Epskamp, L. J. Waldorp, & D. Borsboom (Eds.), *Network Psychometrics with R: A Guide for Behavioral and Social Scientists*. Routledge, Taylor & Francis Group.

Burger, J., Stroebe, M. S., Perrig-Chiello, P., Schut, H. A., Spahni, S., Eisma, M. C., & Fried, E. I. (2020). Bereavement or breakup: Differences in networks of depression. *Journal of Affective Disorders*, *267*, 1–8. https://doi.org/10.1016/j.jad.2020.01.157

Calugi, S., Dametti, L., Chimini, M., Dalle Grave, A., & Dalle Grave, R. (2021). Change in eating-disorder psychopathology network structure in patients with anorexia nervosa treated with intensive cognitive behavior therapy. *International Journal of Eating Disorders*, *54*, 1800–1809. https://doi.org/10.1002/eat.23590

Camacho, T. C., Roberts, R. E., Lazarus, N. B., Kaplan, G. A., & Cohen, R. D. (1991). Physical Activity and Depression: Evidence from the Alameda County Study. *American Journal of Epidemiology*, *134*(2), 220–231. https://doi.org/10.1093/oxfordjournals.aje.a116074

Carleton, R. N., Mulvogue, M. K., Thibodeau, M. A., McCabe, R. E., Antony, M. M., & Asmundson, G. J. G. (2012). Increasingly certain about uncertainty: Intolerance of uncertainty across anxiety and depression. *Journal of Anxiety Disorders*, *26*(3), 468–479. https://doi.org/10.1016/j.janxdis.2012.01.011

Carleton, R. N., Norton, M. A. P. J., & Asmundson, G. J. G. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders*, *21*(1), 105–117. https://doi.org/10.1016/j.janxdis.2006.03.014

Carney, R. M., Freedland, K. E., Steinmeyer, B. C., Rubin, E. H., & Rich, M. W. (2018). Residual Symptoms Following Treatment for Depression in Patients with Coronary Heart Disease. *Psychosomatic Medicine*, *80*(4), 385–392. https://doi.org/10.1097/PSY.0000000000000570

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). *Stan*: A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Çelik, N., Ceylan, B., Ünsal, A., & Çağan, Ö. (2019). Depression in health college students: Relationship factors and sleep quality. *Psychology, Health & Medicine*, *24*(5), 625–630. https://doi.org/10.1080/13548506.2018.1546881

Cervin, M., Storch, E. A., Piacentini, J., Birmaher, B., Compton, S. N., Albano, A. M., Gosch, E., Walkup, J. T., & Kendall, P. C. (2020). Symptom-specific effects of cognitive-behavioral therapy, sertraline, and their combination in a large randomized controlled trial of pediatric anxiety disorders. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *61*(4), 492–502. https://doi.org/10.1111/jcpp.13124

Chen, G., Glen, D. R., Saad, Z. S., Paul Hamilton, J., Thomason, M. E., Gotlib, I. H., & Cox, R. W. (2011). Vector autoregression, structural equation modeling, and their synthesis in neuroimaging data analysis. *Computers in Biology and Medicine*, *41*(12), 1142–1155. https://doi.org/10.1016/j.compbiomed.2011.09.004

Cheney, W. D. P., Carl D. (2017). *Behavior Analysis and Learning: A Biobehavioral Approach, Sixth Edition* (6th ed.). Routledge. https://doi.org/10.4324/9781315200682

Cheng, C., Ebrahimi, O. V., & Lau, Y. (2021). Maladaptive coping with the infodemic and sleep disturbance in the COVID-19 pandemic. *Journal of Sleep Research*, *30*(4), e13235. https://doi.org/10.1111/jsr.13235

Chesney, E., Goodwin, G. M., & Fazel, S. (2014). Risks of all-cause and suicide mortality in mental disorders: A meta-review. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *13*(2), 153–160. https://doi.org/10.1002/wps.20128

Chow, S.-M. (2019). Practical Tools and Guidelines for Exploring and Fitting Linear and Nonlinear Dynamical Systems Models. *Multivariate Behavioral Research*, *54*(5), 690–718. https://doi.org/10.1080/00273171.2019.1566050

Ciraki, D. (2007). *Dynamic structural equation models: Estimation and interference*.

Clarke, D. M., & Kissane, D. W. (2002). Demoralization: Its Phenomenology and Importance. *Australian & New Zealand Journal of Psychiatry*, *36*(6), 733–742. https://doi.org/10.1046/j.1440-1614.2002.01086.x

Collins, A. C., Lass, A. N. S., Jordan, D. G., & Winer, E. S. (2021). Examining rumination, devaluation of positivity, and depressive symptoms via community-based network analysis. *Journal of Clinical Psychology*, *77*(10), 2228–2244. https://doi.org/10.1002/jclp.23158

Copen, C. E., & Mosher, W. D. (2012). *First Marriages in the United States: Data From the 2006-2010 National Survey of Family Growth*. *49*.

Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mõttus, R., Waldorp, L. J., & Cramer, A. O. J. (2015). State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality*, *54*, 13–29. https://doi.org/10.1016/j.jrp.2014.07.003

Cramer, A. O. J., Borkulo, C. D. van, Giltay, E. J., Maas, H. L. J. van der, Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major Depression as a Complex Dynamic System. *PLOS ONE*, *11*(12), e0167490. https://doi.org/10.1371/journal.pone.0167490

Cramer, A. O. J., Borsboom, D., Aggen, S. H., & Kendler, K. S. (2012). The pathoplasticity of dysphoric episodes: Differential impact of stressful life events on the pattern of depressive symptom inter-correlations. *Psychological Medicine*, *42*(5), 957–965. https://doi.org/10.1017/S003329171100211X

Curran, P. J., & Bauer, D. J. (2011). The Disaggregation of Within-Person and Between-Person Effects in Longitudinal Models of Change. *Annual Review of Psychology*, *62*(1), 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

Curtiss, J. E., Wallace, B., Fisher, L. B., Nyer, M., Jain, F., Cusin, C., & Pedrelli, P. (2021). Change processes in cognitive behavioral therapy and motivational interviewing for depression and heavy alcohol use: A network approach. *Journal of Affective Disorders Reports*, *3*, 100040. https://doi.org/10.1016/j.jadr.2020.100040

Dablander, F. (2020). *An Introduction to Causal Inference*. PsyArXiv. https://doi.org/10.31234/osf.io/b3fkw

Dablander, F., & Hinne, M. (2019). Node centrality measures are a poor substitute for causal inference. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-43033-9

David, S. J., Marshall, A. J., Evanovich, E. K., & Mumma, G. H. (2018). Intraindividual Dynamic Network Analysis—Implications for Clinical Assessment. *Journal of Psychopathology and Behavioral Assessment*, *40*(2), 235–248. https://doi.org/10.1007/s10862-017-9632-8

de Ron, J., Fried, E. I., & Epskamp, S. (2020). Psychological Networks in Clinical Populations: A tutorial on the consequences of Berkson's Bias. *Psychological Medicine*, *51*(1), 168–176. https://doi.org/10.1017/S0033291719003209

Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019). Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature Human Behaviour*, *3*(5), 478–491. https://doi.org/10.1038/s41562-019-0555-0

REFERENCES

Deloitte Access Economics. (2020). *Social and economic cost of eating disorders in the United States of America*. 92.

Demakakos, P. (2021). Importance of population-based longitudinal studies to understanding the impact of COVID-19. *J Epidemiol Community Health*, *75*(9), 815–816. https://doi.org/10.1136/jech-2021-217114

Deserno, M. K., Borsboom, D., Begeer, S., Bork, R. van, Hinne, M., & Geurts, H. M. (2020). Highways to happiness for autistic adults? Perceived causal relations among clinicians. *PLOS ONE*, *15*(12), e0243298. https://doi.org/10.1371/journal.pone.0243298

Deserno, M. K., Borsboom, D., Begeer, S., & Geurts, H. M. (2017). Multicausal systems ask for multicausal approaches: A network perspective on subjective well-being in individuals with autism spectrum disorder. *Autism*, *21*(8), 960–971.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, *74*(366a), 427–431. https://doi.org/10.1080/01621459.1979.10482531

Djelantik, A. A. A. M. J., Robinaugh, D. J., Kleber, R. J., Smid, G. E., & Boelen, P. A. (2020). Symptomatology following loss and trauma: Latent class and network analyses of prolonged grief disorder, posttraumatic stress disorder, and depression in a treatment-seeking trauma-exposed sample. *Depression and Anxiety*, *37*(1), 26–34. https://doi.org/10.1002/da.22880

Dotterer, H. L., Beltz, A. M., Foster, K. T., Simms, L. J., & Wright, A. G. C. (2020). Personalized models of personality disorders: Using a temporal network method to understand symptomatology and daily functioning in a clinical sample. *Psychological Medicine*, *50*(14), 2397–2405. https://doi.org/10.1017/S0033291719002563

Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous Time Structural Equation Modeling with *R* Package **ctsem**. *Journal of Statistical Software*, *77*(5). https://doi.org/10.18637/jss.v077.i05

Dujmić, Z., Machielse, E., & Treur, J. (2019). A Temporal-Causal Modeling Approach to the Dynamics of a Burnout and the Role of Physical Exercise: 9th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2018. *Biologically Inspired Cognitive Architectures 2018*, 88–100. https://doi.org/10.1007/978-3-319-99316-4_12

Ebrahimi, O. V., Burger, J., Hoffart, A., & Johnson, S. U. (2021). Within- and across-day patterns of interplay between depressive symptoms and related psychopathological processes: A dynamic network approach during the COVID-19 pandemic. *BMC Medicine*, *19*(1), 317. https://doi.org/10.1186/s12916-021-02179-y

Ebrahimi, O. V., Hoffart, A., & Johnson, S. U. (2021). Physical Distancing and Mental Health During the COVID-19 Pandemic: Factors Associated With Psychological Symptoms and Adherence to Pandemic Mitigation Strategies. *Clinical Psychological Science*, *9*(3), 489–506. https://doi.org/10.1177/2167702621994545

Eckert. (2022). *ParallelDist*. R package version 0.2.6.

Eells, T. D. (2022). *Handbook of Psychotherapy Case Formulation*. Guilford Publications.

Efron, B. (1979). Computers and the Theory of Statistics: Thinking the Unthinkable. *SIAM Review*, *21*(4), 460–480. https://doi.org/10.1137/1021092

Eichler, M. (2007). Granger causality and path diagrams for multivariate time series. *Journal of Econometrics*, *137*(2), 334–353. https://doi.org/10.1016/j.jeconom.2005.06.032

Eisma, M. C. (2018). Public stigma of prolonged grief disorder: An experimental study. *Psychiatry Research*, *261*, 173–177. https://doi.org/10.1016/j.psychres.2017.12.064

Elliott, H., Jones, P. J., & Schmidt, U. (2020). Central Symptoms Predict Posttreatment Outcomes and Clinical Impairment in Anorexia Nervosa: A Network Analysis. *Clinical Psychological Science*, *8*(1), 139–154. https://doi.org/10.1177/2167702619865958

Epskamp, S. (2019). Reproducibility and replicability in a fast-paced methodological world. *Advances in Methods and Practices in Psychological Science*, *2*(2), 145–155. https://doi.org/10.1177/2515245919847421

Epskamp, S. (2020a). *graphicalVAR: Graphical VAR for experience sampling data [R package version 0.3]. Https://CRAN.R-project.org/package=graphicalVAR* [Computer software].

Epskamp, S. (2020b). Psychometric network models from time-series and panel data. *Psychometrika*, *85*(1), 206–231. https://doi.org/10.1007/s11336-020-09697-3

Epskamp, S. (2020c). *Software updates: Bootnet 1.3 and psychonetrics 0.5, retrieved from http://psychonetrics.org/2020/01/26/software-updates-bootnet-1-3-and-psychonetrics-0-5/#New_default_behavior*.

Epskamp, S. (2022). *psychonetrics: Structural Equation Modeling and Confirmatory Network Analysis*.

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1

Epskamp, S., & Bringmann, M. K. D. and L. F. (2019). *mlVAR: Multi-Level Vector Autoregression*. R package version 0.4.4. https://CRAN.R-project.org/package=mlVAR

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software*, *48*, 1–18. https://doi.org/10.18637/jss.v048.i04

Epskamp, S., Deserno, M. K., & Bringmann, L. F. (2021). *Mlvar: Multi-level vector autoregression [R package version 0.5]. Https://CRAN.R-project.org/package= mlVAR* [Computer software].

Epskamp, S., Deserno, M. K., Bringmann, L. F., & Veenman, M. (2019). *mlVAR: Multi-Level Vector Autoregression* R package version 0.5.1. https://cran.r-project.org/web/packages/mlVAR/index.html

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*(4), 617–634. https://doi.org/10.1037/met0000167

Epskamp, S., Haslbeck, J. M. B., Isvoranu, A.-M., & Van Borkulo, C. D. (2022). Pairwise Markov Random Fields. In *Network Psychometrics with R*. Routledge.

Epskamp, S., Isvoranu, A.-M., & Cheung, M. W.-L. (2022). Meta-analytic Gaussian Network Aggregation. *Psychometrika*, *87*(1), 12–46. https://doi.org/10.1007/s11336-021-09764-3

Epskamp, S., Kruis, J., & Marsman, M. (2017). Estimating psychopathological networks: Be careful what you wish for. *PLoS ONE*, *12*(6). https://doi.org/10.1371/journal.pone.0179891

Epskamp, S., Maris, G., Waldorp, L. J., & Borsboom, D. (2018). Network Psychometrics. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development* (pp. 953–986).

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized Network Psychometrics: Combining Network and Latent Variable Models. *Psychometrika*, *82*(4), 904–927. https://doi.org/10.1007/s11336-017-9557-x

Epskamp, S., Van Borkulo, C. D., Van Der Veen, D. C., Servaas, M. N., Isvoranu, A.-M., Riese, H., & Cramer, A. O. J. (2018b). Personalized Network Modeling in Psychopathology: The Importance of Contemporaneous and Temporal Connections. *Clinical Psychological Science*, *6*(3), 416–427. https://doi.org/10.1177/2167702617744325

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The Gaussian Graphical Model in Cross-Sectional and Time-Series Data. *Multivariate Behavioral Research*, *53*(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823

Epstein, J. M. (2008). Why Model? *Journal of Artificial Societies and Social Simulation*, *11*(4). https://econpapers.repec.org/article/jasjasssj/2008-57-1.htm

Esfahlani, F. Z., Visser, K., Strauss, G., & Sayama, H. (2018). A Network-Based Classification Framework for Predicting Treatment Response of Schizophrenia Patients. *Expert Systems with Applications*, *109*. https://doi.org/10.1016/j.eswa.2018.05.005

REFERENCES

Esfahlani, F. Z., Sayama, H., Visser, K. F., & Strauss, G. P. (2017). Sensitivity of the Positive and Negative Syndrome Scale (PANSS) in Detecting Treatment Effects via Network Analysis. *Innovations in Clinical Neuroscience*, *14*(11–12), 59–67.

Ettman, C. K., Abdalla, S. M., Cohen, G. H., Sampson, L., Vivier, P. M., & Galea, S. (2020). Prevalence of Depression Symptoms in US Adults Before and During the COVID-19 Pandemic. *JAMA Network Open*, *3*(9), e2019686. https://doi.org/10.1001/jamanetworkopen.2020.19686

Fairburn, C. G., Cooper, Z., & Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: A "transdiagnostic" theory and treatment. *Behaviour Research and Therapy*, *41*(5), 509–528. https://doi.org/10.1016/S0005-7967(02)00088-8

Finnemann, A., Borsboom, D., Epskamp, S., & van der Maas, H. L. J. (2021). The Theoretical and Statistical Ising Model: A Practical Guide in R. *Psych*, *3*(4), Article 4. https://doi.org/10.3390/psych3040039

First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2015). *Structured Clinical Interview for DSM-5 Disorders, Clinical Trials Version (SCID-5-CT). American Psychiatric Association, Arlington, VA.*

Fisher, A. J. (2015). Toward a dynamic model of psychological assessment: Implications for personalized care. *Journal of Consulting and Clinical Psychology*, *83*(4), 825–836. https://doi.org/10.1037/ccp0000026

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, *115*(27), E6106–E6115. https://doi.org/10.1073/pnas.1711978115

Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the idiographic dynamics of mood and anxiety via network analysis. *Journal of Abnormal Psychology*, *126*, 1044–1056. https://doi.org/10.1037/abn0000311

Flake, J. K., & Fried, E. I. (2019). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *PsyArXiv (Preprint)*. https://doi.org/10.31234/osf.io/hs7wm

Foygel, R., & Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010, NIPS 2010*.

