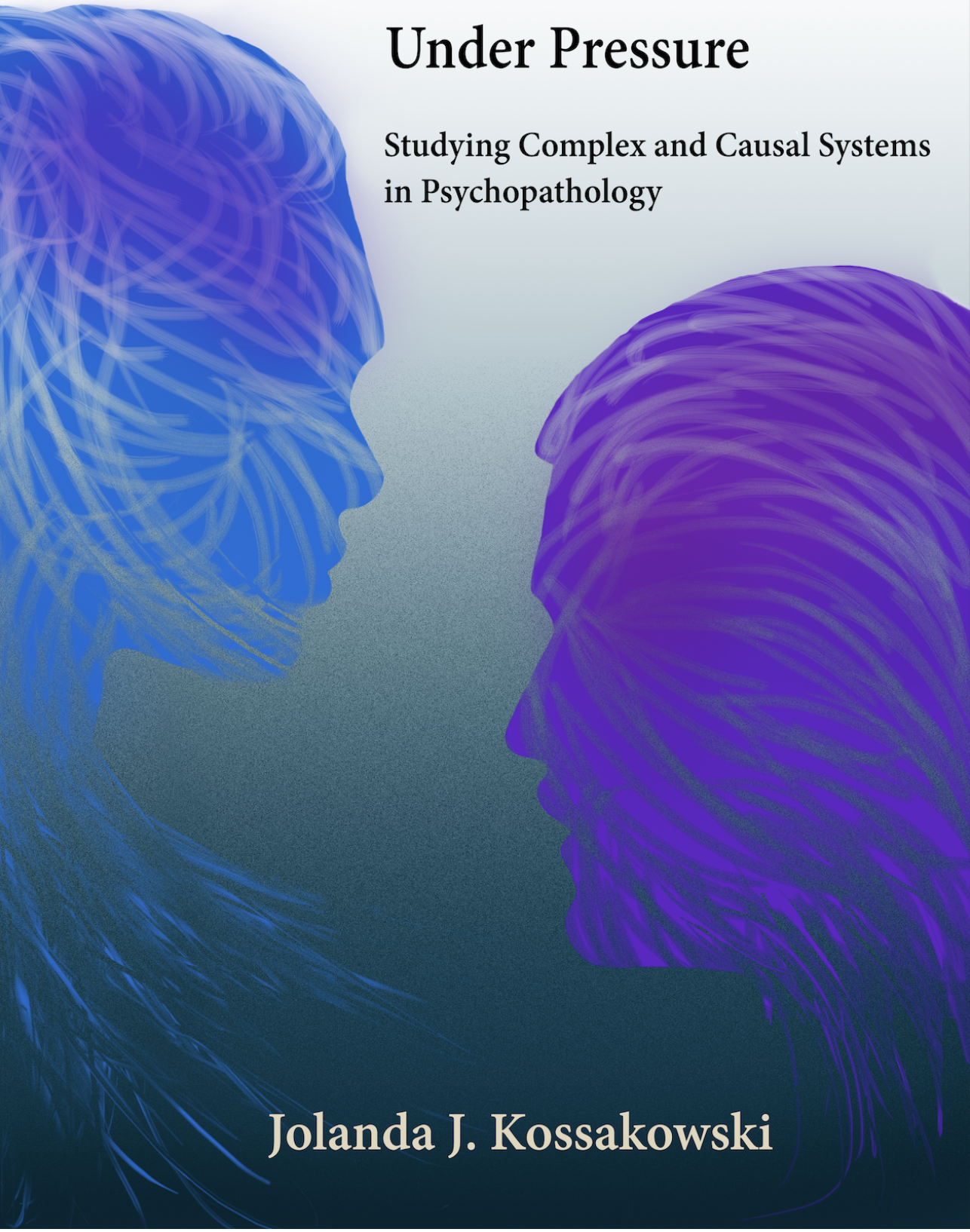
Under Pressure

Studying Complex and Causal Systems in Psychopathology



Jolanda J. Kossakowski

Under Pressure:

Studying Complex and Causal Systems in Psychopathology

Jolanda Jacqueline Kossakowski

© 2020 Jolanda Jacqueline Kossakowski
All rights reserved.

The research in this dissertation was supported by Research Priority Area Yield and an European Research Consolidator Grant received by Prof. Denny Borsboom (grant no. 647209).

Cover design: Brenda van Loenen, https://www.bvlateliers.nl
Printed by: Ipskamp Drukkers
Digital copy: dare.uva.nl

Under Pressure: Studying Complex and Causal Systems in Psychopathology

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. K.I.J. Maex

ten overstaan van een door het College voor Promoties ingestelde commissie

in het openbaar te verdedigen in de Agnietenkapel

op donderdag 14 mei 2020, te 14.00 uur

door Jolanda Jacqueline Kossakowski

geboren te Terneuzen

Promotiecommissie

Promotor:	Prof. Dr. H. L. J. van der Maas	Universiteit van Amsterdam
Promotor:	Dr. L. J. Waldorp	Universiteit van Amsterdam
Overige leden:	Prof. Dr. D. Borsboom	Universiteit van Amsterdam
	Dr. A. O. J. Cramer	Tilburg University
	Prof. Dr. H. M. Huizenga	Universiteit van Amsterdam
	Dr. M. Marsman	Universiteit van Amsterdam
	Prof. Dr. M. C. Wichers	Rijksuniversiteit Groningen
Faculteit:	Faculteit der Maatschappij- en Gedragswetenschappen	

Contents

1

Introduction

Let's play a game. Imagine a board filled with white squares, like the one you see in Figure 1.1 (upper left panel). This board contains nine white cells, and each cell is connected to its neighbouring cells. All cells are in this stage white, indicating that they are dead cells. To start the game, I will randomly select a few cells and bring these to life (denoted by black), as is shown in the upper right panel of Figure 1.1. The rules to the game are pretty simple: (1) a living cell (black) that has less than two live neighbours will die. (2), a living cell with two or three live neighbours will live. (3), a living cell that has more than three living neighbours will die. Lastly (4), a dead cell that has exactly three living neighbours will be brought to life (Gardner, 1970). The lower left and right panels of Figure 1.1 display the first two time steps after initiation. This game is one where the player has a passive role. With the four simple rules I explained earlier, the board may continue to update indefinitely and is thus self-organising. I will leave it to the reader to continue this game; I have placed some empty boards at the end of this dissertation that can be used to play this game. The game can end up in one state where no cells will be brought to life, or die. It is also possible that the board switches (oscillates) between two patterns of living and dead cells (Rendell, 2011).

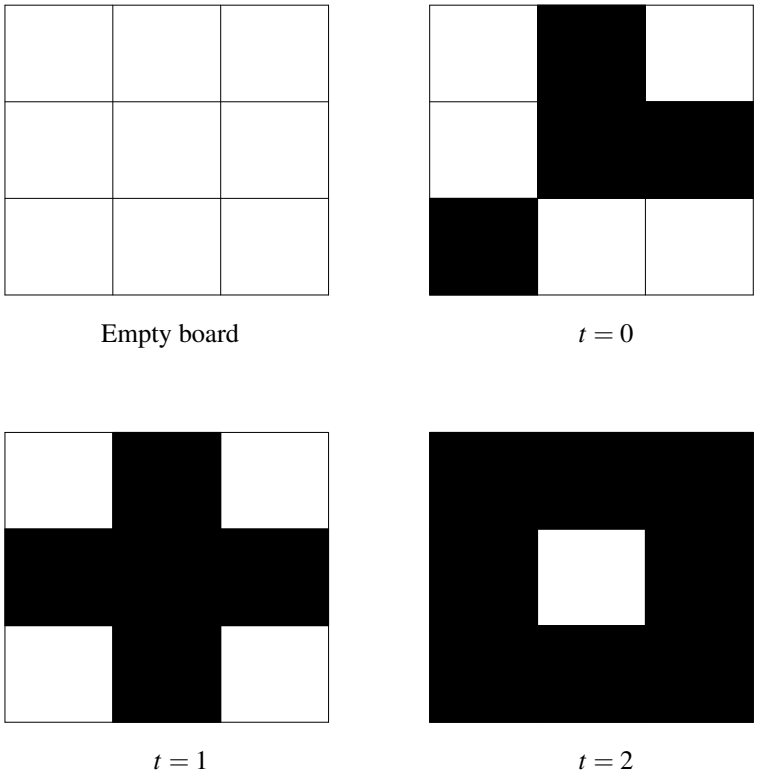


Figure 1.1: An illustration of Conway's game of life (Gardner, 1970). White cells indicate dead cells, and black cells indicate living cells. Top left = empty board with nine dead cells. top right = board at $t = 0$, where four cells were brought to life. Bottom left = board at $t = 1$ with five living cells. Bottom right = board at $t = 2$ with eight living cells.

1.1 Dynamical systems

The small game that I described here is an example of Conway's *game of life* (Gardner, 1970). The game of life reflects real life processes, and their unpredictability. It is also an example of a *dynamical system*: a system in which its components interact with one another and that updates throughout time depending on its earlier states. A dynamical system can best be explained by looking at the definitions of the words 'dynamic' and 'system'. A *system* is the topic of interest, and comprises various components that can interact with each other. Examples of systems that have been studied range from planetary motions (Alligood et al., 1996), to ecosystems (Scheffer et al., 2001), to mental disorders (Kossakowski & Cramer, 2019). A system does not necessarily have to be a topic that is found in the real world. Conway's game of life is also an example of a system, where the cells are the components of the system that can interact with each other. The second word, *dynamic*, is another word for 'change'. This means that we are looking at something that changes or evolves over time (Strogatz, 1994; Hasselblatt & Katok, 2003). The time aspect of a dynamical system can be handled in two ways: continuous or discrete. In a *continuous* dynamical system, time is indexed by a continuous variable than can take any real value. Thus, a continuous dynamical system allows for different intervals between measurements of the dynamical system (Voelkle et al., 2012). In contrast, a *discrete* dynamical system is one where time is explicitly handled and is part of the mathematical model (Holmgren, 1996). Most models that work with discrete time assume that the intervals between the measurements of the dynamical system are equal (Bringmann et al., 2013; Epskamp et al., 2018). In Chapters 3 and 4 I will show that this assumption does not necessarily have to be met in order to accurately interpret the discrete dynamical system. For the remainder of this dissertation, all dynamical systems that will be described are discrete dynamical systems.

One interesting property of a dynamical system is that there are multiple stable states that the system can be in. In such cases, the dynamical system will eventually be in a stable state, or it will be attracted to such a state. In this dissertation we only consider

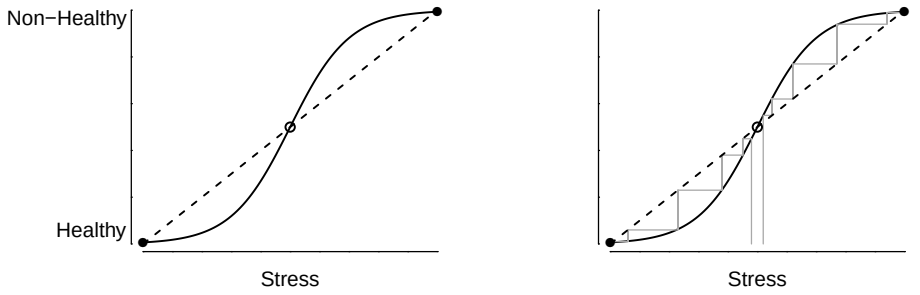


Figure 1.2: Visualisation of attractor points and unstable points. Solid dots = attractor points. Open dot: unstable point. Grey line = path of the system at two different stress values towards an attractor point.

dynamical systems that have two stable states. A visualisation of the attractor points to which a system is drawn to is shown in Figure 1.2 (left panel), where the black solid dots are so-called *attractor points*: these represent the two stable states that the system is attracted to. The open dot in the middle denotes an unstable point that drives the system away. To illustrate, say that we have two stable states (healthy and non-healthy) in a dynamical system that represents major depressive disorder (MDD). No matter the level of stress that we introduce to the system, it will always travel to one of either stable states, even if we come really close to the unstable point. This is shown in the right panel of Figure 1.2.

The way a dynamical system travels from one stable state to another can be either continuously or discontinuously. When changes take place *continuously*, the dynamical system follows a continuous trend to arrive at the stable state. To illustrate, we again look at the dynamical system of MDD that we saw earlier. This exemplary dynamical system has two stable states: healthy and non-healthy. The continuous evolution of the dynamical system is depicted in the left panel of Figure 1.3. In this illustration, an individual whose development of MDD can be represented with such a dynamical system may slowly develop the disorder as time progresses. Some may argue that the development of MDD follows this continuous trend. However, I believe that this is not the case.

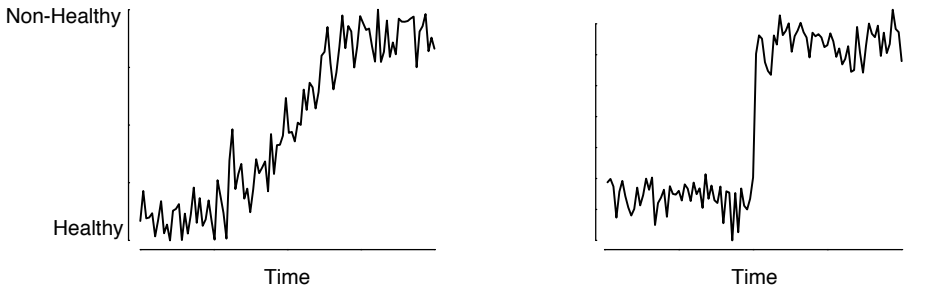


Figure 1.3: An example of a continuous (linear) dynamical system (left panel) and a discontinuous (non-linear) dynamical system (right panel).

I think that it is hard, if not impossible, to find a real-world example in which its development follows a continuous trend. I believe that mental disorders such as MDD develop in a more *discontinuous* manner. In a discontinuous setting, the system travels from one stable state to the other by means of a sudden jump, which we call a *critical transition* (Kuznetsov, 2013). An example of this is shown in the right panel of Figure 1.3. In this illustration, an individual whose development of MDD can be represented with such a dynamical system may at first not experience many MDD symptoms. However, as time progresses, some external event may occur (e.g., the loss of a spouse or getting fired from one's job), and the dynamical system jumps from the healthy to the non-healthy state. It is much easier to find real-world examples of dynamical systems that develop according to this discontinuous trend (also termed a non-linear dynamical system). Real-world examples include MDD (Cramer et al., 2016), development (Emde & Harmon, 1984), but also ecosystems (Scheffer et al., 2001), earthquakes (Sornette, 2002) and even arrhythmia (Owis et al., 2002). These non-linear dynamical systems do not only exist in physics or ecology, but also in psychology. Chapter 4 will describe a non-linear dynamical system for psychopathology in more detail.