Frank, J. D. (1974). Psychotherapy: The Restoration of Morale. *American Journal of Psychiatry*, *131*(3), 271–274. https://doi.org/10.1176/ajp.131.3.271

Frewen, P. A., Allen, S. L., Lanius, R. A., & Neufeld, R. W. J. (2012). Perceived causal relations: Novel methodology for assessing client attributions about causal associations between variables including symptoms and functional impairment. *Assessment*, *19*(4), 480–493. https://doi.org/10.1177/1073191111418297

Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, *208*, 191–197. https://doi.org/10.1016/j.jad.2016.10.019

Fried, E. I. (2020a). Lack of Theory Building and Testing Impedes Progress in The Factor and Network Literature. *Psychological Inquiry*, *31*(4), 271–288. https://doi.org/10.1080/1047840X.2020.1853461

Fried, E. I. (2020b). Theories and Models: What They Are, What They Are for, and What They Are About. *Psychological Inquiry*, *31*(4), 336–344. https://doi.org/10.1080/1047840X.2020.1854011

Fried, E. I., Bockting, C., Arjadi, R., Borsboom, D., Amshoff, M., Cramer, A. O. J., Epskamp, S., Tuerlinckx, F., Carr, D., & Stroebe, M. (2015). From loss to loneliness: The relationship between bereavement and depressive symptoms. *Journal of Abnormal Psychology*, *124*, 256–265. https://doi.org/10.1037/abn0000028

Fried, E. I., & Cramer, A. O. J. (2017). Moving forward: Challenges and directino for psychopathological network theory and methodology. *Perspectives on Psychological Science*, *12*(6), 999–1020. https://doi.org/10.17605/OSF.IO/BNEK

Fried, E. I., Eidhof, M. B., Palic, S., Costantini, G., Dijk, H. M. H., Bockting, C. L. H., Engelhard, I., Armour, C., Nielsen, A. B. S., & Karstoft, K. (2018). Replicability and Generalizability of Posttraumatic Stress Disorder (PTSD) Networks: A Cross-Cultural Multisite Study of PTSD Symptoms in Four Trauma Patient Samples. *Clinical Psychological Science*, *6*(3), 335–351. https://doi.org/10.1177/2167702617745092

Fried, E. I., Epskamp, S., Veenman, M., & van Borkulo, C. D. (2022). Network Stability, Comparison, and Replicability. In A.-M. Isvoranu, S. Epskamp, L. J. Waldorp, & D. Borsboom (Eds.), *Network Psychometrics with R: A Guide for Behavioral and Social Scientists*.

Fried, E. I., & Nesse, R. M. (2014). The impact of individual depressive symptoms on impairment of psychosocial functioning. *PloS One*, *9*(2), e90311. https://doi.org/10.1371/journal.pone.0090311

Fried, E. I., & Nesse, R. M. (2015). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders*, *172*, 96–102. https://doi.org/10.1016/j.jad.2014.10.010

Fried, E. I., Nesse, R. M., Guille, C., & Sen, S. (2015). The differential influence of life stress on individual symptoms of depression. *Acta Psychiatrica Scandinavica*, *131*(6), 465–471. https://doi.org/10.1111/acps.12395

Fried, E. I., Papanikolaou, F., & Epskamp, S. (2022). Mental Health and Social Contact During the COVID-19 Pandemic: An Ecological Momentary Assessment Study. *Clinical Psychological Science*, *10*(2), 340–354. https://doi.org/10.1177/21677026211017839

Fried, E. I., van Borkulo, C. D., Cramer, A. O. J., Boschloo, L., Schoevers, R. A., & Borsboom, D. (2017). Mental disorders as networks of problems: A review of recent insights. *Social Psychiatry and Psychiatric Epidemiology*, *52*(1), 1–10. https://doi.org/10.1007/s00127-016-1319-z

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, *21*(11), 1129–1164. https://doi.org/10.1002/spe.4380211102

Frumkin, M. R., Piccirillo, M. L., Beck, E. D., Grossman, J. T., & Rodebaugh, T. L. (2021). Feasibility and Utility of Idiographic Models in the Clinic: A Pilot Study. *Psychotherapy Research : Journal of the Society for Psychotherapy Research*, *31*(4), 520–534. https://doi.org/10.1080/10503307.2020.1805133

Funkhouser, C. J., Chacko, A. A., Correa, K. A., Kaiser, A. J. E., & Shankman, S. A. (2021). Unique longitudinal relationships between symptoms of psychopathology in youth: A cross-lagged panel network analysis in the ABCD study. *Journal of Child Psychology and Psychiatry*, *62*(2), 184–194. https://doi.org/10.1111/jcpp.13256

Furusawa, C., & Kaneko, K. (2012). A dynamical-systems view of stem cell biology. *Science (New York, N.Y.)*, *338*(6104), 215–217. https://doi.org/10.1126/science.1224311

Gábor, A., & Banga, J. R. (2015). Robust and efficient parameter estimation in dynamic models of biological systems. *BMC Systems Biology*, *9*(1), 74. https://doi.org/10.1186/s12918-015-0219-2

Gates, K. M., & Molenaar, P. C. M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, *63*(1), 310–319. https://doi.org/10.1016/j.neuroimage.2012.06.026

Gates, K. M., Molenaar, P. C. M., Hillary, F. G., Ram, N., & Rovine, M. J. (2010). Automatic search for fMRI connectivity mapping: An alternative to Granger causality testing using formal equivalences among SEM path modeling, VAR, and unified SEM. *NeuroImage*, *50*(3), 1118–1125. https://doi.org/10.1016/j.neuroimage.2009.12.117

REFERENCES

Gierveld, J. de J. (2004). Remarriage, Unmarried Cohabitation, Living Apart Together: Partner Relationships Following Bereavement or Divorce. *Journal of Marriage and Family*, *66*(1), 236–243.

Giorgino, T. (2009). Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. *Journal of Statistical Software*, *31*, 1–24. https://doi.org/10.18637/jss.v031.i07

Goldberg, J. F., Siu, C., Mao, Y., Tsai, J., Pikalov, A., Calabrese, J. R., & Loebel, A. (2020). Major depressive disorder with mixed features and treatment response to lurasidone: A symptom network model. *Journal of Affective Disorders*, *277*, 1045–1054. https://doi.org/10.1016/j.jad.2020.08.048

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS ONE*, *12*(6), e0174035. https://doi.org/10.1371/journal.pone.0174035

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*(4), 323–343. https://doi.org/10.1037/1082-989X.11.4.323

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424–438. https://doi.org/10.2307/1912791

Grasman, J., Grasman, R. P. P. P., & Maas, H. L. J. van der. (2016). The Dynamics of Addiction: Craving versus Self-Control. *PLOS ONE*, *11*(6), e0158323. https://doi.org/10.1371/journal.pone.0158323

Groen, R. N., Snippe, E., Bringmann, L. F., Simons, C. J. P., Hartmann, J. A., Bos, E. H., & Wichers, M. (2019). Capturing the risk of persisting depressive symptoms: A dynamic network investigation of patients' daily symptom experiences. *Psychiatry Research*, *271*, 640–648. https://doi.org/10.1016/j.psychres.2018.12.054

Guest, O., & Martin, A. E. (2021). How Computational Modeling Can Force Theory Building in Psychological Science. *Perspectives on Psychological Science*, 1745691620970585. https://doi.org/10.1177/1745691620970585

Hagberg, A., Swart, P., & S Chult, D. (2008). *Exploring network structure, dynamics, and function using networkx* (LA-UR-08-05495; LA-UR-08-5495). Los Alamos National Lab. (LANL), Los Alamos, NM (United States). https://www.osti.gov/biblio/960616

Haig, B. D. (2005). An Abductive Theory of Scientific Method. *Psychological Methods*, *10*(4), 371–388. https://doi.org/10.1037/1082-989X.10.4.371

Haig, B. D. (2014). *Investigating the psychological world: Scientific method in the behavioral sciences*. MIT press.

Hamaker, E. L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In *Handbook of research methods for studying daily life* (pp. 43–61). The Guilford Press.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2002). On the Nature of SEM Estimates of ARMA Parameters. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(3), 347–368. https://doi.org/10.1207/S15328007SEM0903_3

Hamaker, E. L., & Grasman, R. P. P. P. (2012). Regime switching state-space models applied to psychological processes: Handling missing data and making inferences. *Psychometrika*, *77*, 400–422. https://doi.org/10.1007/s11336-012-9254-8

Hamaker, E. L., & Grasman, R. P. P. P. (2015). To center or not to center? Investigating inertia with a multilevel autoregressive model. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01492

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*, 102–116. https://doi.org/10.1037/a0038889

Hammen, C. (2005). Stress and Depression. *Annual Review of Clinical Psychology*, *1*(1), 293–319. https://doi.org/10.1146/annurev.clinpsy.1.102803.143938

Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press. https://doi.org/10.1017/CBO9781107049994

Harvey, A., Watkins, E., Mansell, W., & Shafran, R. (Eds.). (2004). *Cognitive Behavioural Processes across Psychological Disorders: A transdiagnostic approach to research and treatment*. Oxford University Press. https://doi.org/10.1093/med:psych/9780198528883.002.0005

Haslbeck, J. M. B. (2020). Estimating Group Differences in Network Models using Moderation Analysis. *PsyArXiv (Preprint)*. https://doi.org/10.31234/osf.io/926pv

Haslbeck, J. M. B., Borsboom, D., & Waldorp, L. J. (2019). Moderated Network Models. *Multivariate Behavioral Research*, 1–32. https://doi.org/10.1080/00273171.2019.1677207

Haslbeck, J. M. B., Bringmann, L. F., & Waldorp, L. J. (2021). A Tutorial on Estimating Time-Varying Vector Autoregressive Models. *Multivariate Behavioral Research*, *56*(1), 120–149. https://doi.org/10.1080/00273171.2020.1743630

Haslbeck, J. M. B., & Fried, E. I. (2017). How predictable are symptoms in psychopathological networks? A reanalysis of 18 published datasets. *Psychological Medicine*, *47*(16), 2767–2776. https://doi.org/10.1017/S0033291717001258

Haslbeck, J. M. B., & Ryan, O. (2022). Recovering Within-Person Dynamics from Psychological Time Series. *Multivariate Behavioral Research*, *57*(5), 735–766. https://doi.org/10.1080/00273171.2021.1896353

Haslbeck, J. M. B., Ryan, O., & Dablander, F. (2020). The Sum of All Fears: Comparing Networks Based on Symptom Sum-Scores. *Psychological Methods*, *accepted*.

Haslbeck, J. M. B., Ryan, O., Robinaugh, D. J., Waldorp, L. J., & Borsboom, D. (2021). Modeling psychopathology: From data models to formal theories. *Psychological Methods*, *27*(6), 930. https://doi.org/10.1037/met0000303

Haslbeck, J. M. B., Ryan, O., van der Maas, H. L. J., & Waldorp, L. J. (2022). Modeling Change in Networks. In *Network Psychometrics with R*. Routledge. https://doi.org/10.4324/9781003111238-14

Haslbeck, J. M. B., & Waldorp, L. J. (2018). How well do network models predict observations? On the importance of predictability in network models. *Behavior Research Methods*, *50*, 853–861. https://doi.org/10.3758/s13428-017-0910-x

Haslbeck, J. M. B., & Waldorp, L. J. (2020). MGM: Estimating time-varying mixed graphical models in high-dimensional data. *Journal of Statistical Software*, *93*(8). https://doi.org/10.18637/jss.v093.i08

Hautzinger, M., & Geue, K. (2016). Allgemeine Depressionsskala. In K. Geue, B. Strauß, & E. Brähler (Eds.), *Diagnostische Verfahren in der Psychotherapie* (pp. 33–35). Hogrefe. https://doi.org/10.1026/02700-00

Hayes, A. M., Laurenceau, J.-P., Feldman, G., Strauss, J. L., & Cardaciotto, L. (2007). Change is not always linear: The study of nonlinear and discontinuous patterns of change in psychotherapy. *Clinical Psychology Review*, *27*(6), 715–723. https://doi.org/10.1016/j.cpr.2007.01.008

Hayes, S. C., & Hofmann, S. G. (2018). *Process-based CBT: The science and core clinical competencies of cognitive behavioral therapy* (pp. vi, 452). New Harbinger Publications, Inc.

Hayes, S. C., Hofmann, S. G., Stanton, C. E., Carpenter, J. K., Sanford, B. T., Curtiss, J. E., & Ciarrochi, J. (2019). The role of the individual in the coming era of process-based therapy. *Behaviour Research and Therapy*, *117*, 40–53. https://doi.org/10.1016/j.brat.2018.10.005

Hebbrecht, K., Stuivenga, M., Birkenhäger, T., Morrens, M., Fried, E. I., Sabbe, B., & Giltay, E. J. (2020). Understanding personalized dynamics to inform precision medicine: A dynamic time warp analysis of 255 depressed inpatients. *BMC Medicine*, *18*(1), 400. https://doi.org/10.1186/s12916-020-01867-5

Heiligenstein, E., Guenther, G., Hsu, K., & Herman, K. (1996). Depression and Academic Impairment in College Students. *Journal of American College Health*, *45*(2), 59–64. https://doi.org/10.1080/07448481.1996.9936863

Hennig, C. (2015). What are the true clusters? *Pattern Recognition Letters*, *64*(15), 53–62. https://doi.org/10.1016/j.patrec.2015.04.009

REFERENCES

Henry, T. R., Robinaugh, D. J., & Fried, E. I. (2020). *On the control of psychological networks*. PsyArXiv. https://doi.org/10.31234/osf.io/7vpz2

Hewitt, B., & Turrell, G. (2011). Short-term Functional Health and Well-Being After Marital Separation: Does Initiator Status Make a Difference? *American Journal of Epidemiology*, *173*(11), 1308–1318. https://doi.org/10.1093/aje/kwr007

Hilbert, A., Herpertz, S., Zipfel, S., Tuschen-Caffier, B., Friederich, H.-C., Mayr, A., & de Zwaan, M. (2020). Psychopathological Networks in Cognitive-Behavioral Treatments for Binge-Eating Disorder. *Psychotherapy and Psychosomatics*, *89*(6), 379–385. https://doi.org/10.1159/000509458

Hittner, J. B., May, K., & Silver, N. C. (2003). A Monte Carlo Evaluation of Tests for Comparing Dependent Correlations. *The Journal of General Psychology*, *130*(2), 149–168. https://doi.org/10.1080/00221300309601282

Hoekstra, R. H. A., Epskamp, S., & Borsboom, D. (2022). Heterogeneity in Individual Network Analysis: Reality or Illusion? *Multivariate Behavioral Research*, 1–25. https://doi.org/10.1080/00273171.2022.2128020

Hoffart, A. (2014). The Need to Study Within-Person Processes in Psychotherapy: A Discussion of Alliance Studies. *SOJ Psychology*, *1*(2). https://symbiosisonlinepublishing.com/psychology/

Hoffart, A. (2018). The interplay between experiential states in first generation schema therapy of Cluster C personality problems: A network approach. *Clinical Psychology & Psychotherapy*, *25*(6), 809–817. https://doi.org/10.1002/cpp.2316

Hoffart, A., & Johnson, S. U. (2020). Within-person networks of clinical features of social anxiety disorder during cognitive and interpersonal therapy. *Journal of Anxiety Disorders*, *76*, 102312. https://doi.org/10.1016/j.janxdis.2020.102312

Hoffart, A., Johnson, S. U., & Ebrahimi, O. V. (2020). Loneliness and Social Distancing During the COVID-19 Pandemic: Risk Factors and Associations With Psychopathology. *Frontiers in Psychiatry*, *11*. https://www.frontiersin.org/articles/10.3389/fpsyt.2020.589127

Hoffart, A., Johnson, S. U., & Ebrahimi, O. V. (2022). Loneliness during the COVID-19 pandemic: Change and predictors of change from strict to discontinued social distancing protocols. *Anxiety, Stress, & Coping*, *35*(1), 44–57. https://doi.org/10.1080/10615806.2021.1958790

Hoffart, A., Langkaas, T. F., Øktedalen, T., & Johnson, S. U. (2019). The temporal dynamics of symptoms during exposure therapies of PTSD: A network approach. *European Journal of Psychotraumatology*, *10*(1), 1618134. https://doi.org/10.1080/20008198.2019.1618134

Hoffman, L., & Stawski, R. S. (2009). Persons as Contexts: Evaluating Between-Person and Within-Person Effects in Longitudinal Analysis. *Research in Human Development*, *6*(2–3), 97–120. https://doi.org/10.1080/15427600902911189

Hofmann, S. G., Asnaani, A., Vonk, I. J. J., Sawyer, A. T., & Fang, A. (2012). The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. *Cognitive Therapy and Research*, *36*(5), 427–440. https://doi.org/10.1007/s10608-012-9476-1

Hofmann, S. G., Curtiss, J. E., & Hayes, S. C. (2020). Beyond linear mediation: Toward a dynamic network approach to study treatment processes. *Clinical Psychology Review*, *76*, 101824. https://doi.org/10.1016/j.cpr.2020.101824

Hofmann, S. G., & Hayes, S. C. (2019). The Future of Intervention Science: Process-Based Therapy. *Clinical Psychological Science : A Journal of the Association for Psychological Science*, *7*(1), 37–50. https://doi.org/10.1177/2167702618772296

Holzhauer, C. G., Hildebrandt, T., Epstein, E., McCrady, B., Hallgren, K. A., & Cook, S. (2020). Mechanisms of change in female-specific and gender-neutral cognitive behavioral therapy for women with alcohol use disorder. *Journal of Consulting and Clinical Psychology*, *88*(6), 541–553. https://doi.org/10.1037/ccp0000492

Hoorelbeke, K., Van den Bergh, N., Wichers, M., & Koster, E. H. W. (2019). Between vulnerability and resilience: A network analysis of fluctuations in cognitive risk and protective factors following remission from depression. *Behaviour Research and Therapy*, *116*, 1–9. https://doi.org/10.1016/j.brat.2019.01.007

Hoyle, R. H. (2012). *Handbook of structural equation modeling*. Guilford press.

Hoyle, R. H., & Isherwood, J. C. (2013). Reporting results from structural equation modeling analyses in Archives of Scientific Psychology. *Archives of Scientific Psychology*, *1*(1), 14–22. https://doi.org/10.1037/arc0000004

Huth, K., Ron, J. de, Luigjes, J., Goudriaan, A., Mohammadi, R., Holst, R. van, Wagenmakers, E.-J., & Marsman, M. (2023). *Bayesian Analysis of Cross-sectional Networks: A Tutorial in R and JASP*. PsyArXiv. https://doi.org/10.31234/osf.io/ub5tc

Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift Für Physik*, *31*, 253–258. https://doi.org/10.1007/BF02980577

Isvoranu, A.-M., & Epskamp, S. (2021). Which Estimation Method to Choose in Network Psychometrics? Deriving Guidelines for Applied Researchers. *Psychological Methods*.

Isvoranu, A.-M., Epskamp, S., Waldorp, L. J., & Borsboom, D. (Eds.). (2022). *Network Psychometrics with R: A Guide for Behavioral and Social Scientists*. Routledge, Taylor & Francis Group.

Isvoranu, A.-M., Guloksuz, S., Epskamp, S., van Os, J., Borsboom, D., Investigators, G., & others. (2020). Toward incorporating genetic risk scores into symptom networks of psychosis. *Psychological Medicine*, *50*(4), 636–643.

Isvoranu, A.-M., Ziermans, T., Schirmbeck, F., Borsboom, D., Geurts, H. M., de Haan, L., van Amelsvoort, T., Bartels-Velthuis, A. A., Simons, C. J. P., van Os, J., & others. (2021). Autistic Symptoms and Social Functioning in Psychosis: A Network Approach. *Schizophrenia Bulletin*.

Johnson, S. U., & Hoffart, A. (2018). Metacognitive Therapy Versus Cognitive Behavioral Therapy:A Network Approach. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02382

Jones, A., Remmerswaal, D., Verveer, I., Robinson, E., Franken, I. H. A., Wen, C. K. F., & Field, M. (2019). Compliance with ecological momentary assessment protocols in substance users: A meta-analysis. *Addiction (Abingdon, England)*, *114*(4), 609–619. https://doi.org/10.1111/add.14503

Jones, P. (2018). *Tools for Identifying Important Nodes in Networks* [Computer software].

Jones, P. J. (2017). *Networktools: Assorted Tools for Identifying Important Nodes in Networks. R package version 1.0.0. Https://CRAN.R-project.org/package=networktools*.

Jones, P. J., Ma, R., & McNally, R. J. (2019). Bridge centrality: A network approach to understanding comorbidity. *Multivariate Behavioral Research*, 1–15. https://doi.org/10.1080/00273171.2019.1614898

Jones, P. J., Ma, R., & McNally, R. J. (2021). Bridge Centrality: A Network Approach to Understanding Comorbidity. *Multivariate Behavioral Research*, *56*(2), 353–367. https://doi.org/10.1080/00273171.2019.1614898

Jones, P. J., Mair, P., & McNally, R. J. (2018). Visualizing psychological networks: A tutorial in R. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01742

Jongerling, J., Epskamp, S., & Williams, D. R. (2022). Bayesian Uncertainty Estimation for Gaussian Graphical Models and Centrality Indices. *Multivariate Behavioral Research*, *0*(0), 1–29. https://doi.org/10.1080/00273171.2021.1978054

Joormann, J., & Gotlib, I. H. (2010). Emotion regulation in depression: Relation to cognitive inhibition. *Cognition and Emotion*, *24*(2), 281–298. https://doi.org/10.1080/02699930903407948

Jordan, D. G., Winer, E. S., & Salem, T. (2020). The current status of temporal network analysis for clinical science: Considerations as the paradigm shifts? *Journal of Clinical Psychology*, *76*(9), 1591–1612. https://doi.org/10.1002/jclp.22957

Judd, L. L., Kessler, R. C., Paulus, M. P., Zeller, P. V., Wittchen, H. U., & Kunovac, J. L. (1998). Comorbidity as a fundamental feature of generalized anxiety disorders: Results from the National Comorbidity Study (NCS). *Acta Psychiatrica Scandinavica. Supplementum*, *393*, 6–11. https://doi.org/10.1111/j.1600-0447.1998.tb05960.x

Kaiser, T., Boschloo, L., Berger, T., Meyer, B., Späth-Nellissen, C., Schröder, J., Hohagen, F., Moritz, S., & Klein, J. P. (2021). Maintaining Outcomes of Internet-Delivered Cognitive-Behavioral Therapy for Depression: A Network Analysis of Follow-Up Effects. *Frontiers in Psychiatry*, *12*, 598317. https://doi.org/10.3389/fpsyt.2021.598317

Kaiser, T., Herzog, P., Voderholzer, U., & Brakemeier, E.-L. (2021). Unraveling the comorbidity of depression and anxiety in a large inpatient sample: Network analysis to examine bridge symptoms. *Depression and Anxiety*, *38*(3), 307–317. https://doi.org/10.1002/da.23136

Kaiser, T., & Laireiter, A.-R. (2018). Process-symptom-bridges in psychotherapy: An idiographic network approach. *Journal for Person-Oriented Research*, *4*(2), Article 2. https://doi.org/10.17505/jpor.2018.06

Keser, E., Kahya, Y., & Akın, B. (2020). Stress generation hypothesis of depressive symptoms in interpersonal stressful life events: The roles of cognitive triad and coping styles via structural equation modeling. *Current Psychology*, *39*(1), 174–182. https://doi.org/10.1007/s12144-017-9744-z

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 593. https://doi.org/10.1001/archpsyc.62.6.593

Kievit, R., Frankenhuis, W., Waldorp, L., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00513

Kim, J., Zhu, W., Chang, L., Bentler, P. M., & Ernst, T. (2007). Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. *Human Brain Mapping*, *28*(2), 85–93. https://doi.org/10.1002/hbm.20259

Klintwall, L., Bellander, M., & Cervin, M. (2023). Perceived Causal Problem Networks: Reliability, Central Problems, and Clinical Utility for Depression. *Assessment*, *30*(1), 73–83. https://doi.org/10.1177/10731911211039281

Klipstein, L. von, Servaas, M., Lamers, F., Schoevers, R., Wardenaar, K., & Riese, H. (2022). *Can Floor Effects Explain Increased Affective Reactivity Among Depressed Individuals Found in Experience Sampling Research?* PsyArXiv. https://doi.org/10.31234/osf.io/6vtew

Kolde. (2019). *Pheatmap*. R package version 1.0.12.

Komulainen, K., Airaksinen, J., Savelieva, K., Gluschkoff, K., García Velázquez, R., Elovainio, M., & Jokela, M. (2021). Network dynamics of depressive symptoms in antidepressant medication treatment: Secondary analysis of eight clinical trials. *Molecular Psychiatry*, *26*, 3328–3335. https://doi.org/10.1038/s41380-020-00884-3

Körppen, D., Ropers, N., & Giessmann, H. J. (Eds.). (2011). *The Non-Linearity of Peace Processes: Theory and Practice of Systemic Conflict Transformation* (1st ed.). Verlag Barbara Budrich. https://doi.org/10.2307/j.ctvbkjz66

Kraft, B., Jonassen, R., Heeren, A., Harmer, C., Stiles, T., & Landrø, N. I. (2019). Attention bias modification in remitted depression is associated with increased interest and leads to reduced adverse impact of anxiety symptoms and negative cognition. *Clinical Psychological Science*, *7*, 530–544. https://doi.org/10.1177/2167702618822480

Kravitz, R., & Duan, N. (2014). *Design and Implementation of N-of-1 Trials: A User's Guide*.

Kreiter, D., Drukker, M., Mujagic, Z., Vork, L., Rutten, B. P. F., van Os, J., Masclee, A. A. M., Kruimel, J. W., & Leue, C. (2021). Symptom-network dynamics in irritable bowel syndrome with comorbid panic disorder using electronic momentary assessment: A randomized controlled trial of escitalopram vs. placebo. *Journal of Psychosomatic Research*, *141*, 110351. https://doi.org/10.1016/j.jpsychores.2020.110351

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kushner, M. G., Abrams, K., & Borchardt, C. (2000). The relationship between anxiety disorders and alcohol use disorders: A review of major perspectives and findings. *Clinical Psychology Review*, *20*(2), 149–171. https://doi.org/10.1016/S0272-7358(99)00027-6

Kuyken, W., Padesky, C. A., & Dudley, R. (2009). *Collaborative case conceptualization: Working effectively with clients in cognitive-behavioral therapy* (pp. xvii, 366). Guilford Press.

Lafit, G., Adolf, J. K., Dejonckheere, E., Myin-Germeys, I., Viechtbauer, W., & Ceulemans, E. (2021). Selection of the Number of Participants in Intensive Longitudinal Studies: A User-Friendly Shiny App and Tutorial for Performing Power Analysis in Multilevel Regression Models That Account for Temporal Dependencies. *Advances in Methods and Practices in Psychological Science*, *4*(1), 251524592097873. https://doi.org/10.1177/2515245920978738

Lancee, J., Harvey, A. G., Morin, C. M., Ivers, H., van der Zweerde, T., & Blanken, T. F. (2022). Network Intervention Analyses of cognitive therapy and behavior therapy for insomnia: Symptom specific effects and process measures. *Behaviour Research and Therapy*, *153*, 104100. https://doi.org/10.1016/j.brat.2022.104100

Lane, S. T., Gates, K. M., Pike, H. K., Beltz, A. M., & Wright, A. G. C. (2019). Uncovering general, shared, and unique temporal patterns in ambulatory assessment data. *Psychological Methods*, *24*(1), 54–69. https://doi.org/10.1037/met0000192

Laposa, J. M., & Fracalanza, K. (2019). Does intolerance of uncertainty mediate improvement in anger during group CBT for GAD? A preliminary investigation. *Behavioural and Cognitive Psychotherapy*, *47*(5), 585–593. https://doi.org/10.1017/S1352465819000249

Lauritzen, S. L. (1996). *Graphical models* (Vol. 17). Clarendon Press.

Lee, S. A. (2020). How much "Thinking" about COVID-19 is clinically dysfunctional? *Brain, Behavior, and Immunity*, *87*, 97–98. https://doi.org/10.1016/j.bbi.2020.04.067

Leng, T., White, C., Hilton, J., Kucharski, A., Pellis, L., Stage, H., Davies, N. G., Centre for Mathematical Modelling of Infectious Disease 2019 nCoV Working Group, Keeling, M. J., & Flasche, S. (2020). The effectiveness of social bubbles as part of a Covid-19 lockdown exit strategy, a modelling study. *Wellcome Open Research*, *5*, 213. https://doi.org/10.12688/wellcomeopenres.16164.2

Levenson, J. C., Shensa, A., Sidani, J. E., Colditz, J. B., & Primack, B. A. (2016). The association between social media use and sleep disturbance among young adults. *Preventive Medicine*, *85*, 36–41. https://doi.org/10.1016/j.ypmed.2016.01.001

Levine, S. Z., & Leucht, S. (2016). Identifying a system of predominant negative symptoms: Network analysis of three randomized clinical trials. *Schizophrenia Research*, *178*(1–3), 17–22. https://doi.org/10.1016/j.schres.2016.09.002

Levinson, C. A., Brosof, L. C., Vanzhula, I., Christian, C., Jones, P., Rodebaugh, T. L., Langer, J. K., White, E. K., Warren, C., Weeks, J. W., & others. (2018). Social anxiety and eating disorder comorbidity and underlying vulnerabilities: Using network analysis to conceptualize comorbidity. *International Journal of Eating Disorders*, *51*(7), 693–709.

Levinson, C. A., Cash, E., Welch, K., Epskamp, S., Hunt, R. A., Williams, B. M., Keshishian, A. C., & Spoor, S. P. (2020). Personalized networks of eating disorder symptoms predicting eating disorder outcomes and remission. *International Journal of Eating Disorders*, *53*(12), 2086–2094.

Levinson, C. A., Christian, C., Ram, S. S., Vanzhula, I., Brosof, L. C., Michelson, L. P., & Williams, B. M. (2020). Eating disorder symptoms and core eating disorder fears decrease during online imaginal exposure therapy for eating disorders. *Journal of Affective Disorders*, *276*, 585–591. https://doi.org/10.1016/j.jad.2020.07.075

Levinson, C. A., Cusack, C., Brown, & Smith. (in press). Conceptualizing Eating Disorders Through A Network Lens. *Nature Psychology Reviews*.

Levinson, C. A., Hunt, R. A., Christian, C., Williams, B. M., Keshishian, A. C., Vanzhula, I. A., & Ralph-Nearman, C. (2022). Longitudinal group and individual networks of eating disorder symptoms in individuals diagnosed with an eating disorder. *Journal of Psychopathology and Clinical Science*, *131*(1), 58–72. https://doi.org/10.1037/abn0000727

Levinson, C. A., Hunt, R. A., Keshishian, A. C., Brown, M. L., Vanzhula, I., Christian, C., Brosof, L. C., & Williams, B. M. (2021). Using Individual Networks to Identify Treatment Targets for Eating Disorder Treatment: A Proof-of-Concept Study and Initial Data. *Under Review*.