When a dynamical system is more likely to experience a transition, it approaches a *tipping point*. The closer a system is to a tipping point, the more likely it is that a transition from one stable state to another takes place. A metaphor that is often used to

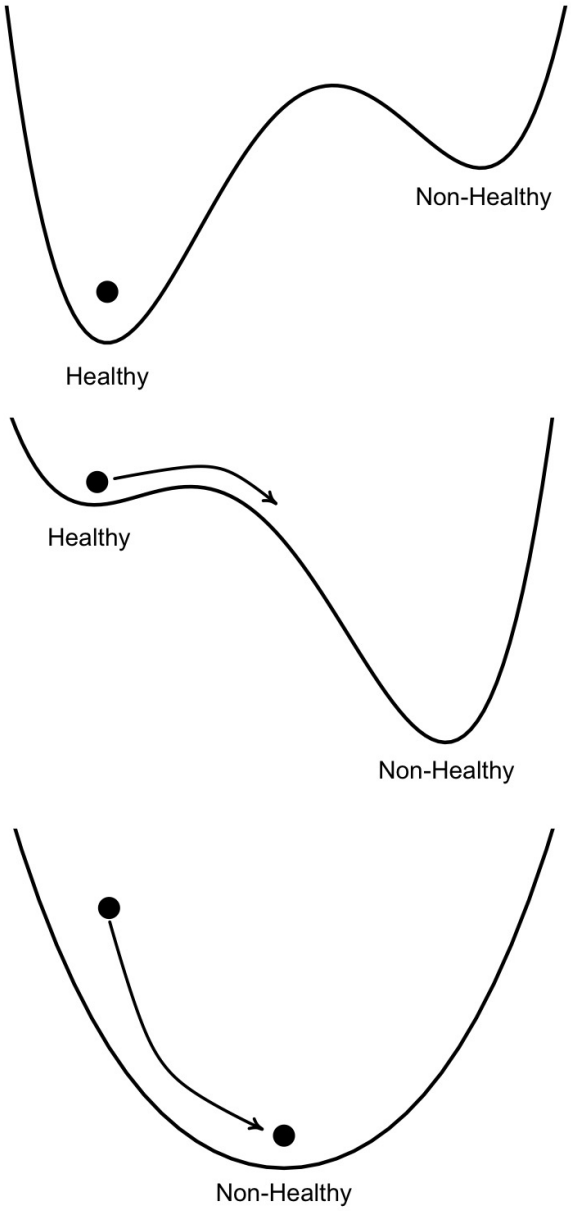


Figure 1.4: Visualisation of the landscape when a dynamical system is in one stable system (upper panel), when a dynamical system approaches a tipping point (middle panel), and when a dynamical system experienced a transition into the other stable state (lower panel).

explain tipping points is the "ball-in-a-cup" metaphor (van de Leemput et al., 2014). An illustration of the idea behind tipping points is shown in Figure 1.4. Here we have three situations. In the first (upper panel), the system - visualised by the black ball - is in a stable state ("healthy"), and a large push is needed to bring the system to the other stable state ("non-healthy"), as exemplified by the depths of the cups. In the second situation (middle panel), the system is approaching a tipping point. The cup that analogous to the healthy stable state is more shallow, and thus it is easier to push the ball from this state to the non-healthy stable state. In the third situation (lower panel), we see a similar landscape as in the first situation, but here the system has reached the tipping point and transitioned from the healthy stable state to the non-healthy stable state. It is shown here that a large push, larger than in the first situation, is needed to push the system out of this stable state.

1.2 Causality

Systems consist of various components that interact with one another. While these interactions are often reciprocal (i.e., X influences Y and Y influences X in a similar manner), There are also cases in which these interactions work as a one-way traffic street. Here, X influences Y, but not the other way around. This is an example of a *causal* relation between two variables. Causality can be seen in many different lights, and therefore it is difficult to set a definition of a causal relation that everyone agrees with. It is beyond the scope of this dissertation to summarise the discussion, but I do want to provide the reader with some background knowledge on causality.

A causal relation from one variable to another indicates that this relation is not reciprocal (Simon, 1952). The one-way traffic street is a good metaphor for this assumption: causal influence can only flow in one direction. A causal relation also contains a time aspect (Granger, 1980): the past and present may influence the future, but the future cannot influence the present and past. A causal relation can be *deterministic* or *probabilistic*. With the former, when we use the example $X \longrightarrow Y$, when X occurs, Y must occur as

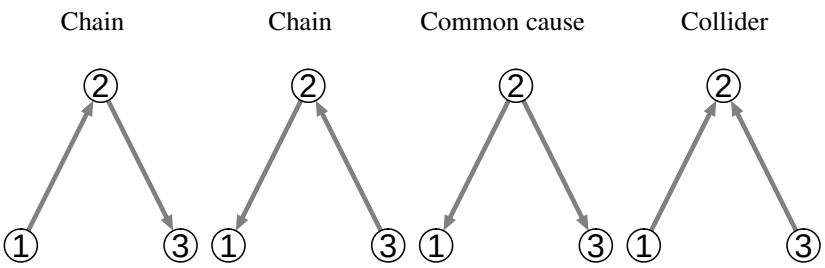


Figure 1.5: The different causal structures that can be detected with (conditional) dependencies and independencies. The chain structures and the common cause structure are statistically equivalent, whereas the collider structure is statistically unique.

well, and when Y is observed, we know that X occurred as well (Granger, 1980). In a probabilistic causal relation, these two cases do not apply. In Chapter 5, I provide an example of probabilistic causality. There, I define a causal relation as one where, when we manipulate the cause, a change must occur in its effect (given all other variables). The change here is probabilistic, as the amount of change may vary. Due to the asymmetry of causality, this definition also implies that, when no change is observed in the effect, no manipulation has taken place on its cause. Pearl (2009), amongst others, calls this a *counterfactual* definition of causality.

An approach that is often used to estimate causal relations is the one developed by Pearl (2009) and Spirtes et al. (2000). Here, we use the notion of (conditional) dependence and independence between sets of variables to estimate a causal relation. By calculating raw and partial correlations, we can draw four different causal structures, depicted in Figure 1.5. The first three structures, the two chain structures and the common cause structure, are statistically equivalent. In these structures, variables 1 and 3 have a nonzero raw correlation, but when we condition on variable 2, the partial correlation becomes zero. As this applies to all three structures the same way, we cannot distinguish them statistically. The raw/partial correlation combination is reverse for the fourth causal structure, the collider. In this structure, the raw correlation between variables 1 and 3 is zero, but the partial correlation when conditioning on variable 2 becomes nonzero. In the approach

by Pearl (2009) and Spirtes et al. (2000), the collider structures are first identified, as they are statistically unique. The "blanks" are filled in with the information from the other three structures.

1.3 Networks

Since we mostly deal with a set of variables, we are also interested in the interactions among these variables. Network theory gives the opportunity to study these interactions. In recent years, the notion has been put forward that mental disorders (such as MDD) can be presented as a system of mutually interacting variables. Work by van der Maas et al. (2006), Borsboom (2008; 2011; 2017), Cramer (2010; 2012), Epskamp (2012), Bringmann (2013) and many others have paved the way for researchers to take a different perspective not only on mental disorders, but also personality (Cramer et al., 2012, Costantini et al., 2015), attitudes (Dalege et al., 2016), sleep disorder (Blanken et al., 2019), health-related quality of life (Kossakowski et al., 2016) and even diagnostic manuals (Tio et al., 2016).

The left panel of Figure 1.6 depicts a basic example of a network. A network consists of two elements: nodes (circles; observed variables) and edges (lines; relations between variables; Newman, 2010). In this example, the nodes represent people, and the edges an existing connection between two people, indicating that they know each other. This type of network is also known as a social network. Here, the edges are unweighted, meaning that the extent to which two people know each other is not operationalised, and undirected, meaning that a connection between two people is reciprocal. The network on the right is adapted from Kossakowski et al. (2016) and visualises the pairwise interactions between items of the Short Form Health Survey (SF-36; Ware Jr & Sherbourne, 1992). This network is also undirected, but it is no longer unweighted. The colour and thickness of the edges represent the strength of the connection between two items, controlling for all other items in the network. Another striking difference is related to the source of the edges. In the network in the left panel of Figure 1.6, the edges are known and observable:

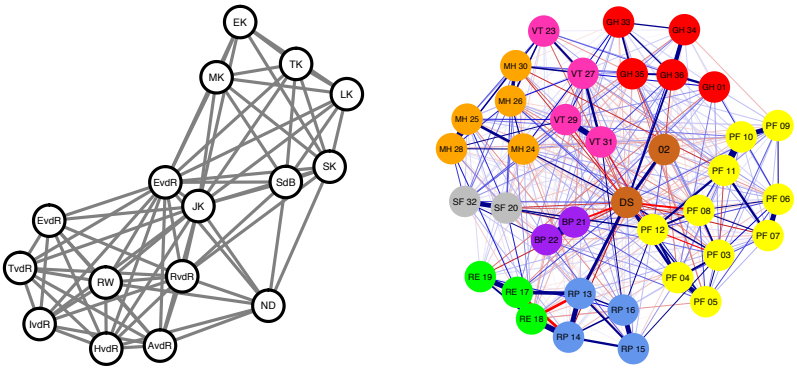


Figure 1.6: Illustration of a social network (left panel) and a network of Health-Related Quality of Life as measured by the Short Form Health Survey (SF-36) in a combined sample of cancer patients and healthy controls (right panel). Blue edges denote positive connections, and red edges negative connections. The strength of a pairwise connection between two items is represented with the colour and thickness of an edge. Node colours correspond to the eight domains (see Kossakowski et al., 2016 for more details).

we can draw up a social network by simply asking people if they know one another. In the network in the right panel of Figure 1.6, one cannot observe the connection between two items; we have to estimate these. There are many different approaches that one can take in estimating a network. A few of these techniques are discussed in Chapter 2.

The networks that we saw earlier were both undirected: the edge between two nodes denotes a reciprocal relation. The edges in a network can also have an arrow, creating directed or causal networks. Directed networks are often used to visualise a set of causal relations. Depending on the model that is used, a causal relation in a directed network may mean something different. Examples of causal networks include citation networks (Newman, 2010), where an arrow shows that one paper cited another. A more abstract example is a cross-lagged panel model (Hamaker et al., 2015) or a structural equation model (Bagozzi & Yi, 1988; Edwards & Bagozzi, 2000), where the arrows indicate how

two variables are related to one another. Chapters 5 and 6 will focus on directed networks and how to estimate them.

1.4 Chapter Outline

The topics that I have discussed here will help the reader to understand the upcoming chapters. I have focused on two subjects that are the subject of this dissertation: complex (dynamical) systems and causal networks. Part one of this dissertation includes three chapters that focus on complex systems in psychopathology. Chapter 2 describes various techniques for estimating networks. It also elaborates on two models that can be used to study complex systems, one of which is based on the work by Cramer et al. (2016). The second model discussed here is the mean field model. With this model we reduce a complex and multidimensional system to a single equation, after which we can study its dynamics. This model is mathematically explained in Chapter 3. By means of a simulation study, we show that reducing a complex system does not result in a reduction in accuracy in terms of assessing the dynamics of the system. In Chapter 4 we show how we can use maximum likelihood estimation, together with the mean field model, to assess the expectancy that an individual may have in transitioning from one stable state to another. This approach is evaluated numerically, after which it is applied to two empirical examples.

Part two of this dissertation includes two chapters that discuss causal networks. Chapter 5 discusses the advantage of combining observational and experimental data when estimating a causal network. Using observational data (where no perturbations have taken place) will not give us the complete causal picture, as some causal patterns are indistinguishable. By adding experimental data (where perturbations did take place), one can separate these causal patterns and thus give a complete overview of a causal system. Several algorithms are discussed and illustrated, after which their performance is evaluated in a simulation study. Two algorithms specifically show great promise when it comes to estimating a causal network. In Chapter 6, we illustrate the potential of these algorithms

by applying them to a dataset of patients who entered treatment for obsessive-compulsive disorder.

This dissertation ends with a discussion on the topics that have been presented. Here, I discuss to what extent the models, as they are presented here, are suitable to be used by other researchers, and what steps we have to take to accomplish this. I elaborate on possible extensions of the models, and what we as a field need to do to move forward.

Part I

Complex Systems in Psychopathology

2

Complexity, Chaos and Catastrophe: Modeling Psychopathology as a Dynamic System

This chapter is published as: Kossakowski, J. J., & Cramer, A. O. J. (2019). Complexity, chaos and catastrophe: Modeling psychopathology as a dynamic system. In M. S. Vitevitch (Eds.), *Network Science in Cognitive Psychology* (pp. 45-79). New York, USA: Routledge.

Abstract

The network perspective is a young competitor in the field of psychopathology. It postulates that psychological constructs, such as mental disorders, arise as a system of mutually interacting variables. Its rise started in 2006 (van der Maas et al., 2006), and has since then been applied to various topics. In this chapter we give an overview of various approaches that can be taken to study psychological constructs as networks. We start by describing three techniques with which networks can be estimated. These techniques differ in the type of item responses (binary, continuous, or a mixture) they can handle, and how they deal with possible spurious connections. Each network estimation technique is illustrated using a freely available dataset (Kossakowski et al., 2017). Two different models for modelling complex dynamical systems are then explained: the Cramer model and the Empirical Mean Field Approximation. These models are designed to model systems that can make a sudden transition between two stable (mood) states. The usefulness of these models are shown with their own empirical examples. Even though each model has its own challenges that need to be overcome, both models show great potential for future use.

2.1 Introduction

In recent years, the notion that correlations between questionnaire items (e.g., insomnia and fatigue; hereafter called: variables) are the result of direct interactions between these variables (e.g., insomnia $\rightarrow$ fatigue) has grown in popularity. The rise of this network perspective started in 2006, when van der Maas et al. (2006) demonstrated that the positive manifold, the phenomenon that scores on cognitive tasks (e.g., verbal comprehension and working memory) are positively correlated, can be explained by means of a network approach. In this groundbreaking study, van der Maas et al. argued that, instead of explaining the positive manifold by means of a latent (i.e., unobserved) variable called g (i.e., general intelligence), the positive manifold can be explained by means of a mutualism model, in which variables have mutual, reinforcing, relations (e.g., verbal comprehension $\rightarrow$ working memory). Figure 2.1 shows an example of both the latent variable and the

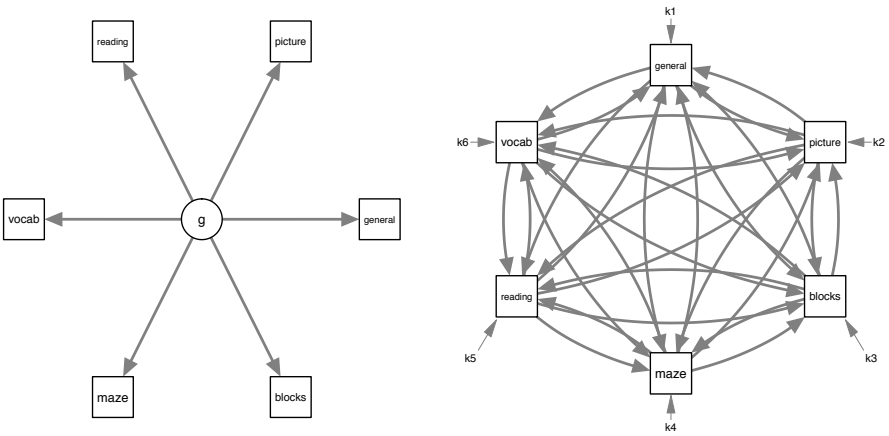


Figure 2.1: A simplified example of a latent variable model (left) and the mutualism model (right). $x1 \ldots x6$ represent variables, and $k1 \ldots k6$ environmental resources. Figure is adapted from van der Maas et al. (2006).

mutualism model (data used for this example are freely available in the statistical software program R; R Core Team, 2016). Here, we have six variables (denoted by squares) that represent six test batteries of an intelligence test: Cattell's culture-fair test (general), vocabulary (vocab), reading comprehension (reading), maze (maze), block design (blocks) and a picture completion test (picture). In the left panel, the latent variable example, the latent variable g is the cause of the six variables, whereas in the mutualism model (right panel of Figure 2.1), each of the six variables has direct and mutual relations with every other variable in the model. Note that $k1 \ldots k6$ represent environmental resources (such as the parents' level of education or a child's age) that may influence individual variables. Van der Maas and colleagues showed that, under certain circumstances, the latent variable model and the mutualism model were statistically equivalent, even though they are conceptually very different. That is, these results told us that you do not always need a latent variable model to explain the existence of a positive manifold between psychological variables. The primary aim of the present chapter is to elaborate on this network perspective as a theory whilst also providing the reader with a set of methods (including

practical examples) that can be easily used for one's own data.

The positive manifold is not unique to the topic of intelligence. For example, symptoms of psychological disorders (e.g., symptoms of major depression) are also consistently positively correlated with one another (i.e., positive manifold; e.g., Hartman et al., 2001); the same manifold was observed for variables that pertain to dimensions of normal personality (e.g., neuroticism; Dolan et al., 2009). As such, van der Maas and colleagues (2006) paved the way for exploring the feasibility of a network perspective in other areas of psychological research: e.g., psychopathology (e.g., Borsboom et al., 2011; Cramer et al., 2010; McNally et al., 2014), personality (Cramer et al., 2012), health-related quality of life (Kossakowski et al., 2016), attitudes (Dalege et al., 2016) the International Classification of Diseases and Related Health Problems (ICD) and the Diagnostic and Statistical Manual of Mental Disorders (DSM; Tio et al., 2016). In particular, the network perspective on psychopathology (e.g., psychological disorders) has not only resulted in a more mature conceptualisation of disorders (and comorbidity between them) as networks of directly interacting symptoms (e.g., Cramer et al., 2010), it has also inspired the development of a host of new methods with which one is able to estimate networks for psychopathological data (see Fried et al., 2017 for an extensive review of the empirical psychopathological network literature).

As previously stated, many psychological constructs can be studied from a network perspective. In a network, individual variables are represented as nodes (circles); the mutual relation between two variables is visualised as a line between these two variables (hereafter called: edge; Newman, 2010). Figure 2.2 depicts two ways of constructing a network, using the same intelligence data as was used for Figure 2.1. The left panel in Figure 2.2 shows a binary network where an edge can be either present ("1") or absent ("0"). As such, all edges have the same weight (i.e., "1") and thus the same thickness and colour. The right panel features blue edges with different widths. Blue edges represent positive relations between nodes, whereas red edges, not currently shown in Figure 2.2, represent negative relations between variables/nodes (Costantini et al., 2015). Blue edges

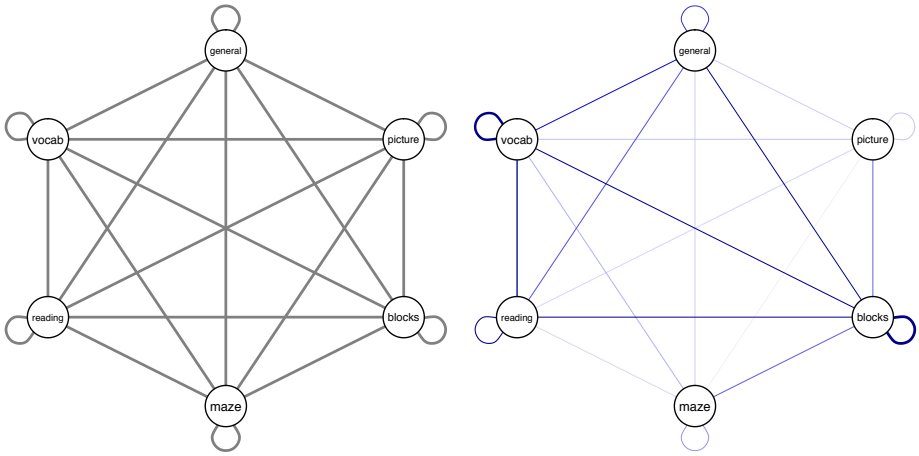


Figure 2.2: Six tests that measure different aspects of intelligence. Left = unweighted network. Right = weighted network. General = Cattell's culture-fair test. Vocab = vocabulary. Reading = reading comprehension. Maze = maze. Blocks = block design. Picture = picture completion test.

indicate that two variables reinforce each other: as one variable increases/decreases, the other increases/decreases as well. In contrast, red edges indicate that, as one variable increases/decreases, the other variable decreases/increases. The width and colour saturation of the edge denotes the strength of the relation: the more saturated and thicker the edge, the stronger the relation between two nodes (Epskamp et al., 2012). Edges do not necessarily need to connect two nodes: it is also possible for a node to reinforce itself (e.g., insomnia results in more insomnia). A so-called *self-loop* is then present in the network: an edge that starts and ends at the same node. In Figure 2.2 for example, the nodes that denote the vocabulary test and the block design test have a strong self-loop.

In networks such as the one shown in the right panel of Figure 2.2, mutual relations between variables – that is, the *connectivity* of a network – can differ. This means that some variables have a stronger or weaker relation than others. For example, in Figure 2.2, the vocabulary test (vocab) and the block design test (blocks) have a strong relation in comparison to the other relations in the network, such as the relation between the variables

reading test (reading) and the maze test (maze). This means that they, relatively speaking, have a strong influence on each other: when individuals score high on the vocabulary test, they also tend to score high on the block design test. The stronger the relations between the variables, the stronger the variables influence each other. *High connectivity* refers to a network with generally strong relations between variables, whereas *low connectivity* refers to a network with generally weak relations between variables. In other words, the higher the connectivity in a network, the stronger the relations between the variables in the network are. Considering intelligence, more strongly connected networks can be beneficial: in a strongly connected network, different aspects of an intelligence test reinforce each other more, which may lead to higher scores on an intelligence test. Thus: stronger connections may result in higher intelligence. Considering the development of psychopathology, this characteristic may not be so beneficial. For example, van Borkulo et al. (2015) showed that depressed participants, who still had the diagnosis of major depressive disorder (MDD) after two years, had a more strongly connected network than depressed participants who were in remission after two years. In this context, strong network connectivity can be seen as a measure of *vulnerability*: individuals who are more vulnerable for developing an episode of MDD tend to have a more strongly connected network than individuals who do not have this vulnerability for MDD (Cramer et al., 2016).

In this chapter, we aim at providing the reader with the basics of network modelling and applying these modelling principles to studying vulnerability in psychopathological constructs. We will start by providing an overview of various methods for estimating individual networks from empirical data (section 2.2), so that we can investigate vulnerability in psychological constructs at the level of the individual. Next, in section 2.3, we will elaborate on catastrophe theory, which is a central theory pertaining to studying vulnerability from a network perspective. Following this, we explain the Cramer model (Cramer et al., 2016) in section 2.4, accompanied by an application of this model to symptoms of mania (Kessler et al., 2014). We will proceed with the explanation of the empirical

mean field approximation (Kossakowski, Gordijn, Riese & Waldorp, 2019; Waldorp & Kossakowski, 2020) in section 2.5, which is also accompanied by an application of the model to various positive and negative affects. We will conclude this chapter by taking a critical perspective: the network approach currently faces several challenges that have not been met (yet) and are important for advancing network modelling in psychology in general, and psychopathology in particular, in the next few years.

2.2 Constructing Networks

Various methods exist for estimating the connectivity of a network (i.e., the strength of the relations between variables in a network). Networks can be estimated for an entire group, or for individuals. When one estimates a network for a group of individuals, one can draw conclusions, for example, about specific relations between variables, that may be generalisable to the population of which we drew a sample. For this type of network, we only need one measurement per individual (cross-sectional data). It is also possible to estimate an intra-individual network, based upon which one can draw conclusions that only pertain to that specific individual. For this type of network, we need several measurements of one individual (time-series data). The applicability of each estimation method described in the remainder of this section heavily relies on the type of empirical data (categorical, continuous, or a mixture of the two), the number of participants that are being measured, and the number of measurements that have been collected per participant. The methods that we elaborate on in this section can be used to estimate intra-individual networks. In this section, we will discuss three methods: 1) *IsingFit* for binary data; 2) *Graphical Vector Auto Regression* (VAR) for continuous data, and 3) *Mixed Graphical Model* (MGM) for data with both binary and continuous variables.

A general disadvantage of network estimation is that networks can have many edges: in a network with six nodes, we can estimate up to 15 unique edges. However, in a network with 25 nodes, we can estimate as much as 300 unique edges. Many of these edges can be spurious. If we look at Figure 2.2 for example, an edge is present between

the reading test (reading) and the block design test (blocks), while these two tests might in fact not be directly related at all. It could very well be that, when one scores high on the reading test, one also scores high on the vocabulary test (vocab), and as a result, also scores high on the block design test. Thus, it is possible that the relation between the reading test and the block design test is 'caused' by the vocabulary test.

All three methods control for such potentially spurious edges by using the *least absolute shrinkage and selection operator* (LASSO; Tibshirani, 1996). The LASSO imposes a parameter *lambda* (λ) that controls the level of sparsity (a sparse network is one with relatively few connections): a high λ will result in many weak edges that will be removed from the network, while a low λ will result in fewer weak edges that will be removed from the network. Choosing this parameter is therefore not trivial as it directly influences the structure of the resulting network. All methods that will be discussed estimate the network using different values for λ (Zhao & Yu, 2006). Then, the most optimal λ parameter is selected using the *Extended Bayesian Information Criterion* (EBIC; Chen & Chen, 2008; Foygel & Drton, 2010), where the network model with the lowest EBIC is chosen. The EBIC uses a parameter of its own *gamma* (γ), that controls whether or not the EBIC prefers simpler models, where simpler indicates a sparse network with fewer edges. A high γ parameter ensures that EBIC prefers the sparse network model with a high λ parameter and fewer edges. The hyperparameter γ ranges between 0 and 1.5 and needs to be manually set. For more information about this method for controlling sparsity, see Epskamp (2016) and Epskamp & Fried (2018).

Each method will be demonstrated with the same empirical dataset. These data are time-intensive measurements of one 57-year old male with a history of depression. Over the course of 239 days, this participant monitored his daily life experiences and affects several times a day, resulting in 1479 time points (for details about these data see Kossakowski et al. 2017; Wichers et al. 2016). Table 2.1 shows the items, their meanings, their intended ranges and the response ranges.

Item labels	Item meaning	Item range	Range of responses
1	I feel relaxed	1 to 7	1 to 7
2	I feel down	-3 to +3	-3 to +3
3	I feel irritated	1 to 7	1 to 7
4	I feel satisfied	1 to 7	1 to 6
5	I feel lonely	-3 to +3	-3 to +3
6	I feel anxious	-3 to +3	-3 to +1
7	I feel enthusiastic	1 to 7	1 to 6
8	I feel suspicious	1 to 7	1 to 5
9	I feel cheerful	1 to 7	1 to 6
10	I feel guilty	-3 to +3	-3 to +2
11	I feel indecisive	1 to 7	1 to 5
12	I feel strong	1 to 7	1 to 6
13	I feel restless	1 to 7	1 to 6
14	I feel agitated	1 to 7	1 to 6
15	I worry	1 to 7	1 to 6
16	I can concentrate well	1 to 7	1 to 6
17	I like myself	1 to 7	1 to 6
18	I am ashamed of myself	1 to 7	1 to 6
19	I doubt myself	1 to 7	1 to 5
20	I can handle anything	1 to 7	1 to 6
21	I am hungry	1 to 7	1 to 4
22	I am tired	1 to 7	1 to 5
23	I am in pain	1 to 7	1 to 3
24	I feel dizzy	1 to 7	1 to 3
25	I have a dry mouth	1 to 7	1 to 2
26	I feel nauseous	1 to 7	1 to 3
27	I have a headache	1 to 7	1 to 4
28	I am sleepy	1 to 7	1 to 7

Table 2.1: Items that were included in the analysis and their assigned item labels.

2.2.1 Binary Data: IsingFit

The IsingFit method is based on the Ising model (Ising, 1925). This model originated in statistical physics and was developed to model ferromagnetic materials, such as the configuration of atoms and their corresponding spins. Nodes in an Ising model can only be in one out of two possible states (i.e., 'active' or 'inactive'). Furthermore, only pairwise interactions are allowed in an Ising model, which means that only interactions between two nodes are allowed, and not, for example, three-way interactions (for example, the interaction between 'vocab' and 'blocks' in Figure 2.2 cannot be positive when 'reading' is active, and negative when 'reading' is inactive). The Ising model consists of two node-specific parameters: an *interaction parameter*, which denotes the strength of the interaction between two nodes and thus the edge weight, and a *threshold parameter* that represents the preference of a node to be either active or inactive regardless of the activity of its neighbouring nodes (a neighbour is defined as all nodes that a particular node has a connection with; van Borkulo et al., 2014). A positive threshold corresponds to a preference of being active, whereas a negative threshold corresponds to a preference of being inactive. A threshold parameter of exactly 0 corresponds to having no preference.

Essentially, the IsingFit method regresses one node on all other nodes in an iterative manner, using the optimal penalty parameter λ. This means that twice as many regressions will be performed as there are nodes in the network. For example, when we consider Figure 2.2, the vocab node will be regressed on reading, maze, blocks, picture and general, reading will be regressed on vocab, maze, blocks, picture and general, and so on. After this node-wise regression, we have an interaction parameter from vocab to reading, and from reading to vocab. In order to calculate the final edge weight, the so-called AND-rule is used: only when both interaction parameters are nonzero will an edge be drawn between the two nodes, with an edge weight that is the mean of these two interaction parameters. It is also possible to use the OR-rule, which states that at least one of the two parameters must be nonzero in order for an edge to be drawn between the two nodes. This edge will then have the weight that is associated with the nonzero interaction

parameter. For clarity of presentation, all networks in this chapter will be estimated using the AND-rule.

Figure 2.3 demonstrates the IsingFit method. To improve visual comparison, the layout of the networks is fixed. In the left panel, a network with a γ parameter of 0 is estimated, while in the right panel, a network with a γ parameter of 1 is estimated. It can be seen that item 12 (I feel strong) has many negative interactions with the other variables. This means that, as item 12 increases, items such as item 1 (I feel relaxed), 16 (I can concentrate well) and 26 (I have a headache) decrease, and vice versa. The two networks depicted in Figure 2.3 demonstrate why setting the γ parameter is not trivial: the network in the left panel (estimated with $\gamma = 0$) has (somewhat) more edges than the network in the right panel (estimated with $\gamma = 1$). More specifically, in the network in the left panel, approximately 34% of the edges are nonzero, whereas in the network in the right panel, approximately 30% of the edges are nonzero. This difference in networks estimated with different values for γ will decrease as the sample size increases.

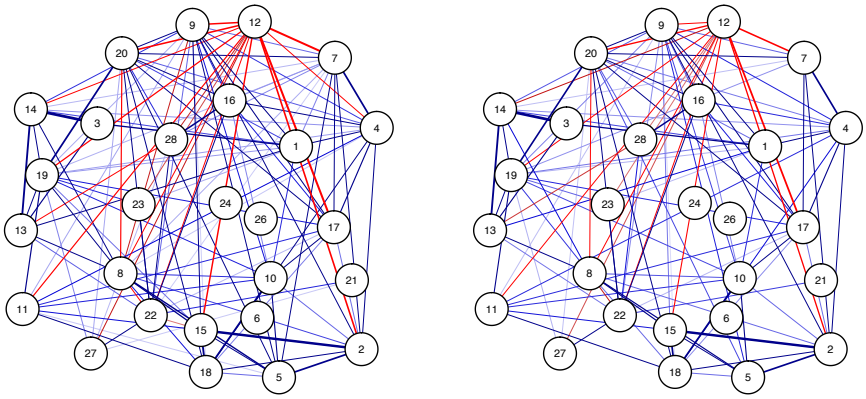


Figure 2.3: Example networks estimated with the IsingFit estimation procedure. Left panel: network structure estimated with $\gamma = 0$. Right panel: network structure estimated with $\gamma = 1$.

The IsingFit method is initially intended for cross-sectional data. However, the pro-

cedure is also suitable for time-series data if we assume that all measurements are independent of one another. This means that the scores on measurement t do not rely on the scores that were obtained at measurement $t-1$. To illustrate, suppose an individual completes the same questionnaire twice. The assumption then states that the answers given the second time are not (partially) determined by the answers given the first time. It is debatable whether this assumption holds in time-series data, but it is beyond the scope of this chapter to go into further detail.

2.2.2 Continuous Data: Graphical Vector Auto Regression

When dealing with continuous time-series data, we can estimate two types of networks: a contemporaneous network that shows the relations between variables that occur within the same time point, and a temporal network that shows the relations between variables that occur across time points (Epskamp, Waldorp, Mottus & Borsboom, 2018). For example, Figure 2.4 shows hypothetical examples of these two types of networks for affect variables: feeling distressed, active, interested, and alert. The left panel presents a contemporaneous network, which shows that feeling active and feeling interested have a negative relation. This means that, as one feels more active, one will feel less interested *within the same time point*, and vice versa. In a contemporaneous network, relations are mutualistic (i.e., no arrows) because these relations unfold within the same time point. The right panel of Figure 2.4 presents a temporal network. This network shows that feeling distressed positively influences feeling alert at the next time point, but that feeling alert negatively influences feeling distressed *at the next time point*.