Levinson, C. A., Vanzhula, I. A., Smith, T. W., & Stice, E. (2020). Group and longitudinal intra-individual networks of eating disorder symptoms in adolescents and young adults at-risk for an eating disorder. *Behaviour Research and Therapy*, *135*, 103731. https://doi.org/10.1016/j.brat.2020.103731

Levinson, C. A., Vanzhula, I., & Brosof, L. C. (2018). Longitudinal and personalized networks of eating disorder cognitions and behaviors: Targets for precision intervention a proof of concept study. *International Journal of Eating Disorders*, *51*(11), 1233–1243. https://doi.org/10.1002/eat.22952

Levinson, C. A., & Williams, B. M. (2020). Eating disorder fear networks: Identification of central eating disorder fears. *International Journal of Eating Disorders*, *53*(12), 1960–1973. https://doi.org/10.1002/eat.23382

Levinson, C. A., Williams, B. M., Christian, C., Hunt, R. A., Keshishian, A. C., Brosof, L. C., Vanzhula, I. A., Davis, G. G., Brown, M. L., Bridges-Curry, Z., Sandoval-Araujo, L. E., & Ralph-Nearman, C. (2023). Personalizing eating disorder treatment using idiographic models: An open series trial. *Journal of Consulting and Clinical Psychology*, *91*(1), 14–28. https://doi.org/10.1037/ccp0000785

Levinson, C. A., Zerwas, S., Calebs, B., Forbush, K., Kordy, H., Watson, H., Hofmeier, S., Levine, M., Crosby, R. D., Peat, C., Runfola, C. D., Zimmer, B., Moesner, M., Marcus, M. D., & Bulik, C. M. (2017). The core symptoms of bulimia nervosa, anxiety, and depression: A network analysis. *Journal of Abnormal Psychology*, *126*(3), 340–354. https://doi.org/10.1037/abn0000254

Liebovitch, L. S., Peluso, P. R., Norman, M. D., Su, J., & Gottman, J. M. (2011). Mathematical model of the dynamics of psychotherapy. *Cognitive Neurodynamics*, *5*(3), 265–275. https://doi.org/10.1007/s11571-011-9157-x

Loerinc, A. G., Meuret, A. E., Twohig, M. P., Rosenfield, D., Bluett, E. J., & Craske, M. G. (2015). Response rates for CBT for anxiety disorders: Need for standardized criteria. *Clinical Psychology Review*, *42*, 72–82. https://doi.org/10.1016/j.cpr.2015.08.004

Logan, E. L., Thornton, J. A., & Breen, L. J. (2018). What determines supportive behaviors following bereavement? A systematic review and call to action. *Death Studies*, *42*(2), 104–114. https://doi.org/10.1080/07481187.2017.1329760

Lorimer, B., Delgadillo, J., Kellett, S., & Brown, G. (2020). Exploring relapse through a network analysis of residual depression and anxiety symptoms after cognitive behavioural therapy: A proof-of-concept study. *Psychotherapy Research: Journal of the Society for Psychotherapy Research*, *30*(5), 650–661. https://doi.org/10.1080/10503307.2019.1650980

Lunansky, G., Naberman, J., van Borkulo, C. D., Chen, C., Wang, L., & Borsboom, D. (2022). Intervening on psychopathology networks: Evaluating intervention targets through simulations. *Methods*, *204*, 29–37. https://doi.org/10.1016/j.ymeth.2021.11.006

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis* [Springer Books]. Springer. https://econpapers.repec.org/bookchap/sprsprbok/978-3-540-27752-1.htm

Lutz, W., Schwartz, B., Hofmann, S. G., Fisher, A. J., Husen, K., & Rubel, J. A. (2018). Using network analysis for the prediction of treatment dropout in patients with mood and anxiety disorders: A methodological proof-of-concept study. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-25953-0

Lydon-Staley, D. M., Leventhal, A. M., Piper, M. E., Schnoll, R. A., & Bassett, D. S. (2021). Temporal networks of tobacco withdrawal symptoms during smoking cessation treatment. *Journal of Abnormal Psychology*, *130*, 89–101. https://doi.org/10.1037/abn0000650

Lydon-Staley, D. M., Schnoll, R. A., Hitsman, B., & Bassett, D. S. (2020). The network structure of tobacco withdrawal in a community sample of smokers treated with nicotine patch and behavioral counseling. *Nicotine & Tobacco Research*, *22*, 408–414. https://doi.org/10.1093/ntr/nty250

Madhoo, M., & Levine, S. Z. (2016). Network analysis of the Quick Inventory of Depressive Symptomatology: Reanalysis of the STAR*D clinical trial. *European Neuropsychopharmacology*, *26*(11), 1768–1774. https://doi.org/10.1016/j.euroneuro.2016.09.368

Malgaroli, M., Maccallum, F., & Bonanno, G. A. (2018). Symptoms of persistent complex bereavement disorder, depression, and PTSD in a conjugally bereaved sample: A network analysis. *Psychological Medicine*, *48*(14), 2439–2448. https://doi.org/10.1017/S0033291718001769

Manea, L., Gilbody, S., & McMillan, D. (2012). Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): A meta-analysis. *CMAJ: Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne*, *184*(3), E191-196. https://doi.org/10.1503/cmaj.110829

Mansueto, A. C., Wiers, R. W., van Weert, J. C. M., Schouten, B. C., & Epskamp, S. (2022). Investigating the feasibility of idiographic network models. *Psychological Methods*. https://doi.org/10.1037/met0000466

Mariotti, E. C., Bolden, J., & Finn, M. T. M. (2021). Unifying treatment for mild anxiety and depression in preadolescence. *Psychotherapy (Chicago, Ill.)*, *58*(1), 121–132. https://doi.org/10.1037/pst0000350

Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L. J., Maas, H. L. J. van der, & Maris, G. (2018). An Introduction to Network Psychometrics: Relating Ising Network Models to Item Response Theory Models. *Multivariate Behavioral Research*, *53*(1), 15–35. https://doi.org/10.1080/00273171.2017.1379379

Marsman, M., Maris, G., Bechger, T., & Glas, C. (2015). Bayesian inference for low-rank Ising networks. *Scientific Reports*, *5*(1), 1–7.

Martin, L. L., & Tesser, A. (1989). Toward a motivational and structural theory of ruminative thought. In *Unintended thought* (pp. 306–326). The Guilford Press.

Maser, J. D., & Cloninger, C. R. (1990). *Comorbidity of mood and anxiety disorders* (pp. xvii, 869). American Psychiatric Association.

Mathisen, T. F., Bratland-Sanda, S., Rosenvinge, J. H., Friborg, O., Pettersen, G., Vrabel, K. A., & Sundgot-Borgen, J. (2018). Treatment effects on compulsive exercise and physical activity in eating disorders. *Journal of Eating Disorders*, *6*, 43. https://doi.org/10.1186/s40337-018-0215-1

McElroy, E., Napoleone, E., Wolpert, M., & Patalay, P. (2019). Structure and Connectivity of Depressive Symptom Networks Corresponding to Early Treatment Response. *EClinicalMedicine*, *8*, 29–36. https://doi.org/10.1016/j.eclinm.2019.02.009

McLean, D. C., Nakamura, J., & Csikszentmihalyi, M. (2017). Explaining system missing: Missing data and experience sampling method. *Social Psychological and Personality Science*, *8*, 434–441. https://doi.org/10.1177/1948550617708015

McNally, R. J. (2016). Can network analysis transform psychopathology? *Behaviour Research and Therapy*, *86*, 95–104. https://doi.org/10.1016/j.brat.2016.06.006

McNeish, D., & Hamaker, E. L. (2020). A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. *Psychological Methods*, *25*(5), 610–635. https://doi.org/10.1037/met0000250

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence* (pp. x, 149). University of Minnesota Press. https://doi.org/10.1037/11281-000

Meehl, P. E. (1990). Appraising and Amending Theories: The Strategy of Lakatosian Defense and Two Principles that Warrant It. *Psychological Inquiry*, *1*(2), 108–141. https://doi.org/10.1207/s15327965pli0102_1

Meinshausen, N., Bühlmann, P., & others. (2006). High-dimensional graphs and variable selection with the lasso. *Annals of Statistics*, *34*(3), 1436–1462. https://doi.org/10.1214/009053606000000281

Mercier, H., & Sperber, D. (2018). *The Enigma of Reason: A new theory of human understanding*. Penguin Books.

Mestdagh, M., & Dejonckheere, E. (2021). Ambulatory assessment in psychopathology research: Current achievements and future ambitions. *Current Opinion in Psychology*, *41*, 1–8. https://doi.org/10.1016/j.copsyc.2021.01.004

Meyer, T. D., & Hautzinger, M. (2001). Allgemeine Depressions-Skala (ADS). *Diagnostica*, *47*(4), 208–215. https://doi.org/10.1026//0012-1924.47.4.208

Miller, W. R., & Seligman, M. E. (1975). Depression and learned helplessness in man. *Journal of Abnormal Psychology*, *84*, 228–238. https://doi.org/10.1037/h0076720

Molenaar, P. C. M. (2004). A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Measurement: Interdisciplinary Research and Perspectives*, *2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Molenaar, P. C. M., & Campbell, C. G. (2009). The New Person-Specific Paradigm in Psychology. *Current Directions in Psychological Science*, *18*(2), 112–117. https://doi.org/10.1111/j.1467-8721.2009.01619.x

Monteleone, A. M., Cardi, V., Ambwani, S., Cascino, G., Albano, G., Pellegrino, F., & Treasure, J. (2021). Network intervention analysis to assess the trajectory of change and treatment effects associated with the use of online guided self-help for anorexia nervosa. *Early Intervention in Psychiatry*, *15*(5), 1210–1216. https://doi.org/10.1111/eip.13064

Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time Series Missing Value Imputation in R. *The R Journal*, *9*(1), 207. https://doi.org/10.32614/RJ-2017-009

Morris, L., & Mansell, W. (2018). A systematic review of the relationship between rigidity/flexibility and transdiagnostic cognitive and behavioral processes that maintain psychopathology. *Journal of Experimental Psychopathology*, *9*(3), 2043808718779431. https://doi.org/10.1177/2043808718779431

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*(11), 2074–2102. https://doi.org/10.1002/sim.8086

Mullarkey, M. C., Stein, A. T., Pearson, R., & Beevers, C. G. (2020). Network analyses reveal which symptoms improve (or not) following an Internet intervention (Deprexis) for depression. *Depression and Anxiety*, *37*(2), 115–124. https://doi.org/10.1002/da.22972

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry*, *17*(2), 123–132. https://doi.org/10.1002/wps.20513

Nature Medicine. (2020). Keep mental health in mind. *Nature Medicine*, *26*(5), Article 5. https://doi.org/10.1038/s41591-020-0914-4

Newman, M. (2018). *Networks*. Oxford university press.

Nolte, T., Guiney, J., Fonagy, P., Mayes, L., & Luyten, P. (2011). Interpersonal Stress Regulation and the Development of Anxiety Disorders: An Attachment-Based Developmental Framework. *Frontiers in Behavioral Neuroscience*, *5*. https://www.frontiersin.org/articles/10.3389/fnbeh.2011.00055

Normann, N., & Morina, N. (2018). The Efficacy of Metacognitive Therapy: A Systematic Review and Meta-Analysis. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02211

O'Driscoll, C., Buckman, J. E. J., Fried, E. I., Saunders, R., Cohen, Z. D., Ambler, G., DeRubeis, R. J., Gilbody, S., Hollon, S. D., Kendrick, T., Kessler, D., Lewis, G., Watkins, E., Wiles, N., & Pilling, S. (2021). The importance of transdiagnostic symptom level assessment to understanding prognosis for depressed adults: Analysis of data from six randomised control trials. *BMC Medicine*, *19*(1), 109. https://doi.org/10.1186/s12916-021-01971-0

O'Hagan, A. (2006). *Uncertain judgements: Eliciting experts' probabilities*. Wiley. https://doi.org/10.1002/0470033312

O'Hagan, A. (2019). Expert Knowledge Elicitation: Subjective but Scientific. *The American Statistician*, *73*(sup1), 69–81. https://doi.org/10.1080/00031305.2018.1518265

Olatunji, B. O., Levinson, C., & Calebs, B. (2018). A network analysis of eating disorder symptoms and characteristics in an inpatient sample. *Psychiatry Research*, *262*, 270–281. https://doi.org/10.1016/j.psychres.2018.02.027

Olthof, M., Hasselman, F., & Lichtwarck-Aschoff, A. (2020). Complexity in psychological self-ratings: Implications for research and practice. *BMC Medicine*, *18*(1), 317. https://doi.org/10.1186/s12916-020-01727-2

Ono, M., Schneider, S., Junghaenel, D. U., & Stone, A. A. (2019). What Affects the Completion of Ecological Momentary Assessments in Chronic Pain Research? An Individual Patient Data Meta-Analysis. *Journal of Medical Internet Research*, *21*(2), e11398. https://doi.org/10.2196/11398

Ooi, H., de Vries, A., & Microsoft. (2020). *R-package checkpoint*.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251. https://doi.org/10.1016/j.socnet.2010.03.006

Pachankis, J. E. (2015). A Transdiagnostic Minority Stress Treatment Approach for Gay and Bisexual Men's Syndemic Health Conditions. *Archives of Sexual Behavior*, *44*(7), 1843–1860. https://doi.org/10.1007/s10508-015-0480-x

Page, A. C., & Stritzke, W. G. K. (2014). *Clinical Psychology for Trainees: Foundations of Science-Informed Practice* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139857109

Page, A. C., Stritzke, W. G. K., & Mclean, N. J. (2008). Toward science-informed supervision of clinical case formulation: A training model and supervision method. *Australian Psychologist*, *43*(2), 88–95. https://doi.org/10.1080/00050060801994156

Palgi, Y., Shrira, A., Ring, L., Bodner, E., Avidor, S., Bergman, Y., Cohen-Fridel, S., Keisari, S., & Hoffman, Y. (2020). The loneliness pandemic: Loneliness and other concomitants of depression, anxiety and their comorbidity during the COVID-19 outbreak. *Journal of Affective Disorders*, *275*, 109–111. https://doi.org/10.1016/j.jad.2020.06.036

Pallesen, S., Bjorvatn, B., Nordhus, I. H., Sivertsen, B., Hjørnevik, M., & Morin, C. M. (2008). A New Scale for Measuring Insomnia: The Bergen Insomnia Scale. *Perceptual and Motor Skills*, *107*(3), 691–706. https://doi.org/10.2466/pms.107.3.691-706

Papa, A., Rummel, C., Garrison-Diehn, C., & Sewell, M. T. (2013). Behavioral Activation for Pathological Grief. *Death Studies*, *37*(10), 913–936. https://doi.org/10.1080/07481187.2012.692459

Papini, S., Rubin, M., Telch, M. J., Smits, J. A. J., & Hien, D. A. (2020). Pretreatment Posttraumatic Stress Disorder Symptom Network Metrics Predict the Strength of the Association Between Node Change and Network Change During Treatment. *Journal of Traumatic Stress*, *33*(1), 64–71. https://doi.org/10.1002/jts.22379

REFERENCES

Park, S.-C., Kim, Y., Kim, K., Woo, Y. S., Kim, J.-B., Jang, E. Y., Lee, H.-Y., Yim, H.-W., Ham, B.-J., Kim, J.-M., & Park, Y. C. (2021). Network Analysis of the Symptoms of Depressive Disorders Over the Course of Therapy: Changes in Centrality Measures. *Psychiatry Investigation*, *18*(1), 48–58. https://doi.org/10.30773/pi.2020.0367

Pashler, H., & Wagenmakers, E. (2012). Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? *Perspectives on Psychological Science*, *7*(6), 528–530. https://doi.org/10.1177/1745691612465253

Pe, M. L., Kircanski, K., Thompson, R. J., Bringmann, L. F., Tuerlinckx, F., Mestdagh, M., Mata, J., Jaeggi, S. M., Buschkuehl, M., Jonides, J., Kuppens, P., & Gotlib, I. H. (2015). Emotion-Network Density in Major Depressive Disorder. *Clinical Psychological Science*, *3*(2), 292–300. https://doi.org/10.1177/2167702614540645

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*.

Persons, J. B. (2006). Case formulation–driven psychotherapy. *Clinical Psychology: Science and Practice*, *13*(2), 167–170.

Persons, J. B. (2012). *The case formulation approach to cognitive-behavior therapy*. Guilford Press.

Persons, J. B., & Talbot, L. S. (2019). An idiographic hypothesis-testing approach to psychotherapy: Using case formulation and progress monitoring to guide treatment. In S. Dimidjian (Ed.), *Evidence-based practice in action: Bridging clinical science and intervention* (pp. 187–200). The Guilford Press; US.

Piccirillo, M. L., Beck, E. D., & Rodebaugh, T. L. (2019). A Clinician's Primer for Idiographic Research: Considerations and Recommendations. *Behavior Therapy*, *50*(5), 938–951. https://doi.org/10.1016/j.beth.2019.02.002

Piccirillo, M. L., & Rodebaugh, T. L. (2019). Foundations of idiographic methods in psychology and applications for psychotherapy. *Clinical Psychology Review*, 90–100. https://doi.org/10.1016/j.cpr.2019.01.002

Pinheiro. (2022). *Nmle* [Computer software].

Pisetsky, E. M., Schaefer, L. M., Wonderlich, S. A., & Peterson, C. B. (2019). Emerging Psychological Treatments in Eating Disorders. *The Psychiatric Clinics of North America*, *42*(2), 219–229. https://doi.org/10.1016/j.psc.2019.01.005

Plummer, M. (2003). *JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling*.

Prigerson, H. G., Horowitz, M. J., Jacobs, S. C., Parkes, C. M., Aslan, M., Goodkin, K., Raphael, B., Marwit, S. J., Wortman, C., Neimeyer, R. A., Bonanno, G., Block, S. D., Kissane, D., Boelen, P., Maercker, A., Litz, B. T., Johnson, J. G., First, M. B., & Maciejewski, P. K. (2009). Prolonged Grief Disorder: Psychometric Validation of Criteria Proposed for DSM-V and ICD-11. *PLoS Medicine*, *6*(8), e1000121. https://doi.org/10.1371/journal.pmed.1000121

Primack, B. A., Shensa, A., Sidani, J. E., Escobar-Viera, C. G., & Fine, M. J. (2021). Temporal Associations Between Social Media Use and Depression. *American Journal of Preventive Medicine*, *60*(2), 179–188. https://doi.org/10.1016/j.amepre.2020.09.014

Proctor, E. K., Landsverk, J., Aarons, G., Chambers, D., Glisson, C., & Mittman, B. (2009). Implementation research in mental health services: An emerging science with conceptual, methodological, and training challenges. *Administration and Policy in Mental Health*, *36*(1), 24–34. https://doi.org/10.1007/s10488-008-0197-4

R Core Team. (2015). R Development Core Team. *R: A Language and Environment for Statistical Computing*, *55*, 275–286.

Radloff, L. S. (1977). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, *1*(3), 385–401. https://doi.org/10.1177/014662167700100306

Ravikumar, P., Wainwright, M. J., & Lafferty, J. D. (2010). High-dimensional ising model selection using ℓ1-regularized logistic regression. *Annals of Statistics*, *38*(3), 1287–1319. https://doi.org/10.1214/09-AOS691

Resnick, L. B., Salmon, M., Zeitz, C. M., Wathen, S. H., & Holowchak, M. (1993). Reasoning in Conversation. *Cognition and Instruction*, *11*(3–4), 347–364. https://doi.org/10.1080/07370008.1993.9649029

Riesco, N., Agüera, Z., Granero, R., Jiménez-Murcia, S., Menchón, J. M., & Fernández-Aranda, F. (2018). Other Specified Feeding or Eating Disorders (OSFED): Clinical heterogeneity and cognitive-behavioral therapy outcome. *European Psychiatry: The Journal of the Association of European Psychiatrists*, *54*, 109–116. https://doi.org/10.1016/j.eurpsy.2018.08.001

Riese, H., von Klipstein, L., Schoevers, R. A., van der Veen, D. C., & Servaas, M. N. (2021). Personalized ESM monitoring and feedback to support psychological treatment for depression: A pragmatic randomized controlled trial (Therap-i). *BMC Psychiatry*, *21*(1), 143. https://doi.org/10.1186/s12888-021-03123-3

Rintala, A., Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2019). Response compliance and predictors thereof in studies using the experience sampling method. *Psychological Assessment*, *31*(2), 226–235. https://doi.org/10.1037/pas0000662

Robinaugh, D., Haslbeck, J., Waldorp, L., Kossakowski, J., Fried, E. I., Millner, A., McNally, R. J., Ryan, O., Ron, J. de, Maas, H. van der, Nes, E. H. van, Scheffer, M., Kendler, K. S., & Borsboom, D. (2019). *Advancing the Network Theory of Mental Disorders: A Computational Model of Panic Disorder*. PsyArXiv. https://doi.org/10.31234/osf.io/km37w

Robinaugh, D. J., Haslbeck, J. M. B., Ryan, O., Fried, E. I., & Waldorp, L. J. (2021). Invisible Hands and Fine Calipers: A Call to Use Formal Theory as a Toolkit for Theory Construction. *Perspectives on Psychological Science*, 1745691620974697. https://doi.org/10.1177/1745691620974697

Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R., & Borsboom, D. (2020). The network approach to psychopathology: A review of the literature 2008-2018 and an agenda for future research. *Psychological Medicine*, *50*(3), 353–366. https://doi.org/10.1017/S0033291719003404

Robinaugh, D. J., LeBlanc, N. J., Vuletich, H. A., & McNally, R. J. (2014). Network analysis of persistent complex bereavement disorder in conjugally bereaved adults. *Journal of Abnormal Psychology*, *123*, 510–522. https://doi.org/10.1037/abn0000002

Robinaugh, D. J., Millner, A. J., & McNally, R. J. (2016). Identifying highly influential nodes in the complicated grief network. *Journal of Abnormal Psychology*, *125*, 747–757. https://doi.org/10.1037/abn0000181

Robinson, W. (1950). Ecological correlations and the behavior of individuals. *Am Sociol Rev*, *15*. https://doi.org/10.2307/2087176

Rodebaugh, T. L., Frumkin, M. R., & Piccirillo, M. L. (2020). The long road from person-specific models to personalized mental health treatment. *BMC Medicine*, *18*(1), 365. https://doi.org/10.1186/s12916-020-01838-w

Rodebaugh, T. L., Tonge, N. A., Piccirillo, M. L., Fried, E., Horenstein, A., Morrison, A. S., Goldin, P., Gross, J. J., Lim, M. H., Fernandez, K. C., Blanco, C., Schneier, F. R., Bogdan, R., Thompson, R. J., & Heimberg, R. G. (2018). Does centrality in a cross-sectional network suggest intervention targets for social anxiety disorder? *Journal of Consulting and Clinical Psychology*, *86*(10), 831–844. https://doi.org/10.1037/ccp0000336

Rothman, A. J., Levina, E., & Zhu, J. (2010). Sparse Multivariate Regression With Covariance Estimation. *Journal of Computational and Graphical Statistics : A Joint Publication of American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of North America*, *19*(4), 947–962. https://doi.org/10.1198/jcgs.2010.09188

Rubel, J. A., Fisher, A. J., Husen, K., & Lutz, W. (2018). Translating Person-Specific Network Models into Personalized Treatments: Development and Demonstration of the Dynamic Assessment Treatment Algorithm for Individual Networks (DATA-IN). *Psychotherapy and Psychosomatics*, *87*(4), 249–252.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

REFERENCES

Ryan, O., Bringmann, L., & Schuurman, N. K. (2019). *The Challenge of Generating Causal Hypotheses Using Network Models*. PsyArXiv. https://doi.org/10.31234/osf.io/ryg69

Ryan, O., & Hamaker, E. (2020). *Time to Intervene: A Continuous-Time Approach to Network Analysis and Centrality*. PsyArXiv. https://doi.org/10.31234/osf.io/2ambn

Ryan, O., & Hamaker, E. L. (2022). Time to Intervene: A Continuous-Time Approach to Network Analysis and Centrality. *Psychometrika*, *87*(1), 214–252. https://doi.org/10.1007/s11336-021-09767-0

Ryan, O., Kuiper, R. M., & Hamaker, E. L. (2018). A Continuous-Time Approach to Intensive Longitudinal Data: What, Why, and How? In K. van Montfort, J. H. L. Oud, & M. C. Voelkle (Eds.), *Continuous Time Modeling in the Behavioral and Related Sciences* (pp. 27–54). Springer International Publishing. https://doi.org/10.1007/978-3-319-77219-6_2

Salari, N., Hosseinian-Far, A., Jalali, R., Vaisi-Raygani, A., Rasoulpoor, S., Mohammadi, M., Rasoulpoor, S., & Khaledi-Paveh, B. (2020). Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: A systematic review and meta-analysis. *Globalization and Health*, *16*(1), 57. https://doi.org/10.1186/s12992-020-00589-w

Sanford, B., Ciarrochi, J., Hofmann, S., Chin, F., Gates, K., & Hayes, S. (2022). Toward empirical process-based case conceptualization: An idionomic network examination of the process-based assessment tool. *Journal of Contextual Behavioral Science*, *25*. https://doi.org/10.1016/j.jcbs.2022.05.006

Sbarra, D. A. (2015). Divorce and Health: Current Trends and Future Directions. *Psychosomatic Medicine*, *77*(3), 227. https://doi.org/10.1097/PSY.0000000000000168

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177.

Schaub, H., & Schiepek, G. (1992). Simulation of Psychological Processes: Basic Issues and an Illustration Within the Etiology of a Depressive Disorder. In W. Tschacher, G. Schiepek, & E. J. Brunner (Eds.), *Self-Organization and Clinical Psychology: Empirical Approaches to Synergetics in Psychology* (pp. 121–149). Springer. https://doi.org/10.1007/978-3-642-77534-5_6

Scheffer, M., Hosper, S. H., Meijer, M. L., Moss, B., & Jeppesen, E. (1993). Alternative equilibria in shallow lakes. *Trends in Ecology & Evolution*, *8*(8), 275–279. https://doi.org/10.1016/0169-5347(93)90254-M

Schiepek, G. (2003). A Dynamic Systems Approach to Clinical Case Formulation. *European Journal of Psychological Assessment*, *19*(3), 175–184. https://doi.org/10.1027//1015-5759.19.3.175

Schiepek, G. (2009). Complexity and Nonlinear Dynamics in Psychotherapy. *European Review*, *17*(2), 331–356. https://doi.org/10.1017/S1062798709000763

Schiepek, G., Aas, B., & Viol, K. (2016). The mathematics of psychotherapy: A nonlinear model of change dynamics. *Nonlinear Dynamics, Psychology, and Life Sciences*, *20*(3), 369–399.

Schiepek, G., Fartacek, C., Sturm, J., Kralovec, K., Fartacek, R., & Plöderl, M. (2011). Nonlinear dynamics: Theoretical perspectives and application to suicidology. *Suicide and Life-Threatening Behavior*, *41*, 661–675. https://doi.org/10.1111/j.1943-278X.2011.00062.x

Schiepek, G. K., Viol, K., Aichhorn, W., Hütt, M.-T., Sungler, K., Pincus, D., & Schöller, H. J. (2017). Psychotherapy Is Chaotic—(Not Only) in a Computational World. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00379

Schiepek, G., Stoger-Schmidinger, B., Aichhorn, W., Scholler, H., & Aas, B. (2016). Systemic case formulation, individualized process monitoring, and state dynamics in a case of dissociative identity disorder. *Frontiers in Psychology*.

Scholler, H., Viol, K., Goditsch, H., Aichhorn, W., Hutt, M.-T., & Schiepek, G. (2019). A Nonlinear Dynamic Systems Model of Psychotherapy: First Steps Toward Validation and the Role of External Input. *Nonlinear Dynamics, Psychology, and Life Sciences*, *23*(1), 79–112.

Scholten, S., Lischetzke, T., & Glombiewski, J. A. (2021). Integrating theory-based and data-driven methods to case conceptualization: A functional analysis approach with ecological momentary assessment. *Psychotherapy Research*, *0*(0), 1–13. https://doi.org/10.1080/10503307.2021.1916639

Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, *99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Schumacher, L., Burger, J., Zoellner, F., Zindler, A., Epskamp, S., & Barthel, D. (2021). Using clinical expertise and empirical data in constructing networks of trauma symptoms in refugee youth. *European Journal of Psychotraumatology*, *12*(1), 1920200. https://doi.org/10.1080/20008198.2021.1920200

Schuurman, N. K. (2023). *A "Within/Between Problem" Primer: About (Not) Separating Within-Person Variance and Between-Person Variance in Psychology*. PsyArXiv. https://doi.org/10.31234/osf.io/7zgkx

Schuurman, N. K., Grasman, R. P. P. P., & Hamaker, E. L. (2016). A Comparison of Inverse-Wishart Prior Specifications for Covariance Matrices in Multilevel Autoregressive Models. *Multivariate Behavioral Research*, *51*(2–3), 185–206. https://doi.org/10.1080/00273171.2015.1065398

Schweren, L., van Borkulo, C. D., Fried, E., & Goodyer, I. M. (2018). Assessment of Symptom Network Density as a Prognostic Marker of Treatment Response in Adolescent Depression. *JAMA Psychiatry*, *75*(1), 98–100. https://doi.org/10.1001/jamapsychiatry.2017.3561

Scott, J., Bellivier, F., Manchia, M., Schulze, T., Alda, M., Etain, B., & Collaboration, investigators involved in the C. (2020). Can network analysis shed light on predictors of lithium response in bipolar I disorder? *Acta Psychiatrica Scandinavica*, *141*(6), 522–533. https://doi.org/10.1111/acps.13163

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry*, *59 Suppl 20*, 22-33;quiz 34-57.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.*, *4*, 1–32.

Sim, K., Gwee, K. P., & Bateman, A. (2005). Case formulation in psychotherapy: Revitalizing its usefulness as a clinical tool. *Academic Psychiatry: The Journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry*, *29*(3), 289–292. https://doi.org/10.1176/appi.ap.29.3.289

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Skjerdingstad, N., Johnson, M. S., Johnson, S. U., Hoffart, A., & Ebrahimi, O. V. (2021). *Feelings of Worthlessness Links Depressive Symptoms and Parental Stress: A Network Analysis During the COVID-19 Pandemic*. PsyArXiv. https://doi.org/10.31234/osf.io/w2eg3

Smaldino, P. E. (2017). Models Are Stupid, and We Need More of Them. In R. R. Vallacher, S. J. Read, & A. Nowak (Eds.), *Computational Social Psychology* (1st ed., pp. 311–331). Routledge. https://doi.org/10.4324/9781315173726-14

Smaldino, P. E. (2020). How to translate a verbal theory into a formal model. *Social Psychology*, *51*(4), 207–218. https://doi.org/10.1027/1864-9335/a000425

REFERENCES

Smith, K. E., Mason, T. B., Crosby, R. D., Cao, L., Leonard, R. C., Wetterneck, C. T., Smith, B. E. R., Farrell, N. R., Riemann, B. C., Wonderlich, S. A., & Moessner, M. (2019). A comparative network analysis of eating disorder psychopathology and co-occurring depression and anxiety symptoms before and after treatment. *Psychological Medicine*, *49*(2), 314–324. https://doi.org/10.1017/S0033291718000867

Snippe, E., Viechtbauer, W., Geschwind, N., Klippel, A., de Jonge, P., & Wichers, M. (2017). The Impact of Treatments for Depression on the Dynamic Network Structure of Mental States: Two Randomized Controlled Trials. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/srep46523

Solbakken, O. A., Ebrahimi, O. V., Hoffart, A., Monsen, J. T., & Johnson, S. U. (2023). Emotion regulation difficulties and interpersonal problems during the COVID-19 pandemic: Predicting anxiety and depression. *Psychological Medicine*, *53*(5), 2181–2185. https://doi.org/10.1017/S0033291721001987

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Stan Development Team. (2022). *RStan: The R interface to Stan*. R package version 2.26.6.

Starr, L. R., & Davila, J. (2012). Temporal patterns of anxious and depressed mood in generalized anxiety disorder: A daily diary study. *Behaviour Research and Therapy*, *50*(2), 131–141. https://doi.org/10.1016/j.brat.2011.11.005

Stefan, A. M., Evans, N. J., & Wagenmakers, E.-J. (2020). Practical challenges and methodological flexibility in prior elicitation. *Psychological Methods*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/met0000354

Stein, M. D., & Friedmann, P. D. (2006). Disturbed Sleep and Its Relationship to Alcohol Use. *Substance Abuse*, *26*(1), 1–13. https://doi.org/10.1300/J465v26n01_01

Steinhausen, H.-C. (2009). Outcome of eating disorders. *Child and Adolescent Psychiatric Clinics of North America*, *18*(1), 225–242. https://doi.org/10.1016/j.chc.2008.07.013

Stice, E., Marti, C. N., Spoor, S., Presnell, K., & Shaw, H. (2008). Dissonance and healthy weight eating disorder prevention programs: Long-term effects from a randomized efficacy trial. *Journal of Consulting and Clinical Psychology*, *76*(2), 329–340. https://doi.org/10.1037/0022-006X.76.2.329

Stoger-Schmidinger, B., Aichhorn, W., Scholler, H., Aas, B., & Schiepek, G. (2016). Systemic case formulation and state dynamics in a female patient with structural personality dissociation. *Familiendynamik: Systemische Praxis und Forschung*, *41*(4), 322–332. https://doi.org/10.21706/FD-41-4-322

Stone, A. A., & Shiffman, S. (1994). Ecological Momentary Assessment (Ema) in Behavioral Medicine. *Annals of Behavioral Medicine*, *16*(3), 199–202. https://doi.org/10.1093/abm/16.3.199

Strauss, G. P., Zamani Esfahlani, F., Sayama, H., Kirkpatrick, B., Opler, M. G., Saoud, J. B., Davidson, M., & Luthringer, R. (2020). Network Analysis Indicates That Avolition Is the Most Central Domain for the Successful Treatment of Negative Symptoms: Evidence From the Roluperidone Randomized Clinical Trial. *Schizophrenia Bulletin*, *46*(4), 964–970. https://doi.org/10.1093/schbul/sbz141

Stroebe, M., Stroebe, W., Schut, H., & Boerner, K. (2017). Grief is not a disease but bereavement merits medical awareness. *The Lancet*, *389*(10067), 347–349. https://doi.org/10.1016/S0140-6736(17)30189-7

Strunk, G., & Schiepek, G. (2006). *Systemische Psychologie: Eine Einführung in die komplexen Grundlagen menschlichen Verhaltens*. Elsevier, Spektrum Akad. Verlag.