Both networks are equally important and interesting: when collecting time-series data, one is often interested in the progression of an individual throughout time. At the same time, important associations between variables may occur within a single time point that are not captured when solely considering temporal relations between variables. Figure 2.4 shows this: fictitious relations are present between feeling distressed, active, interested, and alert. Here, we see that a positive, mutual relation between feeling alert

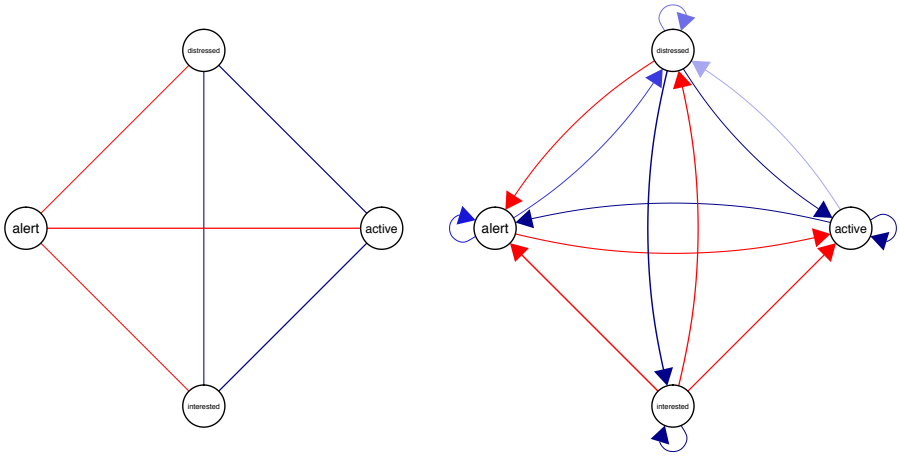


Figure 2.4: Simplified examples of the contemporaneous network (left panel) and the temporal network (right panel) estimated with the graphical VAR estimation procedure. Node labels (in clockwise direction): distressed, active, interested, alert.

and feeling active exists in the contemporaneous network (left panel), which means that, as one feels more active, one also feels more alert and vice versa. At the same time, in the temporal network (right panel), feeling active negatively influences feeling alert, which means that, as one feels more active, one will feel less alert at the next time point. Also note that there is a negative mutual relation between feeling distressed and feeling active in the contemporaneous network, and that the temporal network shows that this relation also exists over time: feeling distressed negatively influences feeling active over time, and vice versa. As Aristotle once said: "the sum is more than its parts"; estimating both a contemporaneous and a temporal network will give us a more complete picture of the progression of a construct, such as a psychological disorder or intelligence, and will yield more information than focusing on either one of the two networks (Epskamp et al., 2018).

We can estimate both these network structures using Graphical Vector Auto Regression (VAR; Epskamp et al., 2018). This method uses complex matrix algebra to estimate the elements of a matrix containing the contemporaneous associations between the items and the elements of a matrix containing the temporal relations between the items (Wild

et al., 2010). It is beyond the scope of this chapter to go into depth with respect to the estimation of the contemporaneous and temporal relations. For a more detailed and technical explanation, see Epskamp et al., (2018) and Wild et al. (2010). Figure 2.5 visualises the essence of the Graphical VAR method: For each node, we regress it on itself at the following time point $t+1$, and we regress it on all other items at the following time point $t+1$. In the example shown in Figure 2.5, node S at time point $t+1$ is regressed on itself and node C, both at time point t. The procedure is repeated for node C (at time point $t+1$), which is regressed on itself and node S at the previous time point t.

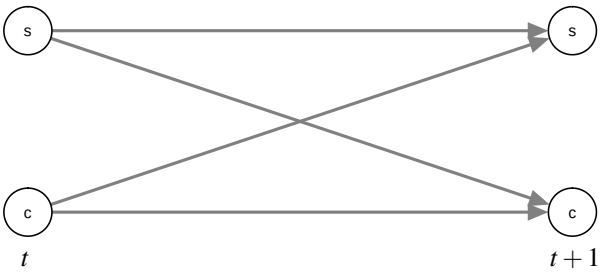


Figure 2.5: A visual representation of the Graphical VAR model. S = sleep problems. C = concentration problems. t = time point.

Figure 2.6 shows network structures for data of the 57-year old male with depression estimated using Graphical VAR. For explanatory purposes, we assumed all variables to be continuous. The left panel shows the contemporaneous network, and the right panel the temporal network, estimated with $\gamma = 0$. We left the network structures for $\gamma = 1$ out for clarity of presentation. In the contemporaneous network (left panel), variable 5 (I feel lonely) has strong positive relations with items 2 (I feel down) and 6 (I feel anxious). Furthermore, a strong negative relation is present between item 12 (I feel strong) and item 16 (I can concentrate well). This means that, as the participant feels more strong, he can concentrate less at the same time point. In the temporal network (right panel), we see strong positive self-loops for items 11 (I doubt myself), 22 (I am tired) and 27 (I have a headache): when the participant doubts himself at one time point the feedback reinforces this feeling, and he will doubt himself more at the next time point. Interest-

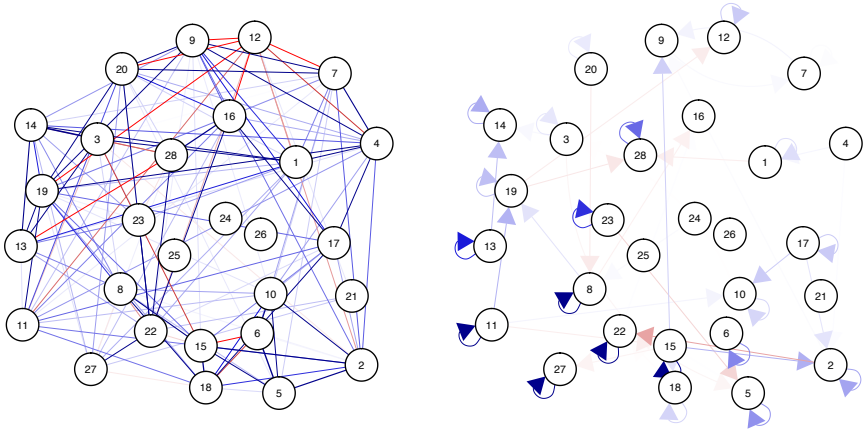


Figure 2.6: Example networks estimated with the Graphical VAR estimation procedure with $\gamma = 0$. Left panel: contemporaneous network. Right panel: temporal network.

ingly, the contemporaneous network has many more edges than the temporal network: in the contemporaneous network, approximately 36% of the edges are nonzero, while in the temporal network, approximately 4% of the edges are nonzero. This shows the importance of looking both at contemporaneous and temporal network structures to uncover the dynamics of an individual throughout time.

2.2.3 Mixed Data: Mixed Graphical Model

In the previous two sections, we discussed methods that estimated one network structure. When we estimate one network structure, we assume that this structure does not change over time, and is thus *stationary*. When dealing with time-series data, we are interested in how an individual potentially changes over time, and therefore we are also interested in how the associated network might change over time. Where the IsingFit and the Graphical VAR method both estimate one (contemporaneous) network, the Mixed Graphical Model (MGM) method estimates multiple (more than 1) network structures that together demonstrate how the relations between variables in a network of an individual may change over time (Haslbeck & Waldorp, 2016a,b). For example, with the MGM method, we can esti-

mate two, five, ten, or more network structures (how many is up to the user to decide) that, when shown sequentially, display potential changes in the network structure throughout time.

In order to estimate time-varying networks using the MGM method, we start out by identifying the type of variables in the time-series: variables can be continuous (e.g., age) or categorical (e.g., sex). The time-series data is then cut into z pieces (the number of pieces is chosen by the user), and for each piece, the network structure is estimated, taking the types of variables (i.e., continuous or categorical) into account (see Haslbeck & Waldorp, 2016b for a more detailed description). Similar to the IsingFit method, each node is regressed on all other nodes. For example, when we again consider Figure 2.2, the vocab node will be regressed on reading, maze, blocks, picture and general, reading will be regressed on vocab, maze, blocks, picture and general, and so on. After this node-wise regression, we obtain two edge weights for the edge between vocab and reading, for example. When both of these edge weights are nonzero, the mean of these edge weights will be used to visualise the relation between, in this example, vocab and reading. When either of the two edge weights is zero, no edge will be drawn between the two variables.

Because one is estimating multiple network structures that together show the potential changes in a network structure throughout time, one needs to decide how many network structures are desired, which is an arbitrary choice. Also, the MGM method assumes that the network structure does not change within each piece of time for which the network structure is estimated. The MGM method assigns each time point within one piece a certain weight, that indicates how important that time point is for the estimation of the edges. Time points that are seen as more important will have a stronger influence on the edge estimation process than time points that are seen as less important. For this chapter, we use the default assignment of weights to time points, as suggested by Haslbeck & Waldorp (2016a). For a more detailed and technical explanation of the MGM method, see Haslbeck & Waldorp (2016a,b).

Figure 2.7 shows network structures for the single patient data estimated using MGM.

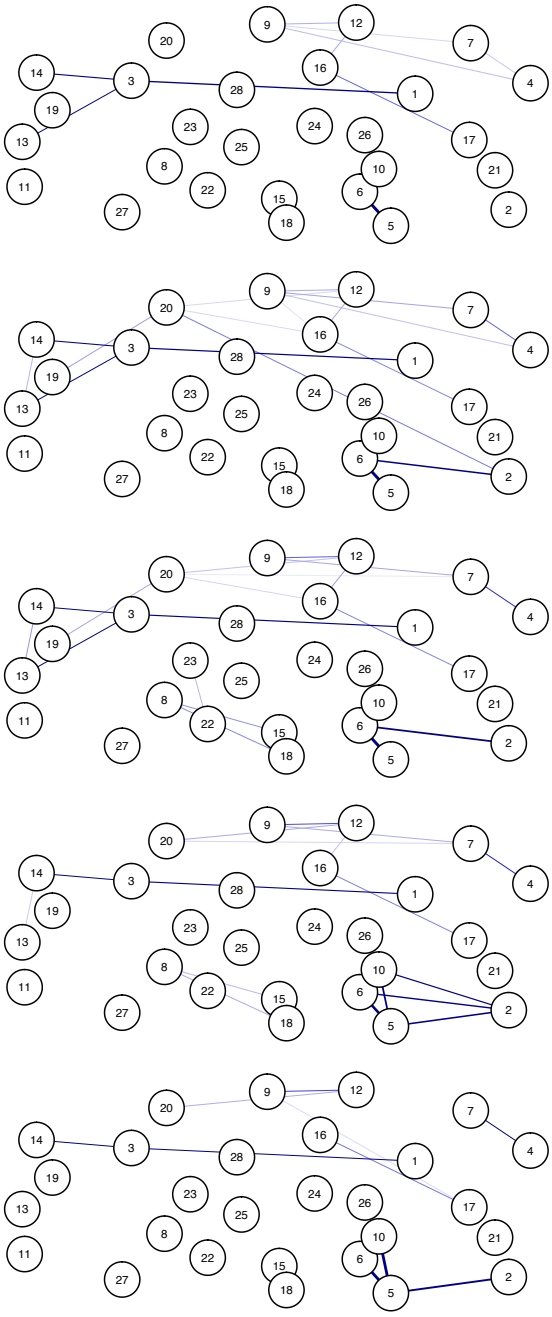


Figure 2.7: Example networks estimated with the MGM estimation procedure. Each row denotes a time step. All networks are estimated with $\gamma = 0$.

In order to estimate an MGM, we dichotomised items 2 (I feel down), 5 (I feel lonely), 6 (I feel anxious) and 10 (I feel guilty), because their scale ranges from -3 to +3 and the MGM method cannot handle negative scores. Furthermore, we assumed that items with seven response categories were continuous, resulting in three continuous items and 25 categorical items. We decided (arbitrarily) to divide the time-series up into five pieces and estimate five networks accordingly. We estimated the five network structures both with $\gamma = 0$ and $\gamma = 1$. We left the network structures estimated with $\gamma = 1$ out for clarity of presentation. Starting at the top network in Figure 2.7 and then going down, the changes in the network structure become visible. For example, the edge between items 5 (I feel lonely) and 10 (I feel guilty) is not present in the first network, but slowly appears in the fourth and fifth network. This means that the relation between feelings of loneliness and feelings of guilt developed over time. Furthermore, it can be seen that the edge between items 2 (I feel strong) and 10 (I feel guilty) appears in the fourth network, but disappears again in the fifth network. This indicates that, at the beginning and at the end of the data collection period, feeling strong and feeling guilty were not related, but this relation developed into a positive relation, and then disappeared again. This shows the importance of estimating not one network structure, but several network structures, as Figure 2.7 shows that we gain more insight into the dynamics of an individual throughout time compared to only one network.

2.3 Catastrophe Theory

One of the problems with networks (also called *systems*) is that they easily become complex. Even a relatively simple model with, say, ten variables already entails more than fifty parameters to be estimated. For a slightly more complex model with 15 variables the number of parameters to be estimated is larger than 100. As such, if one is interested in understanding and/or predicting the behaviour of a system (network) as a whole, then this becomes more difficult as the number of system variables increases and with it the number of parameters. Therefore, in other fields, one reduces the complexity of a network model

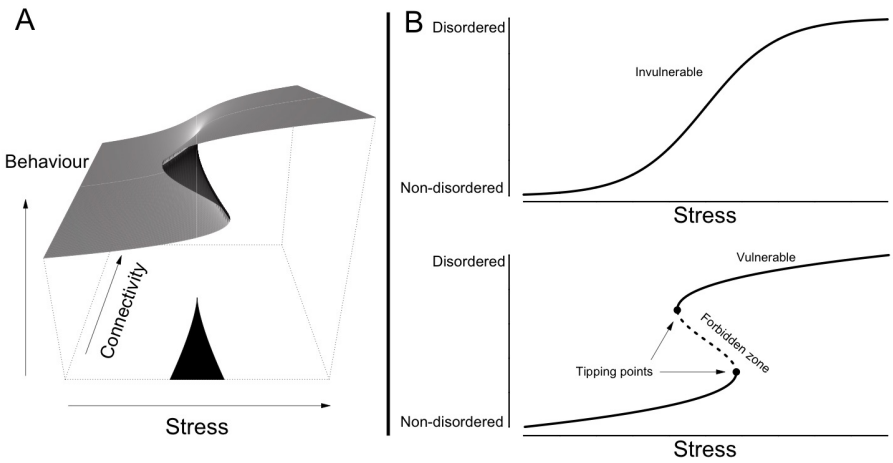


Figure 2.8: Example of a cusp catastrophe model for a psychological disorder (i.e., system). *Panel A*: the 3D representation with stress on the *x*-axis, connectivity on the *y*-axis and the behaviour of the system (ranging from non-disordered to disordered) on the *z*-axis. *Panel B*: the 2D representation of the cusp as depicted in *Panel A*. In the case of weak connectivity (top figure in *Panel B*), the system shows smooth continuous behaviour in response to increasing stress (invulnerable networks). In the case of strong connectivity (bottom figure in *Panel B*), the system shows discontinuous behaviour with sudden jumps from non-disordered to more disordered states and vice versa (vulnerable networks). The system with strong connectivity shows two tipping points (black dots in bottom figure of *Panel B*) with in between a forbidden zone (i.e., dashed part of the black line in bottom figure of *Panel B*): in that zone, the behaviour of the system is unstable to such an extent that even a minor perturbation will force the system out of that state into a stable state (i.e., the solid parts of the black line).

in order to understand its general dynamics by means of using a *cusp catastrophe model*. With this mathematical model, one reduces complexity by considering two orthogonal control variables - instead of the more than 100 parameters when a system is comprised of 15 variables - and their impact on the behaviour of a system.

Figure 2.8 shows a hypothetical cusp catastrophe model with two control variables,

stress (*x*-axis) and connectivity (*y*-axis; Cramer et al. 2016) that have their own way of influencing the behaviour of the system (i.e., a psychological disorder in our example; *z*-axis). Panel A of this figure shows the 3D representation while Panel B shows the 2D representation with stress on the *x*-axis and the behaviour of the system on the *y*-axis for two situations: weak connectivity (top figure of Panel B) and strong connectivity (bottom figure of Panel B). Stress acts as the so-called *normal* variable while connectivity acts as the so-called *splitting* variable.

For sufficiently low values of the splitting (i.e., connectivity) variable, the behaviour of the system responds to increasing values of the normal (i.e., stress) variable in a smooth and continuous fashion (see top figure of Panel B of Figure 2.8): that is, when connectivity of a, say, depression system is relatively weak, then more stress results in the system being more depressed in a continuous fashion. Cramer et al. called such systems *invulnerable* or resilient networks. For higher values of the splitting variable (i.e., strong connectivity) the outcome surface splits and bifurcates from smooth, continuous changes to sudden, discontinuous jumps for increasing values of the normal variable (i.e., stress; see bottom figure of Panel B of Figure 2.8): that is, when connectivity of a psychopathological system is relatively strong, then at a certain point (i.e., the tipping points in the bottom figure of Panel B of Figure 2.8) the system either jumps from a non-disordered to a disordered state, or the other way around, which is called a *critical transition*. Cramer et al. called such systems *vulnerable* networks. In between these tipping points lies a *forbidden zone* (in bottom figure of Panel B of Figure 2.8: the dashed part of the black line): within this zone, the behaviour of the system (i.e., the disorder) is unstable to such an extent that even a tiny amount of stress will already kick the system into a stable (non-)disordered state (i.e., the solid parts of the black line).

A final phenomenon that is present in networks with strong connectivity is *hysteresis*. Figure 2.9 shows a hypothetical example of hysteresis for a symptom network that is strongly connected. The *x*-axis represents stress while the *y*-axis represents the state of the system (from a healthy - none or few symptoms active - to a disordered state - most

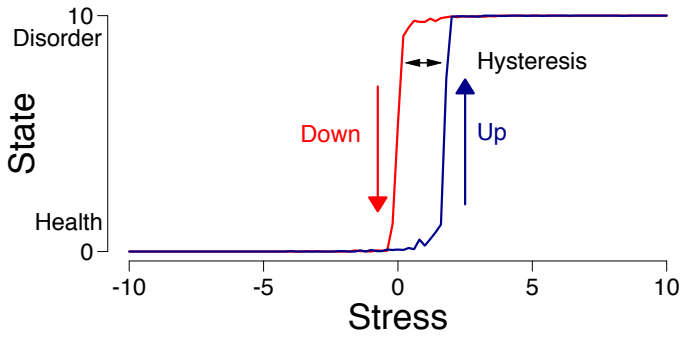


Figure 2.9: An example of hysteresis for a network with strong connectivity. The x-axis represents stress while the y-axis represents the state of the system (from a healthy - none or few symptoms active - to a disordered state - many or all symptoms active). The blue line presents an average number of hypothetical active symptoms when stress was increasing (UP); and the red line represents an average number of hypothetical active symptoms when stress was decreasing (DOWN). Hysteresis is the phenomenon that the amount of stress needed to get a system into a healthy, non-disordered, state (jump from "disorder" to "health" in the red line) exceeds the amount of stress that tipped the system into a disordered state in the first place (jump from "health" to "disorder" in the blue line): i.e., hysteresis is the gap, marked with an arrow, between the red and blue line.

or all symptoms active). The blue line (UP) shows what happens if stress is increasing: as we already saw in Figure 2.8 increasing stress results in a sudden jump from a healthy state to a disordered state. But what happens when, once disordered, we start lowering the amount of stress that is influencing the system? This is shown with the red line (DOWN): the amount of stress needed to get the system back into a healthy, non-disordered, state (jump from "health" to "disorder" in the blue line) exceeds the amount of stress that tipped the system into a disordered state in the first place (jump from "disorder" to "health" in the red line). That is, hysteresis is the gap, marked with an arrow in Figure 2.9, between the blue and the red line.

2.4 The Cramer Model

In an earlier section we elaborated on the theoretical principles of a cusp catastrophe model. We sketched one possibility of such a model in which a disorder is the result of an interaction between two control parameters, the connectivity of the system (i.e., the strength of relations in a symptom network of a disorder) and the amount of stress (i.e., external perturbation) that is put on that system (see Figure 2.8). This may sound appealing but we first need to establish that, indeed, vulnerability of a system to becoming disordered has something to do with the connectivity of that system. In this section, we first provide some information on mania data that we used as an example for showing the workings of the Cramer model. Next, we outline the specifics of this model and show that, indeed, vulnerable networks/systems are those with strong connections between its symptoms. We conclude this section by showing how one might detect critical transitions with this model.

2.4.1 Mania Data

Data for our example came from the National Comorbidity Survey Replication (NCS-R). This is a nationally representative household survey of English speakers 18 years and older in the United States (see Kessler et al., 2004). The NCS-R survey schedule is the version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI) that was developed for the WHO World Mental Health Survey Initiative (WMH-CIDI; Kessler & Ustun 2004). The interviews were conducted between February 2001 and April 2003. A total of 9282 respondents participated in Part 1 of the interview (core diagnostic assessment), the data of which we used for this chapter. Specifically, we used data for the 10 symptoms of mania: elevated mood, restlessness, pressure of speech, flight of ideas, loss of social inhibitions, decreased need for sleep, inflated self-esteem, distractibility, reckless behaviour, and marked sexual energy. For each of these criteria, respondents had to indicate whether or not they had suffered from that symptom in the

past year. As such, the data were dichotomous. Missing data were imputed with zeroes.

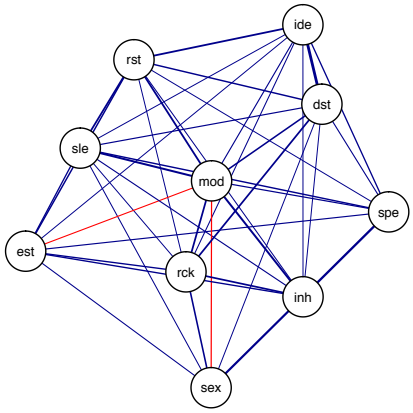
We estimated network parameters for the 10 symptoms of mania in the NCS-R dataset with the IsingFit method as explicated earlier. Figure 2.10 shows the resulting network. Note that the skip structure of the NCS-R schedule is responsible for the fact that elevated mood (*mod* in Figure 2.10) is such a central symptom in this network: a 'yes' response is needed to proceed to the remainder of the questionnaire while a 'no' results in skipping the other nine mania criteria. Other relatively strong connections seem intuitively plausible: for example, a strong connection between loss of social inhibitions and marked sexual energy; and between distractibility and flight of ideas. The empirical parameters of this network are used in the next section as input for the formal dynamic Cramer model.

2.4.2 The Cramer Model

The model assumes the following. First, symptoms X_i can be 'on' (1; active) or 'off' (0; inactive). Second, symptom activation takes place over time t such that, for example, elevated mood at time t may cause marked sexual energy at time $t+1$. Third, a symptom i receives input from symptoms with which it is connected in the empirical mania symptom network based on the NCS-R data. We call these symptoms with which symptom i is connected neighbour symptoms. The empirical weight parameters are collected in a matrix $\mathbf{W}$ for the $J = 10$ mania symptoms: entry W_{ij} thus represents the logistic regression weight between symptoms i and j as estimated from the NCS-R data. Next, the total amount of activation a symptom i receives at time t is the weighted (by $\mathbf{W}$) summation of all the neighbouring symptoms $\mathbf{X}$ - that is, the vector that contains the "0" (inactive) and "1" (active) values - at time $t-1$. Cramer et al. (2016) call this the *total activation function* (boldfaced parameters are estimated from the NCS-R mania data):

$$A_i^t = \sum_{j=1}^{J} \mathbf{W_{ij}} X_j^{t-1} \tag{2.1}$$

An example: suppose we wish to compute the total activation function for marked sexual energy, which is connected to elevated mood with a NCS-R weight parameter of



Symptom	Threshold
Elevated mood (mod)	-2.15
Restlessness (rst)	-9.62
Pressure of speech (spe)	-6.75
Flight of ideas (ide)	-9.66
Loss of social inhibitions (inh)	-9.66
Decreased need for sleep (sle)	-9.61
Inflated self-esteem (est)	-9.96
Distractibility (dst)	9.64
Reckless behavior (rck)	-9.70
Marked sexual energy (sex)	-8.64

Figure 2.10: The inter-individual mania symptom network based on the NCS-R data. Each node in the upper panel of the figure represents one of the 10 symptoms of mania. A line (i.e., edge) between any two nodes represents a logistic regression weight: the line is blue when that weight is positive, and red when negative. An edge becomes thicker as the regression weight becomes larger. The lower part of the figure displays the estimated thresholds for each symptom.

2 and to reckless behaviour with a NCS-R weight parameter of 0.80. Also assume that at time $t-1$ elevated mood is activated while reckless behaviour is not (encoded in the vector $\mathbf{X}$). In this simple example, the total activation at time t for the symptom marked sexual energy is $2 \cdot 1 + 0.80 \cdot 0 = 2$.

In the next step, we develop a function for computing the probability of symptom i becoming active at time t. This logistic probability function states the following: the probability of symptom i becoming active at time t depends on the difference between the total activation of its neighbouring symptoms and the threshold (estimated with the NCS-R mania data) of symptom i (in Formula 2.2: $\mathbf{b_i} - A_i^t$). The more the total activation exceeds the threshold of symptom i at time t, the higher the probability that symptom i becomes active (in Formula 2.2 $P(X_i^t = 1)$) at time t. Cramer et al. (2016) call this the *probability function* (boldfaced parameters are estimated from the NCS-R mania data):

$$P(X_i^t = 1) = \frac{1}{1 + e^{b_i - A_i^t}} \tag{2.2}$$

Please note that the parameter $\mathbf{b_i}$ denotes the absolute value of the threshold of symptom i as estimated from the NCS-R data. So let us return to our simple example above in which marked sexual energy was only connected to elevated mood and reckless behaviour. Suppose that the threshold for marked sexual energy is 1 and we already know that the total activation at time t equals 2. Then the probability of having marked sexual energy at time t equals $\frac{1}{1+e^{1-2}} = 0.73$. This probability becomes much lower when the threshold exceeds the total activation, for example when the threshold would be 4. In that case, the probability becomes $\frac{1}{1+e^{4-2}} = 0.11$. A special case arises when A_i^t is equal to $\mathbf{b_i}$, that is, when the amount of activation of the neighbours of symptom i is exactly equal to the threshold of symptom i. In that case, the probability of symptom i becoming active is exactly $1/2$ (e.g., $\frac{1}{1+e^{2-2}} = 0.50$).

To summarise, the Cramer model is a process model that develops over time. The probability of a symptom becoming active at a particular point in time depends on both its threshold and the amount of activation it receives from its neighbouring symptoms at

that same point in time. The more activation a symptom i receives from its neighbouring symptoms and the lower its threshold, the higher the probability of symptom i becoming active. In the next few sections, we will use this model to study the behaviour of a strongly connected mania system when put under stress. Will this behaviour conform to what we would expect from the cusp catastrophe model we elaborated on earlier (i.e., the bottom graph of Panel B of Figure 2.8)?

2.4.3 Putting the Cramer model under stress: critical transitions?

For the purposes of this section we are particularly interested in the potential behaviour of a strongly connected system that is put under stress. Would we find, as a cusp catastrophe model predicts, that putting stress on a strongly connected system results in sudden discontinuous jumps from one state to the other? More specifically, would there be a *critical transition*, the change that occurs when the condition of a person exceeds the tipping point (dots in bottom graph of Figure 2.8), which transfers the person from one state (e.g., no mania) to another (e.g., mania) in response to a relatively small external force: i.e., in the bottom graph of Panel B of Figure 2.8 only a relatively small amount of stress is needed to get the system past the tipping point into an alternative stable state (i.e., one of the solid parts of the black line in bottom graph of Panel B in Figure 2.8; Olde Rikkert et al. 2016). Cramer et al. (2016) showed that for a strongly connected depression network, it was indeed the case that the behaviour of the depression system was discontinuous, with sudden jumps from a non-depressed state to a depressed state with a forbidden zone in between both stable states. Here, we sought to replicate these findings for the NCS-R mania data.

First, in order for the mania system to be sufficiently strongly connected, we multiplied the weights matrix $\mathbf{W}$ that was estimated with IsingFit on the NCS-R mania data with $c = 1.5$. Second, we extended the total activation formula in 2.1 with a stress parameter S_i^t a number that was added to the total activation of the neighbours of symptom i at time t: the higher S_i^t - that is, the more stress - the higher the total activation function, and

thus the higher the probability (Formula 2.2) that symptom i will become active at time t. This resulted in the following modified total activation function:

$$A_i^t = \sum_{j=1}^{J} c\mathbf{W_{ij}} X_j^{t-1} + S_i^t \tag{2.3}$$

The probability function remained the same as stated in Formula 2.2. We simulated 10000 time points starting with all symptoms being "off" (i.e., vector $\mathbf{X}$ with only zeroes). At each time point, we computed the total activation and the resulting probability of a symptom becoming active. Next, symptom values (either "0" or "1", denoting inactive and active, respectively) were sampled using these probabilities. Meanwhile, over the course of the 10000 points S_i^t was repeatedly gradually increased from -15 to 15 and then decreased from 15 to -15 with small steps of 0.01. The impact of the stress parameter on the behaviour of the mania system was quantified by computing the average state M of the system, the average number of symptoms active at a certain time point t. Specifically, since all the stress parameter values were used multiple times during the simulation - because of the repeated increasing and decreasing of the stress parameter during the course of the simulation - we averaged states within 0.20 ranges of these stress parameter values.

Figure 2.11 shows the main results of the simulation: the x-axis represents stress while the y-axis represents the average state of the system M (note: the higher on the state variable, the more mania symptoms are active, and thus the more 'manic' the system). The grey line (and points) represents the average number of active symptoms when stress was increasing; the black line (and points) represents the average number of active symptoms when stress was decreasing. The results are consistent with what we predicted from the cusp catastrophe model. First, the behaviour of the system was discontinuous with sudden jumps from a non-manic to a more manic state and vice versa. That is, a small increase (decrease) in stress can lead to a disproportional reaction, resulting in a more manic (non-manic) state with more (less) symptoms active. Second, Figure 2.11 clearly shows that during the transition from healthier to more manic states, and vice versa, a 'forbidden zone' (from around 1 to 7 symptoms) was crossed that does not seem to func-

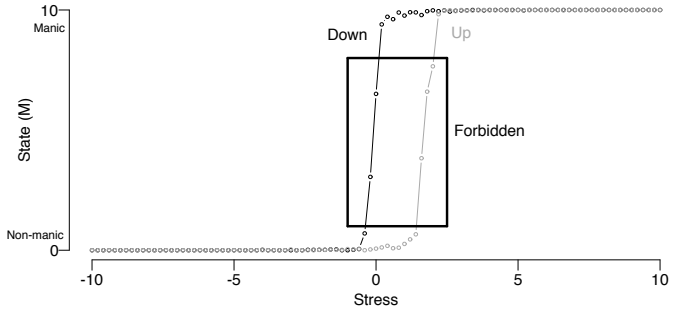


Figure 2.11: The average state of the mania system in response to stress in the case of strong connectivity. The *x*-axis represents stress while the *y*-axis represents the average state of the mania system M: that is, the total number of active symptoms averaged over every 0.20 range of the stress parameter. The grey line (and points) depicts the situation where stress is increasing (UP) whereas the black line (and points) depicts the situation where stress is decreasing (DOWN). The black box with *Forbidden* noted next to it represents an area without data points.

tion as a stable state (i.e., no data points in that area, see black box in Figure 2.11). Third, and final, the results show clear hysteresis: the amount of stress reduction needed to get the system into a non-disordered state (i.e., only a few symptoms active or none at all) exceeds the amount of stress that initially tipped the system into a more manic state.

In sum, when put under stress, strongly connected systems that behave over time according to the equations of the Cramer model (Cramer et al., 2016) show critical transitions from non-disordered states to disordered states and vice versa. These critical transitions take place at tipping points and between these tipping points states are unstable thereby 'forcing' the system to either the upper branch of disorder (solid upper black line in bottom graph of Panel B in Figure 2.8) or the lower branch of relative health (solid bottom black line in bottom graph of Panel B of Figure 2.8). Perhaps more important than knowing that there are critical transitions, is there any way with which we can detect such upcoming critical transitions? Detecting upcoming critical transitions in real data

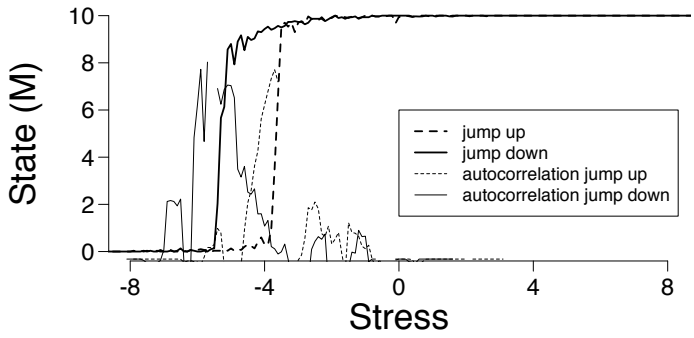


Figure 2.12: Increasing autocorrelation as a potential early warning signal in the mania network with strong connectivity. The x-axis represents stress while the y-axis represents the average state: that is, the total number of active mania symptoms averaged over every 0.20 range of the stress parameter value. The dashed lines depict the situation where stress is increasing whereas the solid lines depict the situation where stress is decreasing. The "jump" lines show the total number of active symptoms (i.e., the state M), the "autocorrelation" lines track the autocorrelation between these states over time.

of individuals might prove beneficial, for example because it would allow for early intervention, before a critical transition to a disordered state takes place, that might help in preventing a diagnosis.