Swoyer, C. (1991). Structural representation and surrogative reasoning. *Synthese*, *87*(3), 449–508. https://doi.org/10.1007/BF00499820

Taylor, J. A. (1953). A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology*, *48*(2), 285–290. https://doi.org/10.1037/h0056264

Tesli, M. S., Handal, M., Kirkøen, B., Torvik, F. A., Knudsen, A. K. S., Odsbu, I., Gustavson, K., Nesvåg, R., Hauge, L. J., & Reneflot, A. (2016). *Mental illness among adults*. Norwegian Institute of Public Health. https://www.fhi.no/en/op/hin/mental-health/psykisk-helse-hos-voksne/

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2). Reading, Mass.

Van Borkulo, C., Boschloo, L., Borsboom, D., Penninx, B. W. J. H., Waldorp, L. J., & Schoevers, R. A. (2015). Association of Symptom Network Structure With the Course of Depression. *JAMA Psychiatry*, *72*(12), 1219. https://doi.org/10.1001/jamapsychiatry.2015.2079

Van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, T. F., Boschloo, L., Schoevers, R. A., & Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Scientific Reports*, *4*. https://doi.org/10.1038/srep05918

van Borkulo, C. D., Boschloo, L., Kossakowski, J. J., Tio, P., Schoevers, R. A., Borsboom, D., & Waldorp, L. J. (2017). Comparing network structures on three aspects: A permutation test. *Manuscript Submitted*. https://doi.org/10.13140/RG.2.2.29455.38569

van Borkulo, C. D., Epskamp, S., & Robitzsch, A. (2014). IsingFit: Fitting Ising models using the eLasso method. *R Package Version 2.0*.

van Borkulo, C. D., Millner, A., & van Borkulo, C. D. (2017). *NetworkComparisonTest: Statistical Comparison of Two Networks Based on Three Invariance Measures*.

van Borkulo, C. D., van Bork, R., Boschloo, L., Kossakowski, J. J., Tio, P., Schoevers, R. A., Borsboom, D., & Waldorp, L. J. (2022). Comparing network structures on three aspects: A permutation test. *Psychological Methods*. https://doi.org/10.1037/met0000476

van Os, J., Delespaul, P., Wigman, J., Myin-Germeys, I., & Wichers, M. (2013). Beyond DSM and ICD: Introducing "precision diagnosis" for psychiatry using momentary assessment technology. *World Psychiatry*, *12*(2), 113–117. https://doi.org/10.1002/wps.20046

van Rooij, I., & Baggio, G. (2021). Theory Before the Test: How to Build High-Verisimilitude Explanatory Theories in Psychological Science. *Perspectives on Psychological Science*, 1745691620970604. https://doi.org/10.1177/1745691620970604

van Rooij, I., & Blokpoel, M. (2020). Formalizing verbal theories: A tutorial by dialogue. *Social Psychology*, *51*(5), 285–298. https://doi.org/10.1027/1864-9335/a000428

Vannucci, A., Flannery, K. M., & Ohannessian, C. M. (2017). Social media use and anxiety in emerging adults. *Journal of Affective Disorders*, *207*, 163–166. https://doi.org/10.1016/j.jad.2016.08.040

Vansteenkiste, M., Lens, W., Soenens, B., & Luyckx, K. (2006). Autonomy and Relatedness among Chinese Sojourners and Applicants: Conflictual or Independent Predictors of Well-Being and Adjustment? *Motivation and Emotion*, *30*(4), 273–282. https://doi.org/10.1007/s11031-006-9041-x

Vittengl, J. R., Clark, L. A., Thase, M. E., & Jarrett, R. B. (2022). Does symptom linkage density predict outcomes in cognitive therapy for recurrent depression? *Journal of Psychopathology and Behavioral Assessment*, *44*, 469–480. https://doi.org/10.1007/s10862-021-09914-y

Von Kentzinsky, H., Wijtsma, S., & Treur, J. (2020). A Temporal-Causal Modelling Approach to Analyse the Dynamics of Burnout and the Effects of Sleep. In X.-S. Yang, S. Sherratt, N. Dey, & A. Joshi (Eds.), *Fourth International Congress on Information and Communication Technology* (Vol. 1027, pp. 219–232). Springer Singapore. https://doi.org/10.1007/978-981-32-9343-4_18

von Klipstein, L., Riese, H., van der Veen, D. C., Servaas, M. N., & Schoevers, R. A. (2020). Using person-specific networks in psychotherapy: Challenges, limitations, and how we could use them anyway. *BMC Medicine*, *18*(1), 345. https://doi.org/10.1186/s12916-020-01818-0

Wade, M., Prime, H., & Browne, D. T. (2020). Why we need longitudinal mental health research with children and youth during (and after) the COVID-19 pandemic. *Psychiatry Research*, *290*, 113143. https://doi.org/10.1016/j.psychres.2020.113143

Wagner, A. A., Hall, M., Scherner, P., & Rubel, J. (2023). *If You Give a Therapist a Network: A Qualitative Analysis of Therapists' Reactions to their Patients' EMA-Based Network Models* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/hqyft

Waller, G. (2009). Evidence-based treatment and therapist drift. *Behaviour Research and Therapy*, *47*(2), 119–127. https://doi.org/10.1016/j.brat.2008.10.018

Waller, G. (2016). Treatment Protocols for Eating Disorders: Clinicians' Attitudes, Concerns, Adherence and Difficulties Delivering Evidence-Based Psychological Interventions. *Current Psychiatry Reports*, *18*, 36. https://doi.org/10.1007/s11920-016-0679-0

Walsh, B. T., Xu, T., Wang, Y., Attia, E., & Kaplan, A. S. (2021). Time Course of Relapse Following Acute Treatment for Anorexia Nervosa. *American Journal of Psychiatry*, *178*(9), 848–853. https://doi.org/10.1176/appi.ajp.2021.21010026

Wang, Y., Hu, Z., Feng, Y., Wilson, A., & Chen, R. (2020). Changes in network centrality of psychopathology symptoms between the COVID-19 outbreak and after peak. *Molecular Psychiatry*, *25*(12), Article 12. https://doi.org/10.1038/s41380-020-00881-6

Wangersky, P. J. (1978). Lotka-Volterra Population Models. *Annual Review of Ecology and Systematics*, *9*(1), 189–218. https://doi.org/10.1146/annurev.es.09.110178.001201

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), Article 6684. https://doi.org/10.1038/30918

Wegerif, R., Mercer, N., & Dawes, L. (1999). From social interaction to individual reasoning: An empirical investigation of a possible socio-cultural model of cognitive development. *Learning and Instruction*, *9*(6), 493–516. https://doi.org/10.1016/S0959-4752(99)00013-4

Wells, A. (2009). *Metacognitive therapy for anxiety and depression* (pp. xvii, 316). Guilford Press.

Wen, C. K. F., Schneider, S., Stone, A. A., & Spruijt-Metz, D. (2017). Compliance With Mobile Ecological Momentary Assessment Protocols in Children and Adolescents: A Systematic Review and Meta-Analysis. *Journal of Medical Internet Research*, *19*(4), e132. https://doi.org/10.2196/jmir.6641

Wendt, L. P., Wright, A. G. C., Pilkonis, P. A., Woods, W. C., Denissen, J. J. A., Kühnel, A., & Zimmermann, J. (2020). Indicators of Affect Dynamics: Structure, Reliability, and Personality Correlates. *European Journal of Personality*, *34*(6), 1060–1072. https://doi.org/10.1002/per.2277

Wensing, M., & Grol, R. (2019). Knowledge translation in health: How implementation science could contribute more. *BMC Medicine*, *17*(1), 88. https://doi.org/10.1186/s12916-019-1322-9

Wichers, M. (2014). The dynamic nature of depression: A new micro-level perspective of mental disorder that meets current challenges. *Psychological Medicine*, *44*(7), 1349–1360. https://doi.org/10.1017/S0033291713001979

Wichers, M., Riese, H., Hodges, T. M., Snippe, E., & Bos, F. M. (2021a). A Narrative Review of Network Studies in Depression: What Different Methodological Approaches Tell Us About Depression. *Frontiers in Psychiatry*, *12*(719490). https://doi.org/10.3389/fpsyt.2021.719490

Wichers, M., Riese, H., Hodges, T. M., Snippe, E., & Bos, F. M. (2021b). A Narrative Review of Network Studies in Depression: What Different Methodological Approaches Tell Us About Depression. *Frontiers in Psychiatry*, *12*, 719490. https://doi.org/10.3389/fpsyt.2021.719490

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wigboldus, D. H. J., & Dotsch, R. (2016). Encourage Playing with Data and Discourage Questionable Reporting Practices. *Psychometrika*, *81*, 27–32. https://doi.org/10.1007/s11336-015-9445-1

Wild, B., Eichler, M., Friederich, H.-C., Hartmann, M., Zipfel, S., & Herzog, W. (2010a). A graphical vector autoregressive modelling approach to the analysis of electronic diary data. *BMC Medical Research Methodology*, *10*(1), 28. https://doi.org/10.1186/1471-2288-10-28

Wild, B., Eichler, M., Friederich, H.-C., Hartmann, M., Zipfel, S., & Herzog, W. (2010b). A graphical vector autoregressive modelling approach to the analysis of electronic diary data. *BMC Medical Research Methodology*, *10*(1), 28. https://doi.org/10.1186/1471-2288-10-28

Williams, D. R. (2021). Bayesian Estimation for Gaussian Graphical Models: Structure Learning, Predictability, and Network Comparisons. *Multivariate Behavioral Research*, *0*(0), 1–17. https://doi.org/10.1080/00273171.2021.1894412

Williams, D. R., & Mulder, J. (2020). BGGM: Bayesian Gaussian Graphical Models in R. *The Journal of Open Source Software*, *5*(51). https://doi.org/10.21105/joss.02111

Williams, D. R., Rast, P., Pericchi, L. R., & Mulder, J. (2020). Comparing Gaussian Graphical Models With the Posterior Predictive Distribution and Bayesian Model Selection. *Psychological Methods*, *25*(5), 653–672. https://doi.org/10.1037/met0000254

Williams, D. R., Rhemtulla, M., Wysocki, A. C., & Rast, P. (2019). On Nonregularized Estimation of Psychological Networks. *Multivariate Behavioral Research*, 719–750. https://doi.org/10.1080/00273171.2019.1575716

Winkler, P., Formanek, T., Mlada, K., Kagstrom, A., Mohrova, Z., Mohr, P., & Csemy, L. (2020). Increase in prevalence of current mental disorders in the context of COVID-19: Analysis of repeated nationwide cross-sectional surveys. *Epidemiology and Psychiatric Sciences*, *29*, e173. https://doi.org/10.1017/S2045796020000888

Wójcik, G., Zawisza, K., Jabłońska, K., Grodzicki, T., & Tobiasz-Adamczyk, B. (2021). Transition out of Marriage and its Effects on Health and Health–Related Quality of Life among Females and Males. COURAGE and COURAGE-POLFUS–Population Based Follow-Up Study in Poland. *Applied Research in Quality of Life*, *16*(1), 13–49. https://doi.org/10.1007/s11482-019-09742-z

World Health Organization. (2023). *Mental health Burden*. https://www.who.int/health-topics/mental-health

Worldometers. (2021). *COVID - Coronavirus Statistics—Worldometer*. https://www.worldometers.info/coronavirus/

Wright, A. G. C., & Woods, W. C. (2020). Personalized models of psychopathology. *Annual Review of Clinical Psychology*, 49–74. https://doi.org/10.1146/annurev-clinpsy-102419-125032

Wrzus, C., Hänel, M., Wagner, J., & Neyer, F. J. (2013). Social network changes and life events across the life span: A meta-analysis. *Psychological Bulletin*, *139*, 53–80. https://doi.org/10.1037/a0028601

Xiong, J., Lipsitz, O., Nasri, F., Lui, L. M. W., Gill, H., Phan, L., Chen-Li, D., Iacobucci, M., Ho, R., Majeed, A., & McIntyre, R. S. (2020). Impact of COVID-19 pandemic on mental health in the general population: A systematic review. *Journal of Affective Disorders*, *277*, 55–64. https://doi.org/10.1016/j.jad.2020.08.001

Yang. (2021). *Pompom* [Computer software].

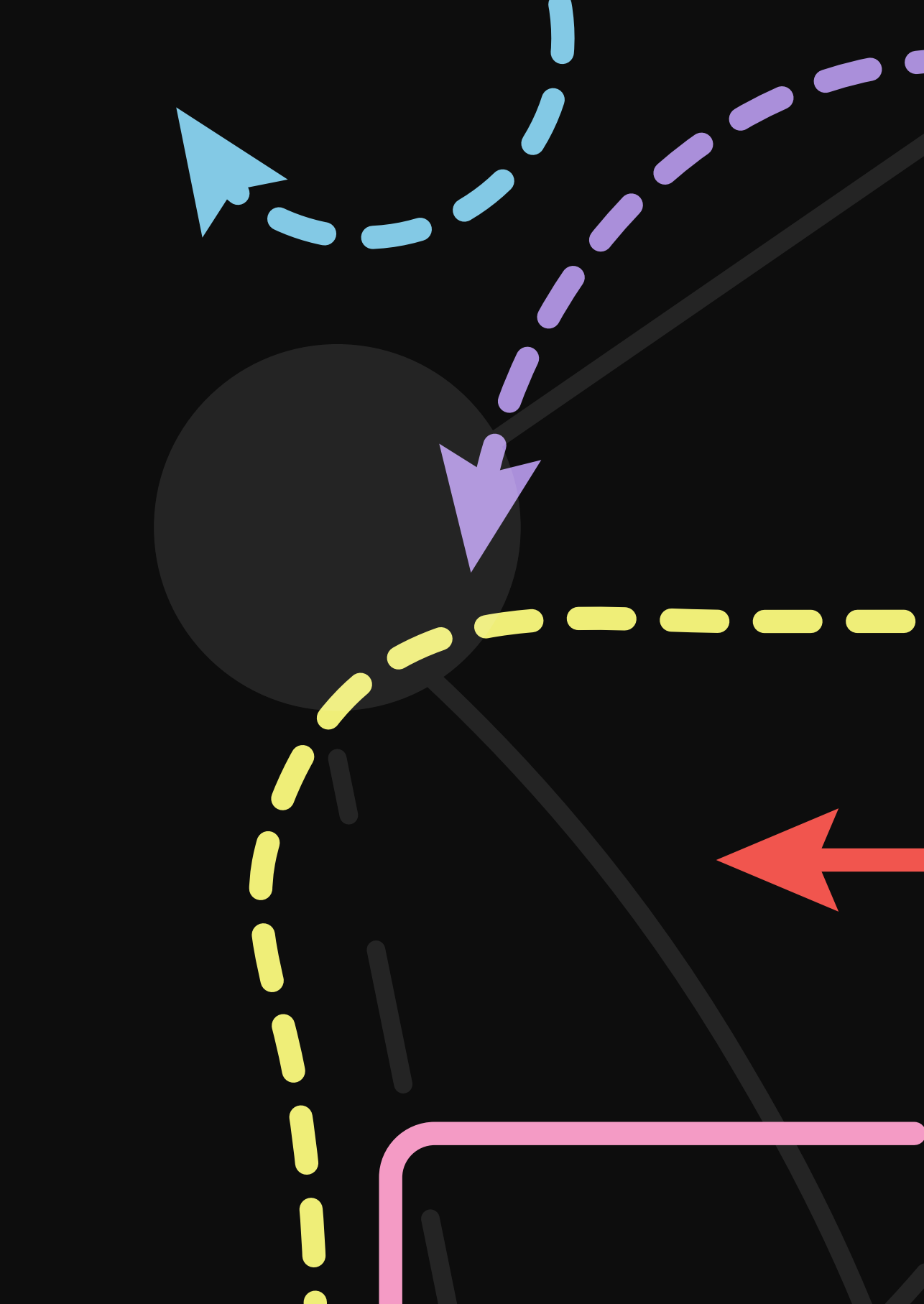
Yao, H., Chen, J.-H., & Xu, Y.-F. (2020). Patients with mental health disorders in the COVID-19 epidemic. *The Lancet Psychiatry*, *7*(4), e21. https://doi.org/10.1016/S2215-0366(20)30090-0