2.4.4 Detecting critical transitions: critical slowing down

There is evidence for the hypothesis that all catastrophic systems, from financial systems to the climate, display *early warning signals* that a system is approaching a tipping point (Carpenter & Brock, 2006; Dakos et al., 2008; Fort et al., 2010; Scheffer et al., 2014). One such early warning signal is called *critical slowing down*: right before a tipping point, the system becomes increasingly slower in recovering from small perturbations. Pertaining to mania, for instance, one might see that someone has more difficulty than usual to recuperate from a relatively minor daily hassle such as an argument with a spouse over which restaurant to go to.

Numerically, this slowing down can be traced by inspecting autocorrelations: that is, the correlations between values of the same variable at multiple time points (e.g., the correlation between 60 measurements of elevated mood over time). These autocorrelations increase when the system slows down: slowing down means that at each time point, the system much resembles the system as it was at the previous time point, meaning that the autocorrelation is relatively high. The question now is: if we compute the autocorrelations between the state variable *M* of the strongly connected mania system, would we find an increase in these autocorrelations?

Again, we used the NCS-R mania data and specifically inspected the autocorrelations between the state variable *M* (i.e., the total number of active mania symptoms) over time for the network with strong connectivity. The setup of the simulation was identical to the one reported in the previous section about putting a network/system under stress. The results are presented in Figure 2.12. As expected, when stress was increasing, the autocorrelations between the states of the strongly connected mania network increased (dashed line increasing) markedly right before the network abruptly switched from a relatively non-manic to a manic state (thicker dashed line jumping from 1 to 10 symptoms). However, Figure 2.12 also showed an increase in active symptoms from 0 to 1 symptoms before the sharp increase in autocorrelations so it appears that for this data example, the early warning was not early enough. Since Cramer et al. (2016) did show a clear early warning signal for major depression data it remains an open question to what extent the present results are a clue to the unsuitability of autocorrelations as window into predicting critical transitions.

2.5 The Empirical Mean Field Approximation

2.5.1 Description

The empirical MFA is based on the work of Waldorp & Kossakowski (2020). As previously stated, networks can quickly become quite complex. By assuming that variables

behave in a similar manner and have the same number of connections to other variables, we can reduce such a complex and multivariate system to a single equation (Kossakowski et al., 2019; Waldorp & Kossakowski, 2020). We can then use this equation to investigate the behaviour of the network as a whole over time.

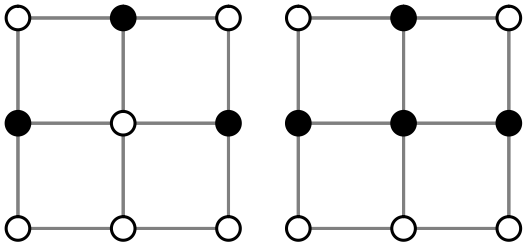


Figure 2.13: Example network. WHITE = inactive nodes. BLACK = active nodes. Grey lines indicate connected nodes.

Suppose we have a network with nine nodes, such as the one depicted in the left panel of Figure 2.13. In this network, nodes are either active (1, black) or inactive (0, white). For example, in the left panel of Figure 2.13, three nodes are active (i.e., black). In the MFA, we check for each node how many of its neighbours (nodes that are directly connected to a specific node) are active. In Figure 2.13 (left panel), when we focus on the middle node, we see that it has four neighbours, and three of them are active. We use this information on the activity of a node's neighbours and whether or not this is the majority of a node's neighbours to set a probability parameter p that decides whether or not a specific node becomes active itself at the next time point $t+1$. As a rule, we state that, when at least the majority of a node's neighbours is active, the probability for that specific node to become active will be $1-p$. When less than the majority of a node's neighbours is active, this probability will be p. Looking again at Figure 2.13, 3 out of 4 neighbours (i.e., the majority) of the middle node are active at time point t, so the probability for the middle node to become active at time point $t+1$ is $1-p$. In the right panel of Figure 2.13, we see that the middle node indeed became active as a result of its active neighbours and

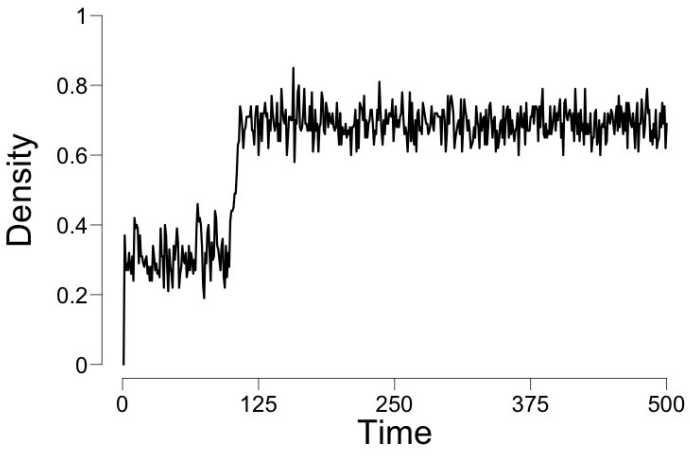


Figure 2.14: Visualisation of the density set against time.

its probability parameter. We can repeat this process for each node within a time point, and for each time point, thereby creating a $t \times n$ matrix that contains 0s and 1s for each time point t and each node n.

For each time point, we can calculate the number of active nodes proportional to the total number of nodes, called the density. These densities can be visualised, as is shown in Figure 2.14 with time on the x-axis and the densities on the y-axis. In this figure, we see how the density, the proportion of active nodes in a network, changes over time. In this example, we see that the density suddenly jumps from one stable state to the other, which is a critical transition. In this case, even though we spotted the transition, we are too late in investigating whether or not the individual was vulnerable for the critical transition before it occurred. Fortunately, the empirical MFA enables us to investigate this vulnerability.

The network in Figure 2.13 is called a *grid* where each node has the same number of neighbours. In psychology, it is hard to come up with an example in which each variable has the same number of neighbours. The MFA assumes that each node has the same number of neighbours, and Waldorp & Kossakowski (2020) investigated the performance of the MFA when this assumption is no longer satisfied, and nodes can have a different number of neighbours. They also investigated two other types of network structures: a

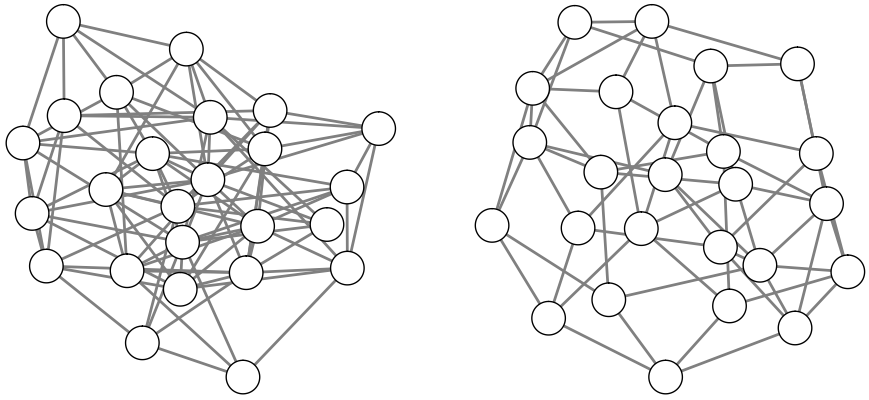


Figure 2.15: Examples of a random graph (left panel) and a small world graph (right panel). Edge probability $p_e = 0.30$. Rewiring probability $p_w = 0.20$.

random graph and a *small world graph*, both of which are displayed in Figure 2.15. A random graph is a network in which each edge has a constant probability to be drawn, p_e which is independent of the size of the network (Bollobás, 2001; Durett, 2007). In Figure 2.15 (left panel), we see a random graph with an edge probability of 0.3, which means that each edge has a probability of 0.3 to be present in the network. This also means that around 30% of the edges are present in the network. In the right panel of 2.15, we see a small world graph, which is a network structure that starts out as a grid (i.e., each node has the same number of neighbours, thus edges), but in which each edge has an constant probability p_w to be rewired to two other, possibly non-neighbouring nodes and added to the grid (Newman & Watts, 1999a), and which is also independent of the size of the network. In Figure 2.15 (right panel) we see a small world graph with a rewiring probability of 0.2, which means that each edge in the original grid has a probability of 0.2 to be rewired and added to the network.

In Waldorp & Kossakowski (2020), the performance of the MFA was shown by means of a simulation study: they theoretically determined the shape of the bifurcation diagram, and determined how well simulated data matched the theoretical bifurcation diagram un-

der various conditions. In Kossakowski et al. (2019), they showed how the MFA can be used in practice. Using maximum likelihood (ML) estimation, it was shown how the probability parameter p can be estimated from the data. This empirical MFA method works as follows: first, the network structure is estimated using IsingFit. As there currently is no valid test that can distinguish between a random graph and a small world graph, it is assumed that the network structure is a random graph, with the ratio of edges that is non-zero and the possible number of edges as the edge probability p_e. The edge probability p_e is needed to create the bifurcation diagram. After that, the probability parameter p is estimated from the data using ML estimation. The estimate of p ($\hat{p}$) is then set off against the bifurcation diagram.

Figure 2.16 shows an example of such a bifurcation diagram. A bifurcation diagram is created by plugging in several values for p (x-axis), and plotting the resulting expected density μ_p (y-axis). A bifurcation diagram consists of two parts: the first is the area where two lines occur (at $p = 0$ to $x \approx 0.4$). This is the area where critical transitions can occur; the two lines in this part of the diagram represent the two stable states that an individual can be in. The second part is the area where only one line exists (at $p \approx 0.4$ to $x = 1$): there is only one line, and therefore, critical transitions cannot occur. The empirical MFA compares $\hat{p}$ with the point at which the line splits into two lines, called the critical value. When $\hat{p}$ is smaller than the critical value, we conclude that the individual has an increased risk for experiencing a critical transition. When $\hat{p}$ is higher than the critical value, we conclude that the individual does not have an increased risk for experiencing a critical transition. For a more detailed and technical description, see Kossakowski et al. (2019).

2.5.2 Application of the Empirical MFA to general affect data

Data for our example were collected from and by the first author (hereafter called participant JK). For testing purposes, the participant monitored her daily affects using the *experience sampling method* (ESM; Csikszentmihalyi & Larson, 1987). With ESM, the

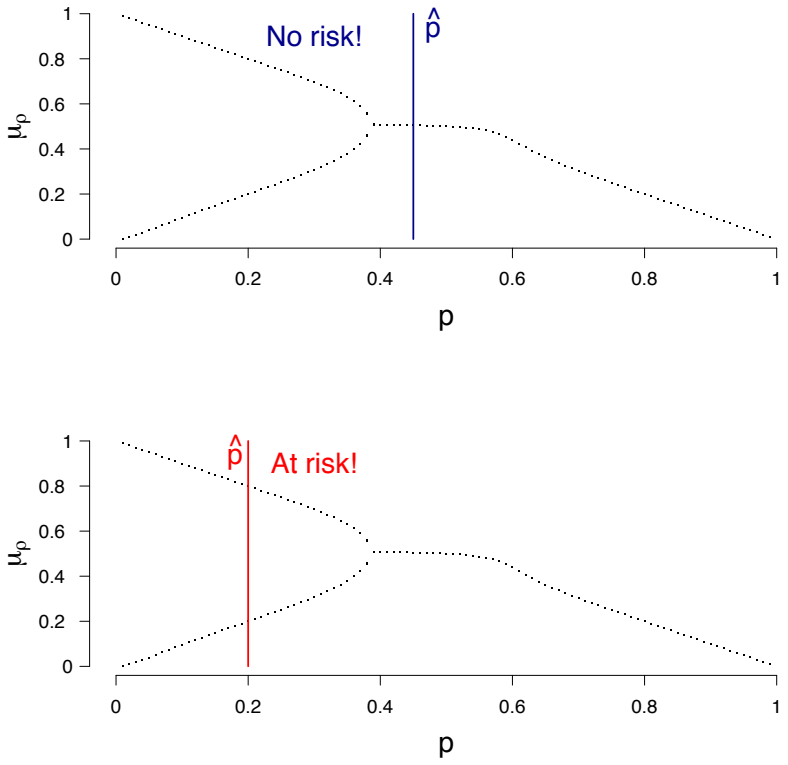


Figure 2.16: Example bifurcation diagram. The upper diagram denotes an example where the estimate $\hat{p}$ is in the area where one line exists. The lower diagram denotes an example where the estimate $\hat{p}$ is in the area where the two prongs exist.

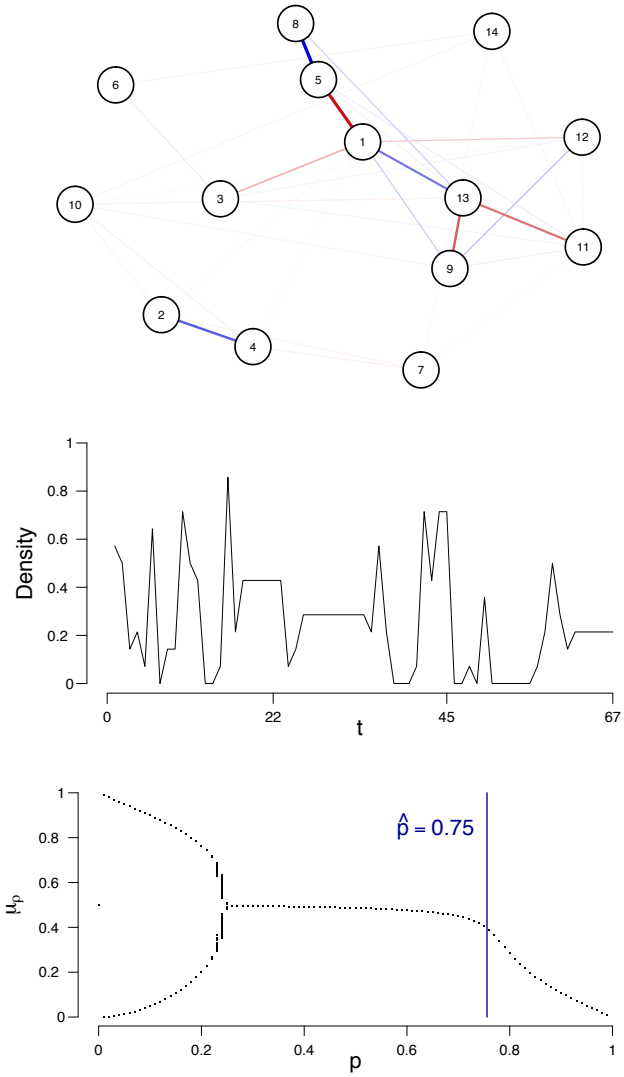


Figure 2.17: Network of participant JK (upper panel), the evolution of the percentage of active nodes for each time point (middle panel), and the bifurcation diagram (lower panel). The size of the associations between two nodes in the network is represented using the colour and thickness of an edge, where positive associations are represented by blue, and negative associations by red. The red line in the bifurcation diagram in the lower panel indicates the estimate p.

participant's waking hours are divided by s intervals of equal size. Within these intervals, the participant received a beep at a random time (within that interval) and is requested to fill out a questionnaire, often on a palmtop or smartphone. Over the course of seven days, participant JK (a 27-year-old female) completed 45 measurements, and missed 22 measurements; in total 67 measurements were collected. Participant JK completed the *Positive Affect Negative Affect Scale* (PANAS; Watson et al., 1988), a 20-item questionnaire that contains ten positive mood states and ten negative mood states, each measured on a 5-point Likert scale, ranging from 'very slightly or not at all' to 'extremely'. Table 2.2 shows the twenty different mood states and their assigned node labels.

Positive variables were recoded, so that high scores indicate a more negative affect. Missing measurements were replaced by the previous measurement, after which the variables were dichotomized using a median split, which is necessary for using IsingFit. Finally, we removed six variables due to observing either of two response categories less than four times.

Figure 2.17 (upper panel) depicts the network structure of participant JK, estimated with the IsingFit method. For example, item 1 (feeling interested) has a strong connection with items 3 (feeling excited), 5 (feeling strong) and 13 (feeling ashamed). Interestingly, the association between items 1 and 5 is negative, which means that, as participant JK feels more interested, she feels less strong and the other way around. The association between items 2 (feeling distressed) and 4 (feeling upset) makes sense: as participant JK feels more upset, she tends to feel more distressed and the other way around.

Figure 2.17 (middle panel) shows the trajectory of the density (the proportion of active nodes) of participant JK over time. It can be seen that sometimes, the density does not change, especially in the middle of the measurement period, but also that, at several occasions, some peaks occur that are indicative of high general negative affect. However, since there is no clear pattern in the densities, we would expect that participant JK is not vulnerable for a critical transition from a positive affective state to a negative affective state. The lower panel of Figure 2.17 shows the bifurcation diagram for participant JK,

Item label	Item meaning
1	To what extent do you feel interested
2	To what extent do you feel distressed
3	To what extent do you feel excited
4	To what extent do you feel upset
5	To what extent do you feel strong
6	To what extent do you feel guilty
7	To what extent do you feel scared
8	To what extent do you feel hostile
9	To what extent do you feel enthusiastic
10	To what extent do you feel proud
11	To what extent do you feel irritable
12	To what extent do you feel alert
13	To what extent do you feel ashamed
14	To what extent do you feel inspired
15	To what extent do you feel nervous
16	To what extent do you feel determined
17	To what extent do you feel attentive
18	To what extent do you feel jittery
19	To what extent do you feel active
20	To what extent do you feel afraid

Table 2.2: Items of the Positive Affect Negative Affect Scale (PANAS) and their assigned item labels.

and $\hat{p}$ that we estimated from the data. It shows that $\hat{p}$ is much higher than the critical value, which lies around 0.26. This means that participant JK is not vulnerable for a critical transition.

2.6 Discussion

In this chapter, we have outlined a network perspective on psychological constructs. We have demonstrated various methods for estimating network structures and, subsequently, we have shown – with the Cramer model and the Empirical MFA – how one can use these network structures to investigate vulnerability and critical transitions: what makes certain people vulnerable for developing psychopathology and can we anticipate critical transitions from, say, a healthy to a disordered state? Pertaining to the latter question: yes, we may. In the Cramer model, we saw that individuals with more strongly connected networks are more vulnerable for the development of psychopathology. Detecting critical transitions by means of autocorrelations, however, was an only partially successful endeavour. In the Empirical MFA, the vulnerability was estimated from the data: the lower the probability parameter, the more vulnerable an individual is for a critical transition. For our general affect examples, we showed that it is possible to see and potentially anticipate critical transitions from one state to another.

Throughout this entire chapter, we investigated vulnerability for a critical transition that an individual may or may not have. In the mania example described in section 2.4 the conceptual meaning of the two possible states and the transition is clear: the system is in either a state of mania or not. However, in the general affect example described in section 2.5 this may not be as clear. In terms of general affect, what are the two states that an individual can be in, and what does the transition between one and the other mean? A vulnerability analysis such as the ones performed in this chapter might be conceptually meaningless if the system's two states are ill-defined or not defined at all. Therefore, not every (psychological) construct might lend itself for such an analysis, and when applying these models, researchers are best advised to first have a clear view on what the two states

are in the system, and what it means if an individual transitions between them. Additionally, some psychopathological constructs might have more than two states. Consider bipolar disorder for example: this disorder is characterised by switching between two states, a depressive state with predominantly depressive symptoms (e.g., feeling blue, insomnia) and a manic state with predominantly manic symptoms (e.g., irritability, sexual promiscuity). In this case it would make sense to hypothesise the existence of not 2 but 3 states: one healthy state (e.g., bipolar patient in remission), a depressive state and a manic state. It remains to be seen if one could model a system with three states at all and if so, if the general principles of, for example catastrophe theory (e.g., hysteresis, discontinuous behaviour), would apply.

In this chapter, we studied critical transitions from a healthy state to a disordered state. We also saw that a strongly connected network is indicative of a vulnerability for a critical transition: when a strongly connected network is put under stress, the system may jump from one state to the other, whereas a weakly connected network, when put under stress, may gradually transition. When healthy individuals are being studied, having a strongly connected network is a disadvantage, for those individuals may experience a critical transition from a healthy state to a disordered state. However, a critical transition can always occur in two ways: from healthy to disordered and from disordered to healthy. In the case of depressed patients, a critical transition may occur from a disordered to a healthy state. In this particular case, having a strongly connected network is an advantage: if some intervention would succeed in turning a few symptoms "off" then strong connectivity might result in the triggering of additional symptom deactivation. In future research, it is therefore important to study critical transitions both from a healthy $\rightarrow$ disorder perspective, and from a disorder $\rightarrow$ healthy perspective.

The ultimate goal of the models shown in this chapter is to assess an individual's vulnerability for a critical transition, before the transition itself occurs. It is quite possible that, by the time individuals enter therapy, it is too late for us to assess their vulnerability; for a critical transition may already have occurred. In such a case, the goal of therapy

also transitions from preventing a critical transition to a disordered state, to helping that individual transition back to a healthy state. If we would solely want to prevent critical transitions, we would need to track virtually every individual from the age of 15 onwards in order to assess their vulnerability for a critical transition from a healthy to a disordered state. This would be an impossible task to carry out. Furthermore, many individuals who may never experience mental problems are in this scenario forced to complete questionnaires whose results will never be used. In studying critical transitions, researchers should therefore not only focus on the prevention of critical transition, but also on how critical transitions can be induced, so that therapists can use our models to help patients return to a healthy state faster.

In the empirical examples of both models, missing data were imputed with either zeroes (Cramer model) or with the previous measurement (Empirical MFA). In any type of data collection, but especially with intensive data collection like ESM, participants may fail to fill out certain questions for numerous reasons. The design of the questionnaire may lead to missing data, as certain questions may only be asked when a specific answer is given to some other question, or participants may skip questions and fail to return to these questions. In the case of ESM data, participants may forget their phone or may not simply be in the mood to complete the questionnaire when it is asked of them several times a day. The effect of missing data on the results of a vulnerability analysis is, while writing this chapter, unclear. The methods for dealing with missing data showed in this chapter led to a decrease in item variance, which may result in removing that specific item from the data that is used for the entire analysis. Future research should therefore focus on investigating the effects of the different types of missing data (missing completely at random, such as missing random measurements; missing at random, such as forgetting the phone with which you completed the questionnaires; or missing not at random, such as not being in the mood to complete a questionnaire before having a cup of coffee in the morning) on the results, and how much data can be missing before results become unreliable. A way of gauging the reliability of network parameter estimates that is currently available for

cross-sectional data and that could potentially be used for time-series data, is implemented in R (Epskamp, Borsboom, & Fried, 2018). This software package enables calculating confidence intervals around edge weights by means of bootstrapping these weights. The resulting confidence intervals indicate how safe one is in concluding that the estimated edge weight is reliable (e.g., potentially less reliable if the confidence interval around an edge weight includes zero).

This chapter described two models for studying vulnerability and critical transitions that are currently still in their developmental phase. Even though there are some serious conceptual and technical challenges that need to be overcome (Fried & Cramer, 2017), both the Cramer model and the Empirical MFA show promising results and have the potential to aid clinical psychologists and psychiatrists in both successfully treating patients and preventing relapses by, for example, tracking individuals during therapy in order to find the window of optimal opportunity for such successful interventions. That is, in our view, network modelling of psychopathological phenomena holds the promise of ultimately leading the field of psychopathology into an era of personalised care.

3

Mean field dynamics of stochastic cellular automata for random and small-world graphs

This chapter is submitted as: Waldorp, L. J., & Kossakowski, J. J. (2020). Mean field dynamics of stochastic cellular automata for random and small-world graphs. *Under review at Journal of Mathematical Psychology*.

Abstract

We investigate dynamical properties of networks by reducing such networks to one-dimensional dynamical systems by simplifications. We aim to provide a theoretical framework to explain the discrete transitions of mood connecting ideas from network theory and dynamical systems theory. It was recently shown how networks (graphs) can be used to represent psychopathologies, where symptoms of, say, depression, affect each other and certain configurations determine whether someone could transition into a depression. To analyse changes over time and characterise possible future behaviour is in general rather difficult for large graphs. We describe the dynamics of graphs using one-dimensional discrete time dynamical systems theory obtained from a mean field approximation to stochastic cellular automata (SCA). Often the mean field approximation is used on a regular graph (a grid or torus) where each node has the same number of edges and the same probability of becoming active. We provide quantitative results on the accuracy of using the mean field approximation for the grid and random and small-world graph to describe the dynamics of the SCA. Bifurcation diagrams for the mean field of the different graphs indicate possible phase transitions for certain parameter settings of the mean field. Simulations confirm for different graph sizes (number of nodes) that the mean field approximation is accurate.

3.1 Introduction

People appear to shift moods, sometimes rather suddenly, in a discrete manner (Hosenfeld et al., 2015). Such 'discontinuities' can be the result of relatively small changes in the environment or person. In some cases such transitions in mood are associated to mental disorders like depression. Recently, mental disorders have been described as a network (graph) of interacting symptoms (Borsboom et al., 2011; van Borkulo et al., 2014). For instance, lack of sleep during the night could lead to poor concentration during the day, which in turn could lead to lack of sleep again by worrying that your job may be on the line. Here we use these ideas and model the dynamics of psychopathology networks as stochastic cellular automata. This connects networks and dynamical systems theory to

provide a framework for an explanation for transitions in mood. To analyse the dynamics we use a mean field approximation to stochastic cellular automata where each node is similar in behaviour to all others. We extend known results for the mean field approach in this context to other types of graphs (random and small-world graphs), where the mean field can be interpreted as a weighted average of all nodes in the graph. We provide quantitative results on the error of approximation for the mean field for the different types of graphs.

Cellular automata are discrete dynamical systems that have deterministic, local rules to move from one generation (time point) to the next (Wolfram, 1984b; Sarkar, 2000). Introduced by Von Neumann (1951), the most famous version is Conway's game of life, popularised by Gardner (1970), and has found many applications from computer science (Wolfram, 1984a) to neuronal population modelling (Kozma et al., 2005) to epidemiology (Kleczkowski & Grenfell, 1999). In a cellular automaton each cell or node in a finite grid (usually a subset of the lattice $\mathbb{Z}^2$) can be 'active' or 'inactive' (1 or 0) and if, for instance, two (direct) neighbours are active, then the node will become active at the next time step. Another example of a cellular automaton is bootstrap percolation, where each node can only become active and cannot be inactivated by its neighbours, and the objective is to determine the initial configuration of active nodes that results in all nodes being active (Janson et al., 2012). In general, a new generation in a cellular automaton is determined by a local and homogeneous update rule ϕ. For each node i in the graph this induces a sequence of states, an orbit. A (random) configuration at time $t = 0$ then determines whether all nodes in the network will be active, inactive, or whether the network will demonstrate periodic or perhaps chaotic behaviour. A generalisation of a cellular automaton is to introduce a probability p to ϕ to decide whether or not a node will become active, using by a node's neighbours. One such rule is the majority rule that gives the probability to switch depending on whether the majority of its neighbours are active. Such a system is called a stochastic cellular automaton (SCA). Here we will investigate the dynamic behaviour of the proportion of active nodes (density) for an SCA with a majority rule that is defined on

toroidal, random and small-world graphs.

Many versions of SCA exist, of particular interest are those that behave similar to the Ising network. The reason is that the Ising network is often used to model realistic phenomena, like magnetisation (Kindermann & Snell, 1980; Sethna et al., 2004) or psychopathologies (van Borkulo et al., 2014). We have in mind the application to psychopathology here. In such systems the symptoms of disorders are the nodes in the graph and edges between the symptoms are estimated from data using the Ising model (van Borkulo et al., 2014) or connectivity is obtained from verbal accounts. A possible application of the work presented here is to apply the models to real data analysis using results from discrete Markov chains and assess the location of the person with respect to the possible stable states.

Watts (1999) showed that a one-dimensional, large-scale cellular automaton (deterministic), where the connectivity between nodes was arranged as a small-world, could execute the density (all zeroes) (end up with a majority of active nodes) and synchronisation (eventually alternating between all active and all inactive nodes regardless of the initial configuration) tasks. Newman & Watts (1999b) gave approximations for path length and clustering on a small-world, to obtain an analytic solution to the threshold above, which a large number of active nodes. Callaway et al. (2000) also studied percolation in different graph topologies in deterministic automata, focusing on the consequences of (randomly) deleting nodes. Here again the objective was to concentrate on stable solutions of the graphs. In a stochastic version, Tomassini et al. (2005) investigated a one-dimensional SCA on a regular and small-world graph in terms of its performance on the density and synchronisation tasks. They determined by using evolutionary algorithms that a small-world topology is most efficient (compared to random graphs and lattices) to solving both tasks, corresponding to the results of Watts (1999) in a deterministic version. Their objective was different from ours in that here we are interested in all types of dynamic behaviour (stable or not), and specifically representing this behaviour for the SCA by the mean field.

Our starting point is the work by Balister et al. (2006) and Kozma et al. (2005) where a two-dimensional (toroidal) grid on an SCA is defined. In a two-dimensional grid the nodes are located on the intersecting points of integers in the plane, where at the edges of the grid nodes are connected to the nodes on the opposite end, so that is in fact a torus (see Figure 3.2). The mean field is then used to determine the unconditional probability distribution of the density (relative number of active nodes). Balister et al. (2006) showed that the mean field model predicts a bifurcation for small values of the probability of a node switching to another state and determine its critical point for a neighbourhood of size five (see also Kozma et al., 2004, 2005). This is of particular interest in our case as it may explain mood disorders (e.g., depression or bipolar disorder) from symptoms and their connectivity. To apply these results to random and small-world graphs we determine the marginal distribution across the possible node degree (number of connections of a node) probabilities given the topology of the random or small-world graph. Extending results of homogeneous graphs has been applied to social networks by Barrat et al. (2008) and to cellular automata (Janson et al., 2019).

We first introduce stochastic cellular automata in Section 3.2. Then in Section 3.3 we show how the traditional version of an SCA on a grid can be reduced to a single discrete time dynamical system, called the mean field. In Section 3.3.2 we show that for the random graph we can use a variation on the formulation for the grid of the dynamical system to describe dynamics. We use these results on the random graph to show in Section 3.3.3 that we can obtain a similar approximation for the small-world graph, again using the formulation for the grid. Having shown that these approximations are appropriate, we see in Section 3.4 what the dynamics of the process is for the different graph topologies. We follow these theoretical results by simulations to verify the accuracy of the mean field in Section 3.5. Proofs can be found in Appendix A.

3.2 Stochastic cellular automata

A cellular automaton is a dynamical system consisting of nodes in a fixed, finite grid where directly connected nodes (neighbours) determine the state of a node at each subsequent time step (Wolfram, 1984a). Each node i in a node set $V = \{1, 2, \ldots, n\}$ is at time t in one of the states of a finite alphabet A, referred to as $x_i = a \in A$, where, for instance, $A = \{0, 1\}$ (in this chapter we only consider binary alphabets). The neighbours in the graph $G = (V, E)$ are given by the edges (i, j) in the edge set E. Often the graph G is the square lattice $\mathbb{Z}^2$, where for each node i the neighbourhood $\Gamma(i)$ are the nodes that have an edge to i and includes i as well, that is $\Gamma(i) = \{j : (i, j) \in E\} \cup \{i\}$. A local rule $\phi : A^\Gamma \to A$ assigns to each configuration of states of the nodes in the neighbourhood $x_{\Gamma(i)}$ a value such that $\phi : x_{\Gamma(i)} \mapsto x_i$, where $x_i = a \in A$. For instance, consider the one-dimensional graph (a graph on a line with neighbours on each side) in Figure 3.1, where in each figure the top row represents time point t and the bottom represents time point $t + 1$. The middle node has two neighbours, and application of ϕ is represented by the row below it indicated by $t + 1$. Each node has a value from the binary alphabet $A = \{0, 1\}$. There are three nodes and the local rule ϕ determines from $x_{\Gamma(2)} = (x_1, x_2, x_3)^\intercal$ at time point t the value $\phi(x_{\Gamma(2)})$ at time point $t + 1$. Many different rules for the update function ϕ exist (e.g., see Wolfram, 1984b). In Figure 3.1 we show so-called Rule 30, so named because when the output values are considered as coefficients of the binary expansion, then they represent the number 30. This Rule 30 determines, for instance, that if all nodes in $\Gamma(2)$ are 1, then at the next time point, the value of the middle node is 0. All 8 possible configurations for three nodes in the neighbourhood $\Gamma(2)$ are considered in Rule 30 and are shown in Figure 3.1. Cellular automata are completely determined by a particular mapping chosen for ϕ.