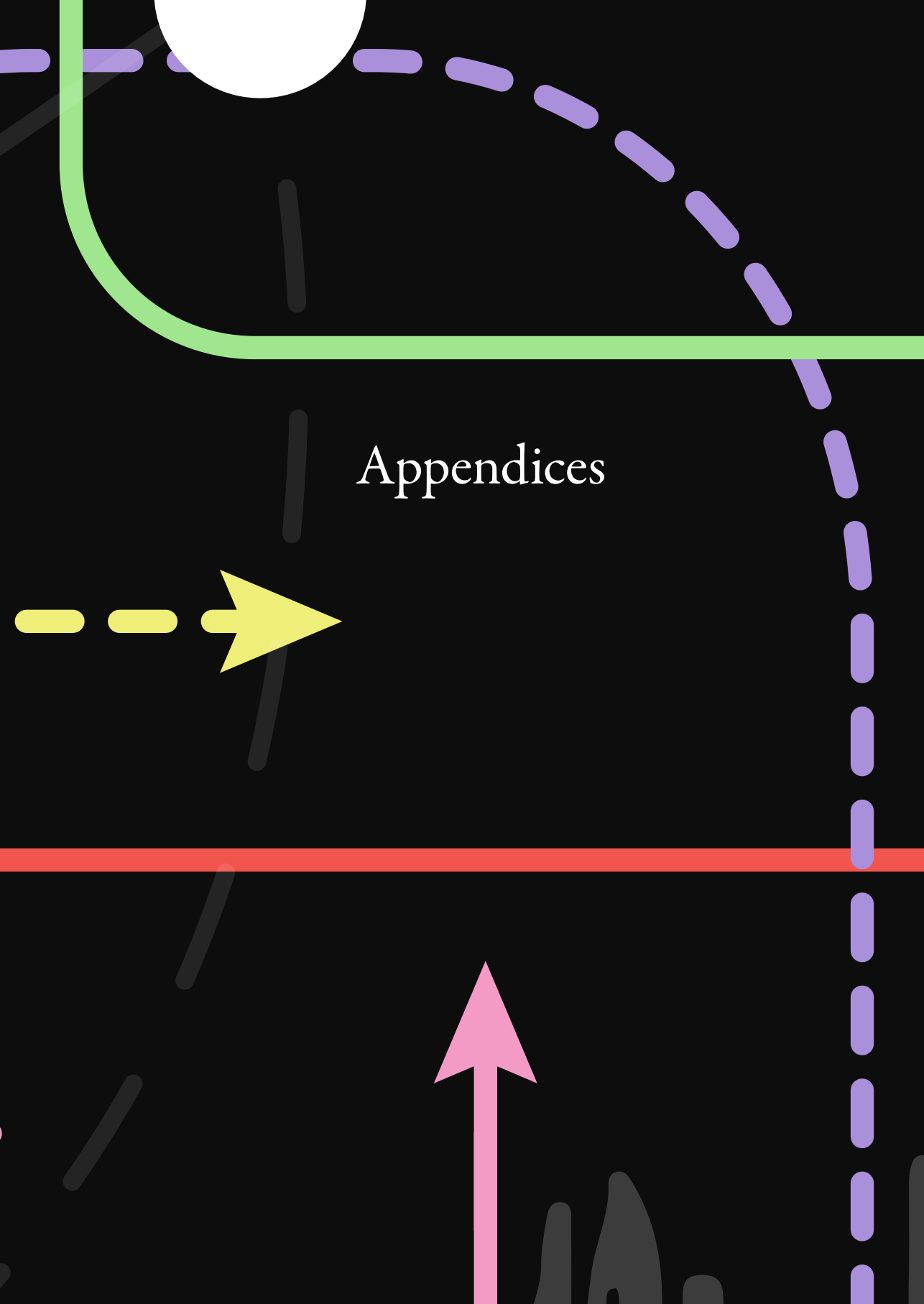
Zhou, J., Liu, S., Mayes, T. L., Feng, Y., Fang, M., Xiao, L., & Wang, G. (2022). The network analysis of depressive symptoms before and after two weeks of antidepressant treatment. *Journal of Affective Disorders*, *299*, 126–134. https://doi.org/10.1016/j.jad.2021.11.059

Zimmerman, M., Ellison, W., Young, D., Chelminski, I., & Dalrymple, K. (2015). How many different ways do patients meet the diagnostic criteria for major depressive disorder? *Comprehensive Psychiatry*, *56*, 29–34. https://doi.org/10.1016/j.comppsych.2014.09.007

Zimmermann, J., Woods, W. C., Ritter, S., Happel, M., Masuhr, O., Jaeger, U., Spitzer, C., & Wright, A. G. C. (2019). Integrating structure and dynamics in personality assessment: First steps toward the development and validation of a personality dynamics diary. *Psychological Assessment*, *31*(4), 516–531. https://doi.org/10.1037/pas0000625

Zuidersma, M., Riese, H., Snippe, E., Booij, S. H., Wichers, M., & Bos, E. H. (2020). Single-Subject Research in Psychiatry: Facts and Fictions. *Frontiers in Psychiatry*, *11*, 539777. https://doi.org/10.3389/fpsyt.2020.539777




Appendices

APPENDICES

Summary

English summary

In this thesis, I discussed what the future of case formulation could look like based on advancements in momentary assessment, statistical network estimation, and simulation-based formal modeling. The thesis is divided in five sections: First, laying out methodological background on estimation, second, discussing how statistical networks can be used for exploration, third, a proposal for integrating case formulations with statistical networks systematically, fourth, introducing the formalization of case formulations, and fifth, discussing future directions of the findings.

Part I: Methodological Background

The thesis started with methodological notes on conceptualizing research questions, creating research designs, estimating models, and reporting results in the context of network analysis. **Chapter 2** discussed the importance of relating longitudinal design choices to characteristics of the data. There are different types of longitudinal data (e.g., single measurement data, $N = 1$ time series data, $N > 1$ time series data, panel data) that can be used to estimate statistical networks. The characteristics of the data, along with the choice of the statistical model to analyze them, determine the precise interpretation of the resulting edges. In addition, interpretation of effects are dependent on specific characteristics of the research design, such as the assessed time scales. **Chapter 3** described how longitudinal networks can be estimated from $N = 1$ and $N > 1$ time series data, such as data collected via Ecological Momentary Assessment (EMA). The graphical vector auto-regressive (GVAR) model can be applied to this type of data to estimate contemporaneous and temporal effects, indicating associations between variables within the same time frame and over time, respectively. The same model can be estimated in a multi-level fashion for multiple individuals, resulting in fixed effects estimates, as well as between subject estimates for averages of person-wise means. There are specific challenges to this type of modeling technique, which include potentially unfeasible power requirements, and strong statistical assumptions such as stationarity. **Chapter 4** shed light on using network analysis for evaluating treatment effects. In the empirical literature, there is a wide variety of design and analysis choices to address this question, ranging from cross-sectional networks estimated from RCT data that include the treatment as a binary node, to personalized time series networks estimated before and after treatment. This systematic review highlighted the need for clear reporting and open science practices in the context of treatment evaluation. **Chapter 5** introduced general reporting standards for cross-sectional network analysis for the most common research aims found in the empirical network analysis literature. The chapter also illustrated that reporting standards are not only important for scientific rigor in writing articles, but also for the research design and planning phase, because they allow to make decisions based on anticipating the precise analytic challenges that may arise.

Part II: Exploration – Statistical networks based on empirical data

The second part of the thesis presented empirical network contributions that provide exploratory insights relevant for case formulations given a specific context. **Chapter 6** presented a multi-level longitudinal network analysis of 1368 individuals who completed 30 daily assessments on anxiety symptoms during the COVID-19 pandemic. The results showed that anxiety symptoms were espe-

cially well predicted by uncontrollability of worry, generalized worry, fear of being infected, fear of significant others being infected, and threat monitoring on the previous day. **Chapter 7** presented a multi-level longitudinal network analysis of the same population in the context of depressive symptoms. The main findings of this study were that depressive symptoms were mostly predicted by experiences of helplessness during the previous day, while within the same day, anhedonia, emotion regulation deficits, and lethargy, were most predictive. **Chapter 8** presented a cross-sectional network analysis of 724 older individuals. The study investigated between-subject relationships of depressive symptoms following two severe adverse life events, spousal loss and separation, and found that separated compared to bereaved individuals were more likely to experience an unfriendly environment and oneself as a failure. For both life events, the network showed strong relations with loneliness, which was in turn connected with a host of other depressive symptoms.

Part III: Integration – Combining clinical prior information with statistical networks

The third part of the thesis introduced a new framework to systematically integrate case formulations with personalized networks. **Chapter 9** presented the Prior Elicitation Module for Idiographic System Estimation (PREMISE). PREMISE is a novel approach that formally integrates case formulations with personalized network estimation via prior elicitation and Bayesian inference. In doing so, it addresses some of the most pressing issues in implementing personalized network modeling: Estimation is more efficient because readily available clinical information are used, and clinicians can include theories and person-specific information in the model. The chapter showcased the clinical utility of PREMISE using the case formulation and EMA data from a patient diagnosed with OCD. **Chapter 10** presented a PREMISE investigation of two patients diagnosed with Anorexia Nervosa, specifically focusing on treatment implications. The main finding was that PREMISE networks for both patients had different implications for centrality-based treatment targets depending on the type of model (PREMISE versus traditional versus case formulation network). In particular, for one of the patients, the PREMISE network could be matched to CBT treatment modules to reduce excessive exercising and exposure approaches for fear of weight gain, whereas cognitive symptoms were more prominent in the data-driven network, calling for cognitive modules within the CBT-E protocol. **Chapter 11** introduced an alternative approach to combining longitudinal assessments with constructing networks, the Longitudinal Perceived Causal Relations (L-PCR) approach. The chapter illustrated how L-PCR networks can circumvent several of the potentially unfeasible assumptions in statistical estimation of personalized networks, such as the issue of restrictive time scales. Using data from 20 participants who completed between 20 to 28 daily assessments of depressive symptom relations, this chapter showed that L-PCR is generally feasible, well accepted, and leads to clinically relevant insights on the structure and stability of perceived causal networks.

Part IV: Formalization – Computational models of case formulations

The fourth part of this thesis discussed the potential of formalizing case formulations, and using a simulation-based approach to evaluate case formulations and clinical treatments for a given patient. **Chapter 12** illustrated the benefits of this simulation-based approach using an example of a patient diagnosed with panic disorder. The resulting computational model showed how case formulations can be evaluated by means of simulation results. In this specific example, the simulations aligned with general dynamics of panic disorder, but specific observations, such as rapid onset and decline of panic attacks were not depicted. The chapter discussed how these inconsistencies can give rise to refining the case formulation, overall contributing to stronger theories.

Part V: Conclusions

The final part of this thesis integrated the findings of the parts and chapters and answered the three main research questions. Based on the current evidence, I see promise in PREMISE and L-PCR-based networks to aid communication in therapy. The validity, as well as the clinical utility of these approaches is yet to be established in randomized controlled trials. Regarding the feasibility and clinical utility of formalized case formulations I remain sceptic, at least in the current landscape. In this final part, I have outlined steps toward making the formalization of case formulations more feasible. The success of this approach will largely depend on if these steps can be realized. Should this indeed be possible at some point in the future, we can look forward to a new era of case formulations.

APPENDICES

Nederlandse Samenvatting

Nederlandse samenvatting

In dit proefschrift getiteld "The Future of Case Formulation in Clinical Psychology – Advancements in Network Modeling and Simulation-based Science" onderzoek ik hoe de toekomst van psychologische casusformulering eruit zou kunnen zien op basis van de huidige ontwikkelingen in statistische netwerkschattingen, formele modellen en computersimulaties. Het proefschrift bestaat uit vijf delen. In het eerste deel beschrijf ik de methodologische achtergrond van statistische netwerkmodellen, zoals statistische schattingen, onderzoeksdesigns, en rapportagestandaarden. In het tweede deel onderzoek ik hoe statistische netwerken gebruikt kunnen worden voor de ontwikkeling van casusformulering. Het derde deel onderzoekt hoe casusformuleringen met statistische netwerken gecombineerd kunnen worden. In het vierde deel introduceer ik de formalisatie van casusformuleringen via computermodellen en simulaties. In het laatste deel bespreek ik tenslotte verdere stappen om de resultaten van dit proefschrift in de praktijk te implementeren en om klinisch te valideren.

Deel I: Methodologische achtergrond

In de inleiding van het proefschrift beschrijf ik de methodologische achtergrond en onderbouwing van de in dit proefschrift gestelde onderzoeksvragen. Hoofdstuk 2 bespreekt het belang van het relateren van longitudinale statistische designs aan kenmerken van de te gebruiken tijdreeks data. Er zijn verschillende soorten data (bijv. cross-sectionele data, $N = 1$ time series data, $N > 1$ time series data, of panel data) die gebruikt kunnen worden om statistische netwerken te schatten. De kenmerken van de data, samen met de keuze van het statistische model om deze te analyseren, bepalen de resultaten en daarmee de specifieke interpretatie van de resulterende connecties. Bovendien zijn interpretaties van effecten afhankelijk van kenmerken van het design, zoals de gekozen tijdschalen waarop de data zijn verzameld.

Hoofdstuk 3 beschrijft hoe longitudinale netwerken geschat kunnen worden uit $N = 1$ en $N > 1$ tijdreeksgegevens, zoals gegevens verzameld via Ecological Momentary Assessment (EMA). Het Graphical Vector Autoregressive (GVAR) model kan worden toegepast op dit type data om gelijktijdige en temporale effecten tussen variabelen te schatten. Eenzelfde model kan worden geschat op data die zijn verzameld door verschillenden individuen in een multi-level GVAR. Hieruit resulteren schattingen van vaste effecten ('fixed effects') en schattingen voor individuen en hun variaties ('random effects'). Er zijn specifieke uitdagingen bij deze modelleringstechniek, zoals potentieel onhaalbare vereisten ten aanzien van noodzakelijke statistische power en sterke statistische aannames zoals stationariteit van de tijdreeks data.

Hoofdstuk 4 onderzoekt, in een systematische literatuurstudie, het gebruik van netwerkanalyse voor het evalueren van behandelingseffecten op mentale klachten. In de empirische literatuur zijn verschillende design- en analysemogelijkheden om deze vraag te beantwoorden, onder andere cross-sectionele netwerken geschat uit RCT-gegevens waarin de behandeling als een binair knooppunt wordt opgenomen, of gepersonaliseerde time-series-netwerken geschat voor en na de behandeling. Deze systematische review benadrukt de noodzaak van precieze rapportage en open wetenschappelijke praktijken in de context van behandelingsbeoordeling.

Hoofdstuk 5 introduceert algemene rapportagestandaarden voor cross-sectionele netwerkanalysen voor de meest voorkomende onderzoeksdoelen die in de empirische netwerkanalyseliteratuur worden gevonden. Het hoofdstuk illustreert ook dat rapportagestandaarden niet alleen belangrijk zijn voor wetenschappelijke nauwkeurigheid bij het schrijven van artikelen, maar ook voor de hieraan voorafgaande fase van design en planning, omdat ze besluitvorming mogelijk maken op basis van het anticiperen van specifieke analytische uitdagingen die kunnen ontstaan. Om toekomstig onderzoek eenvoudiger te kunnen pre-registreren is een pre-registratie template voor cross-sectionele netwerkanalyse ontwikkeld.

Deel II: *Exploratie* - Statistische netwerken gebaseerd op empirische data

Het tweede deel van dit proefschrift onderzoekt empirische netwerkbijdragen die exploratief inzicht kunnen bieden voor psychologische casusformuleringen in de klinische praktijk. Hoofdstuk 6 presenteert een longitudinale multi-level netwerkanalyse van 1368 individuen die gedurende de COVID-19-pandemie, gedurende 30 dagen een vragenlijst over angstsymptomen hebben ingevuld. De resultaten tonen aan dat angstsymptomen vooral voorspeld werden door oncontroleerbare zorgen, zorgen in het algemeen, angst om geïnfecteerd te raken en angst dat naasten geïnfecteerd raken.

Hoofdstuk 7 presenteert een longitudinale multi-level netwerkanalyse van dezelfde populatie als is gebruikt in hoofdstuk 6, voor depressieve symptomen. De belangrijkste bevindingen van dit onderzoek zijn dat depressieve symptomen voornamelijk voorspeld werden door ervaringen van hulpeloosheid tijdens de vorige dag, terwijl binnen dezelfde dag lusteloosheid, emotie regulatieproblemen en traagheid het meest voorspellend waren.

Hoofdstuk 8 presenteert een cross-sectionele netwerkanalyse van 724 oudere volwassenen. Dit hoofdstuk onderzoekt de relaties depressieve symptomen en twee ernstige negatieve levensgebeurtenissen, het verlies van een echtgenoot en scheiding. Er is gevonden dat gescheiden individuen in vergelijking met rouwende individuen vaker een onvriendelijke omgeving ervoeren of en gevoel van persoonlijk falen. Voor beide levensgebeurtenissen toont het netwerk sterke relaties met eenzaamheid, wat op zijn beurt verbonden is met een reeks andere depressieve symptomen.

Deel III: *Integratie* - Combineren van klinische voorkennis met statistische netwerken

In het derde deel van mijn proefschrift introduceer ik een nieuw kader om casusformuleringen systematisch te integreren met gepersonaliseerde netwerken die zijn geschat op basis van EMA data. In hoofdstuk 9 wordt de Prior Elicitation Module for Idiographic System Estimation (PREMISE) beschreven. PREMISE is een nieuw benadering die casusformuleringen formeel integreert met gepersonaliseerde netwerkschatting via klinische voorkennis en Bayesiaanse inferentie. Hierdoor worden enkele problemen bij de implementatie van gepersonaliseerde netwerkmodellering aangepakt. Bijvoorbeeld, omdat direct beschikbare klinische informatie wordt gebruikt is de statistische schatting efficiënter. Dit wil zeggen dat er minder data nodig is om een statistisch model te kunnen schatten (grotere power). Verder kunnen clinici theorieën en voorkennis over verbanden in het model opnemen. Het hoofdstuk demonstreert de klinische bruikbaarheid van PREMISE aan de hand van de casusformulering en EMA-gegevens van een patiënt met de diagnose OCD.

Hoofdstuk 10 presenteert een PREMISE-onderzoek naar twee patiënten met de diagnose Anorexia Nervosa, met focus op de interventies om de klachten te verminderen. De belangrijkste bevinding is dat PREMISE-netwerken voor beide patiënten verschillende implicaties hebben voor behandeldoelen op basis van centraliteit, afhankelijk van het type model (PREMISE-netwerk versus een netwerk op basis van traditionele statistische schatting versus een netwerk op basis van casusformulering). Met name voor één van de patiënten kan het PREMISE-netwerk worden gekoppeld aan CGT-behandelmodules om overmatig bewegen verminderen en de blootstelling aan de angst voor gewichtstoename, terwijl cognitieve symptomen prominenter zijn in het datanetwerk, wat wijst op cognitieve modules binnen het CGT-E-protocol.

Hoofdstuk 11 introduceert een alternatieve benadering om longitudinale beoordelingen te combineren met het construeren van netwerken, de Longitudinal Perceived Causal Relations (L-PCR) benadering. Het hoofdstuk illustreert hoe L-PCR-netwerken enkele van de potentieel onhaalbare aannames in statistische schattingen van gepersonaliseerde netwerken kunnen omzeilen, zoals het probleem van tijdschalen. Met behulp van data van 20 deelnemers die tussen de 20 en 28 dagelijkse beoordelingen van depressieve symptoomrelaties hebben voltooid, toont dit hoofdstuk aan dat L-PCR over het algemeen haalbaar is, goed geaccepteerd wordt en leidt tot klinisch relevante inzichten in de structuur en stabiliteit van waargenomen causale netwerken.

Deel IV: *Formalisatie* – Computersimulaties van casusformuleringen

Het vierde deel van dit proefschrift bespreekt het nut van het formaliseren van casusformuleringen en het gebruik van computersimulaties om casusformuleringen en klinische behandelingen voor een patiënt te evalueren. Hoofdstuk 12 illustreert de voordelen van deze aanpak aan de hand van een voorbeeld van een patiënt met een diagnose paniekstoornis. Het resulterende model toont aan hoe casusformuleringen geëvalueerd kunnen worden aan de hand van simulatieresultaten. In dit specifieke voorbeeld komen de simulaties overeen met algemene dynamieken van paniekstoornis, maar specifieke observaties, zoals plotselinge toename en afname van paniekaanvallen, worden niet weergegeven. Het hoofdstuk bespreekt hoe deze inconsistenties kunnen leiden tot het verfijnen van de casusformulering en zo bijdragen aan sterkere theorieën.

Deel V: Conclusies

Het laatste deel van dit proefschrift integreert de bevindingen van de delen en hoofdstukken en beantwoordt de drie belangrijkste onderzoeksvragen. Op basis van de evidentie zie ik een belofte in PREMISE- en L-PCR-gebaseerde netwerken om communicatie in behandeling te ondersteunen. De validiteit en klinische bruikbaarheid van deze benaderingen moeten echter nog worden vastgesteld in gerandomiseerde gecontroleerde trials (zg., RCT's), of ander wetenschappelijk onderzoek. Wat betreft de haalbaarheid en klinische bruikbaarheid van formalisering van casusformuleringen blijf ik sceptisch, tenminste in de huidige situatie. In dit laatste deel zet ik stappen uiteen om het formaliseren van casusformuleringen haalbaarder te maken door modellen te simplificeren. Het succes van deze benadering zal grotendeels afhangen van de vraag of deze stappen kunnen worden gerealiseerd. Mocht dit inderdaad op enig moment in de toekomst mogelijk zijn, dan kunnen we uitkijken naar een nieuw tijdperk van casusformuleringen ondersteunt door computersimulaties.

APPENDICES

Publications

Publications

Chapter 2 is published as:
Epskamp, S., Hoekstra, R. H. A., **Burger, J**., & Waldorp, L. J. (2022). Chapter 9. Longitudinal design choices: Relating data to analysis. In Isvoranu, A. M., Epskamp, S., Waldorp, L. J., & Borsboom, D. (Eds.). *Network psychometrics with R: A guide for behavioral and social scientists*. Routledge, Taylor & Francis Group.

Chapter 3 is published as:
Burger, J., Hoekstra, R. H. A., Mansueto, A. C., & Epskamp, S. (2022). Chapter 10. Network estimation from time series and panel data. In Isvoranu, A. M., Epskamp, S., Waldorp, L., J. & Borsboom, D. (Eds.). *Network psychometrics with R: A guide for behavioral and social scientists*. Routledge, Taylor & Francis Group.

Chapter 4 is submitted for publication as:
Schumacher, L., **Burger, J.**, Echterhoff, J., & Kriston, L. (2022). Methodological and statistical practices of using symptom networks to evaluate mental health interventions: A systematic review. *PsyArXiv Preprint, under review.*

Chapter 5 is published as:
Burger, J., Isvoranu, A.-M., Lunansky, G., Haslbeck, J. M. B., Epskamp, S., Hoekstra, R. H. A., Fried, E. I., Borsboom, D., & Blanken, T. F. (2022). Reporting standards for psychological network analyses in cross-sectional data. *Psychological Methods.* Advance online publication.

Chapter 6 is published as:
Hoffart, A., **Burger, J.**, Johnson, S. U., & Ebrahimi, O. V. (2023). Daily dynamics and mechanisms of anxious symptomatology in the general population: A network study during the COVID-19 pandemic. *Journal of Anxiety Disorders*, *93*, 102658.

Chapter 7 is published as:
Ebrahimi, O. V., **Burger, J.**, Hoffart, A., & Johnson, S. U. (2021). Within-and across-day patterns of interplay between depressive symptoms and related psychopathological processes: a dynamic network approach during the COVID-19 pandemic. *BMC medicine*, *19*(1), 1-17.

Chapter 8 is published as:
Burger, J., Stroebe, M. S., Perrig-Chiello, P., Schut, H. A., Spahni, S., Eisma, M. C., & Fried, E. I. (2020). Bereavement or breakup: Differences in networks of depression. *Journal of Affective Disorders*, *267*, 1-8.

Chapter 9 is published as:
Burger, J., Epskamp, S., van der Veen, D. C., Dablander, F., Schoevers, R. A., Fried, E. I., & Riese, H. (2022). A clinical PREMISE for personalized models: Toward a formal integration of case formulations and statistical networks. *Journal of Psychopathology and Clinical Science*, *131*(8), 906.

Chapter 10 is published as:
Burger, J., Ralph-Nearman, C., & Levinson, C. A. (2022). Integrating clinician and client case conceptualization with momentary assessment data to construct idiographic networks: Moving toward personalized treatment for eating disorders. *Behaviour Research and Therapy*, *159*, 104221.

Chapter 11 is submitted for publication as:
Burger, J., Andikkhash, V., Jäger, N., Anderbro, T., Blanken, T., & Klintwall, L. (2022). A Novel Approach for Constructing Personalized Networks from Longitudinal Perceived Causal Relations. *PsyArXiv Preprint, under review.*

Chapter 12 is published as:
Burger, J., van der Veen, D. C., Robinaugh, D. J., Quax, R., Riese, H., Schoevers, R. A., & Epskamp, S. (2020). Bridging the gap between complexity science and clinical practice by formalizing idiographic theories: a computational model of functional analysis. *BMC medicine*, *18*, 1-18.

Other publications:

Derks, K., **Burger, J.**, van Doorn, J., Kossakowski, J. J., Matzke, D., Atticciati, L., ... & Wagenmakers, E. J. (2018). Network models to organize a dispersed literature: the case of misunderstanding analysis of covariance. *Journal of European Psychology Students*, *9*(1).

Karnbach, K., Witkowski, M., Ebrahimi, O. V., & **Burger, J.** (2022). Back to life, back to reality: A multi-level dynamic network analysis of student mental health upon return to campus during the COVID-19 pandemic. *Current Psychology*, 1-13.

Leising, D., **Burger, J.**, Zimmermann, J., Bäckström, M., Oltmanns, J. R., & Connelly, B. S. (2020). Why do items correlate with one another? A conceptual analysis with relevance for general factors and network models. *PsyArXiv Preprint, under review.*

Mares, S. H., **Burger, J.**, Lemmens, L. H., van Elburg, A. A., & Vroling, M. S. (2022). Evaluation of the cognitive behavioural theory of eating disorders: A network analysis investigation. *Eating Behaviors*, *44*, 101590.

Schumacher, L., **Burger, J.**, Zoellner, F., Zindler, A., Epskamp, S., & Barthel, D. (2021). Using clinical expertise and empirical data in constructing networks of trauma symptoms in refugee youth. *European Journal of Psychotraumatology*, *12*(1), 1920200.

APPENDICES

Acknowledgements

Acknowledgments

Working on my PhD has been a challenging but also very rewarding journey. I would have not been able to complete it without the support and guidance of my supervision team, many collaborators and mentors, my family, my friends. I want to thank all of you for guiding me scientifically, supporting me personally, and having my back in all the ups and downs of this journey.

First and foremost, Sacha, thank you for starting this journey with me. I could not have wished for a better mentor to guide me through writing the proposal for this PhD. From very early on, you encouraged me to go to conferences and present my work. At the time this was very scary, but in hindsight, I appreciate your encouragement and trust so much. Harriette, thanks for being a fantastic promotor. You always had a clear vision and my best interest at heart, and helped me get this PhD on track. Date, I learnt a ton from your clinical expertise. Our discussions were really formative for the direction of this PhD. Robert, thank you for your guidance in navigating this project. I really appreciate that I was able to work in such an interdisciplinary team.

I want to thank my family who is always here for me, and has always supported anything I wanted to do in my life. Mama und Papa, ich bin so dankbar, dass ihr mir all das ermöglicht habt. Lisa, Leo, Mateo, Maja, Oma Sig und Oma Andernach, danke für all die Unterstützung, ihr seid mein Fundament.

I have had the privilege to work with great collaborators who also became important mentors in my journey. Don, thank you for all your guidance and mentorship during my research visit in Boston. Your thinking was really inspiring in conceptualizing this line of work, and I am so glad I got to work with you. Denny, I cannot appreciate enough that I was able to work in your lab for my PhD. Your vision and passion for systems and networks was a huge inspiration for this project. Tessa, thank you for being a great colleague, collaborator, and mentor. Our coffee walks were immensely helpful in shaping my ideas of what I want in research and for my own future. Lars, working with you on the different versions of the PECAN has been so much fun and so inspiring. I really appreciate your thoughtfulness, and hope to visit you soon again in Stockholm. Saskia, I am so glad we started working together. Your strong clinical point of view really shaped my work, and I am so grateful for all the clinical mentorship you have given me. Eiko, getting to work with you was definitely one of the highlights of my PhD. You are one of the kindest and most inspiring researchers I met, and I am so grateful for your mentorship. Ben, you have had a huge impact on navigating my journey to and in New York. I cannot thank you enough for being my LGBTQ mental health research mentor, and a great friend.

Thank you Ria, Karoline, Maarten van den Ende, Gaby, Jill, Fabian, for being great office mates. Thanks also to all others in the department, Adam, Adela, Alessandra, Jonas H, Jonas D, Lourens, Maarten Marsman, Marie, Omid, Nora, Rene, Riet. Working in this environment was a huge inspiration for my journey that I cannot appreciate enough.

Thank you to my friends from home, Sop and Tüdi. I love to think back to our trip to England and all the good times we had throughout our studies. Thank you also to my Heidelberg friends, Mecky, Yasemin, Eva, Alex, Caro, Kathrin, Isi, Larissa, Dominic, David. Studying with you was one of the most beautiful chapters of my life, and the beginning of my academic journey. I love to think back to all the days spent at the Neckar, and the nights in the Untere.

I want to thank my friends in Amsterdam who have made my life there the most special time. Mirna, Loredana, Bren, Robert, I miss our evenings together, our trips to the nature (even the canoeing). Wesley, Shaakira, my closest improv friends, you helped me get to know a different side of myself and to let go on stage. Stijn and Bart, my loves, the Regina and Karen to my Cady, you have brought so much joy to my life in this past year, and I miss you a lot. Thanks also to all the BB boys/H11 from Netzo: Dani, Alberto, Frank, Miko, Pablo, Remek, Zarko, Imran, Adriano, Jaka, and of course Bert. This community has given me so much and made leaving Amsterdam just really so difficult.

Murat, I am so grateful to have you as my therapist. This journey has been shaped so much by our sessions over the past year-and-a-half, and you helped me grow into the person I am today.

Ivo, my dear Ivo. You are easily one of the most special people I got to meet this year, and I am so lucky to have you in my life. I know I can count on you with anything, and that you will always understand me. I love you, and I cannot wait to be reunited with you again.

Vanessa, ich denke nicht, dass ich jemals einen Menschen getroffen habe, für den Loyalität an höherer Stelle steht als für dich. Dass ich dich in meinem engsten Freundeskreis habe, seit nun mehr als 10 Jahren, vom Physik-Karussell bis hin zum Ende meines PhDs, ist wohl mitunter einer der größten Bereicherungen für diesen Weg gewesen. Ich freue mich auf noch viel mehr Podcasts, und beste-Wochen-deines-Lebens.

Aiden, the final and biggest thank you goes to you. You have been the most important person and biggest support for me in this journey. You were there for me at any step of the way – always, unconditionally, during the best and the most difficult times. I love to think back to our many trips, spending hours in the bath dissecting life, being with your family in England, even our hours-long pandemic walks that we got so sick of. The past years with you have been the most beautiful time. Nothing can take this time away from us, and I will always hold it very close to my heart.

New York City, September 2023