Although (deterministic) cellular automata show extremely interesting and complex behaviour such as periodicity and chaos, we require a stochastic cellular automaton (SCA; e.g., see Paz, 1971) because in psychology we do not have full knowledge of how a symptom changes from inactive to active. Therefore, we make the state of a node $X_{i,t}$ a Bernoulli random variable that is 0 or 1 at time point t. We also define $Z_{i,t} = \sum_{j \in \Gamma(i)} X_{j,t}$

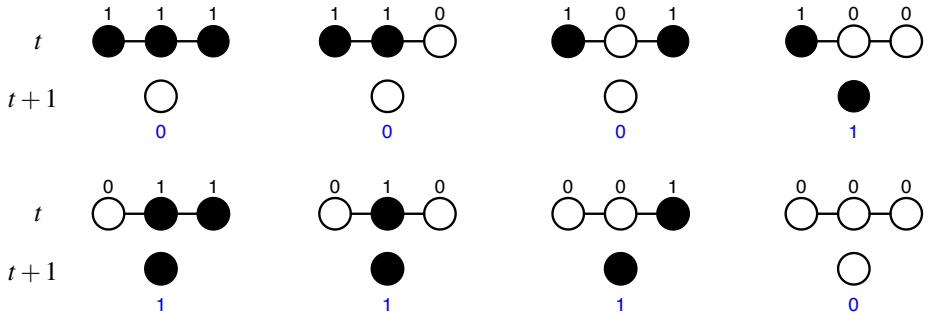


Figure 3.1: The 8 configurations of states $x_{\Gamma(2)}$ (above the three nodes) for the local rule ϕ applied at time t (three nodes at the top) that result in the active or inactive node at time $t+1$ (single node at the bottom with its subsequent state below it). All configurations with the output (at the bottom of each configuration) give the binary expansion $0\cdot 2^7+0\cdot 2^6+0\cdot 2^5+1\cdot 2^4+1\cdot 2^3+1\cdot 2^2+1\cdot 2^1+0\cdot 2^0$ (from left to right) which equals 30, and hence the name Rule 30.

for the sum of active neighbours of node i. In an SCA one of the most common rules to determine the probability of the state of a node at time point $t+1$ is by the majority rule; the majority rule is popular not only because of intuitive appeal, but also because it is stable in the sense that a small number of random flips in the input will not easily lead to different results (O'Donnell, 2014). The majority rule for any node is defined as the probability p of obtaining state 0 if there are fewer than half of the neighbours in state 1 at time point t, and probability $1-p$ of obtaining 1 if there are more than half of the neighbours 1 at time point t. That is,

$$\mathbb{P}(X_{i,t+1}=1 \mid z_{i,t}) = \mathrm{maj}_p(z_{i,t}|\Gamma(i)|) = \begin{cases} p & \text{if } z_{i,t} \leq |\Gamma(i)|/2 \\ 1-p & \text{if } z_{i,t} > |\Gamma(i)|/2 \end{cases} \tag{3.1}$$

where $|\Gamma(i)|$ denotes the cardinality of the neighbourhood set $\Gamma(i)$; we refer to this rule as maj_p. The nodes in graph G are updated simultaneously. The value for node i is obtained by comparing a value u obtained from the uniform distribution on the interval

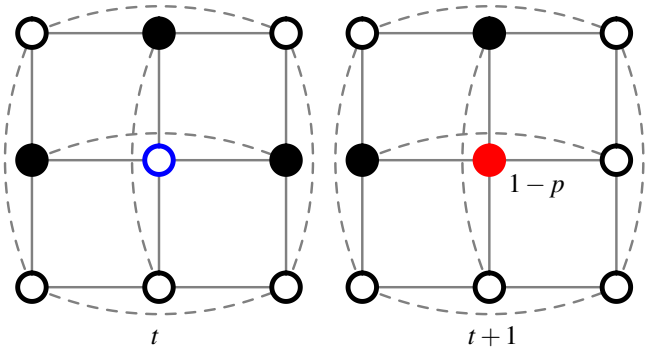


Figure 3.2: Illustration of a two-dimensional grid (lattice) where the middle node had four neighbours and itself at the previous time point (not shown). Each node has the same number of neighbours, and the nodes at the edges of the grid are connected to the nodes on the opposite end (dashed lined), making it a torus. At time point t there are three neighbours active, and at time point $t+1$ the middle node has probability $1-p$ of becoming active.

$[0,1]$ to p or $1-p$, depending on the number of active neighbours $z_{i,t}$. If $u < p$ (or $u < 1-p$) then node i will have state 1 at time $t+1$. As an example, consider a small lattice or grid with boundary conditions (shown in Figure 3.2), such that each node has exactly four other neighbours and itself at time point t. The middle node has three neighbouring nodes in state 1 at time point t and so we obtain the probability $1-p$ of the middle node being in state 1 at time point $t+1$.

We now have an SCA with the majority rule maj_p that can be run and has been shown to lead to complex behaviour. It has proved difficult to describe long term behaviour, like fixed points or bifurcation points, in general for SCA (Kozma et al., 2004). Therefore, approximations have been obtained by assuming homogeneity of the nodes (mean field) such that the long term behaviour of the SCA can be described. Here we follow the approach of Balister et al. (2006) where they use a mean field approach.

3.3 Mean field approximation on graphs

To learn about long term behaviour of the SCA we may turn to the mean field, which uses the assumption of homogeneous nodes, all having the same update rule and same number of neighbours. In doing so we are able to reduce the complicated SCA to a one-dimensional discrete-time Markov chain. This will allow us to analyse its dynamics more easily (see Section 3.4). The key ingredient of the mean field approximation, shown by Balister et al. (2006), is that the properties of interest are uniform over the graph. For the (toroidal) grid topology in two dimensions this is intuitive to see: Any node has the same number of neighbours and each node becomes 0 or 1 by the same local rule (majority rule). We extend results of the grid to the random graph and the small-world graph, where, clearly, the assumption that each node has the same number of nodes is the same is violated. We first consider the case for a grid and then move on to the random and small-world graph.

3.3.1 Mean field on a grid

Let the graph $G_{\text{grid}}(n,\Gamma)$ be a grid with n nodes and boundary conditions such that each node has exactly the same number of neighbours. The mean field approximation for the grid assumes that any set of nodes (of the size of the neighbourhood) could serve as part of the neighbourhood, that is, we have a generic set of nodes Γ from the graph G_{grid} instead of the exact neighbourhood $\Gamma(i)$ for node i because each node has the same number of nodes and the same update rule. For instance, in the two-dimensional lattice $\mathbb{Z}^2$ the number of neighbours is then 5 (including the node itself). This implies that the number of active nodes in Γ, referred to as $Z_t = \sum_{j\in\Gamma} X_{j,t}$ (without the subscript i), depends on the number on the number of nodes in state 1. We call the number of active nodes in the entire graph Y_t defined by $\sum_{j\in V} X_{j,t}$, the sum of all $|V| = n$ nodes in the graph G_{grid}. In order to know the evolution of Y_t over time we need to know the transition probability $\mathbb{P}(y_{t+1} \mid y_t)$ in the setting of the SCA with the majority rule. Since $X_{i,t}$ are

Bernoulli random variables, Y_t is binomial of size n and with the success probability the probability of $X_{i,t}$, which depends on the number of active nodes in the neighbourhood Γ. The probability of obtaining an active node in the generic neighbourhood Γ is determined by y_t/n, which is often referred to as the density and denoted by ρ_t. So, in the mean field we have the probability $\mathbb{P}(X_{i,t} = 1 \mid z_t, \rho_t)$, and we require the probability $\mathbb{P}(X_{i,t} = 1 \mid \rho_t)$, which demands that we know $\mathbb{P}(Z_t = r \mid \rho_t)$ so that we can marginalise over the possible values $r = 0, 1, \ldots, |\Gamma|$ for the number of nodes in state 1. The probability of r nodes in state 1 in the neighbourhood Γ is

$$\mathbb{P}(Z_t = r \mid \rho_t) = \binom{|\Gamma|}{r} \rho_t^r (1-\rho_t)^{|\Gamma|-r}$$

Then the probability of the conditional event $X_{i,t} \mid \rho_t$, that node i will have state 1 at time $t+1$ given the density ρ_t at time t, is obtained by marginalising over all possible numbers of active neighbours r, i.e., the marginal $\mathbb{P}(x_{i,t+1} \mid \rho_t)$ is

$$\sum_{r=0}^{|\Gamma|} \mathbb{P}(x_{i,t+1} \mid z_t, \rho_t) \mathbb{P}(z_t \mid \rho_t) = \sum_{r=0}^{|\Gamma|} \mathrm{maj}_p(r, |\Gamma|) \binom{|\Gamma|}{r} \rho_t^r (1-\rho_t)^{|\Gamma|-r}. \tag{3.2}$$

This marginal can be written in a more intuitive way where the majority function is directly applied as a function of the number of active nodes r (see also Kozma et al., 2005)

$$\sum_{r=0}^{\lfloor |\Gamma|/2 \rfloor} p \binom{|\Gamma|}{r} \rho_t^r (1-\rho_t)^{|\Gamma|-r} + (1-p) \sum_{r=\lceil |\Gamma|/2 \rceil}^{|\Gamma|} p \binom{|\Gamma|}{r} \rho_t^r (1-\rho_t)^{|\Gamma|-r},$$

where $\lfloor x \rfloor$ is the greatest integer less than or equal to x and $\lceil x \rceil$ is the least smallest integer greater than or equal to x. We call this marginal probability

$$p_{\mathrm{grid}}(\rho_t) = \mathbb{P}(x_{i,t+1} \mid \rho_t) \tag{3.3}$$

where the marginal was over the different number of active neighbours $z_t \in \{0, 1, \ldots, |\Gamma|\}$. As Balister et al. (2006) showed, it follows that the number of nodes in state 1 in graph

G_{grid} is then a Binomial random variable. Let $B(n,p)$ denote a binomial random variable with n Bernoulli trials each with success probability p.

Lemma 1 *(Balister et al., 2006, Theorem 2.1) Let $G_{grid}(n,\Gamma)$ be a grid with an SCA as defined above with majority rule maj_p in (3.1) and $\rho_t = Y_t/n$ be the proportion of nodes in state 1 at time t. Let the marginal probability $p_{grid}(\rho_t) = \mathbb{P}(x_{i,t+1} \mid \rho_t)$ in (3.3) be denoted by p_{grid}. Then the evolution of the number of active nodes Y_t on G_{grid} is*

$$Y_{t+1} = B(n, p_{grid}(\rho_t)). \tag{3.4}$$

The mean and variance for the density $\rho_t = Y_t/n$ are, respectively, $\mu_{grid} = p_{grid}$ and $\sigma^2_{grid} = p_{grid}(1-p_{grid})/n$ for any t.

In this binomial process at each time step the proportion $\rho_t = Y_t/n$ is determined, after which the function p_{grid} in (3.3) is applied. This provides a new probability for the binomial process at each time point. Because we are considering an averaging process, the mean field, we are interested in the repeated application $p_{\text{grid}} \circ \cdots \circ p_{\text{grid}}(\rho_t)$, where the composition $p_{\text{grid}} \circ p_{\text{grid}}(\rho_t)$ is defined by $p_{\text{grid}}(p_{\text{grid}}(\rho_t))$. If the mean p_{grid} is a good description of the process (deviations from p_{grid} are not too large), then we can use this mean field as an accurate description of the binomial process. We are then interested in the evolution of p_{grid}. First we consider the accuracy of the approximation.

For each t, $X_{i,t+1}$ is Bernoulli distributed $B(1, p_{\text{grid}}(\rho_t))$ for all nodes $i \in V$, and the number of active nodes Y_{t+1} is the sum of the outcome of Bernoulli trials. Hence, we can apply the law of large numbers so that for large graphs (large n), the mean of the proportion of active nodes $\mu_{\text{grid}} = p_{\text{grid}}$ is close to ρ (where we ignored the subscript t for the moment since this holds for any t) with high probability. Indeed, we can use Chernov's bound to obtain that for any t the difference $|\rho - p_{\text{grid}}|$ is bounded as a function of the size of the graph $|V| = n$ (a proof is in Appendix A).

Lemma 2 *(Accuracy bound of density) Let $Y_t = \sum_{x \in V} X_{i,t}$ be the sum of n Bernoulli trials given by (3.4), with mean of the density $p_{grid}(\rho_t)$. For every $0 < \varepsilon < \min\{p_{grid}, 1 - p_{grid}\}$, let $\delta = 2\exp(-\varepsilon^2/2\sigma^2_{grid})$. We then have with probability at least $1 - \delta$*

$$|\rho - p_{grid}| \leq \sqrt{\frac{p_{grid}(1 - p_{grid})}{n} 2\log(2/\delta)} \tag{3.5}$$

For instance, we obtain the interval with probability at least 0.95 of $[\mu_{\text{grid}} - 2.72\sigma_{\text{grid}}, \mu_{\text{grid}} + 2.72\sigma_{\text{grid}}]$. Another interval can be obtained from the DeMoivre-Laplace central limit theorem. This theorem tells us that for large enough n, $W_{\text{grid}} = (\rho - p_{\text{grid}})/\sigma_{\text{grid}}$ is distributed as $N(0,1)$. In fact, if the third order moment of W_{grid} is $c < \infty$, the Berry-Esseen theorem says that the order of approximation of the distribution of ρ to the normal distribution is $O(3c/\sqrt{n})$ (Venkatesh, 2012). This provides an interval for ρ_{t+1} as a measure of accuracy with $[\mu_{\text{grid}} - 1.96\sigma_{\text{grid}}, \mu_{\text{grid}} + 1.96\sigma_{\text{grid}}]$ with probability 0.95. Clearly, in both limit laws the size of the network n determines the accuracy of the approximation.

3.3.2 Mean field on a random graph

In the original setting of a grid (with boundary conditions) the number of neighbours is fixed and it was seen that the mean field approximation p_{grid} was accurate for the density because each node is identical with respect to a change depending on its neighbours. Here we introduce the neighbourhood size $|\Gamma|$ as a random variable such that we obtain the probability $\mathbb{P}(X_{i,t+1} \mid |\Gamma| = k, \rho_t)$. Then we determine the probability of $X_{i,t+1} \mid \rho_t$ by averaging over all possible sizes of neighbourhoods $k = 0, \ldots, n-1$ weighted by its probability for neighbourhood size (marginalising).

This is done in a random graph where each node has a binomial number of neighbours. Let $G_{\text{rg}}(n, p_e)$ be a random graph with n nodes and (constant) probability p_e of an edge being present, independently (Bollobás, 2001; Durett, 2007). Let the size of the neighbourhood $|\Gamma|$ be a binomial random variable with maximal value $n-1$ neighbours and probability p_e, that is, $B(n-1, p_e)$. Then the probability of obtaining an active node

can be defined conditionally on the event $\{|\Gamma| = k\}$, the neighbourhood having size k. That is, we obtain the probability for each neighbourhood size

$$\mathbb{P}(|\Gamma| = k) = \binom{n-1}{k} p_e^k (1-p_e)^{n-k-1}.$$

Marginalising over the possible neighbourhood sizes, we obtain p_{rg} for the probability of a node being active in the binomial process for the random graph G_{rg}. Then the probability of obtaining $X_{i,t+1}$ is

$$p_{\mathrm{rg}}(\rho_t) = \sum_{k=0}^{n-1} \mathbb{P}(X_{i,t+1} \mid |\Gamma| = k, \rho_t) \mathbb{P}(|\Gamma| = k)$$

Then using (3.2) obtaining

$$\mathbb{P}(x_{i,t+1} \mid |\Gamma| = k, z_t = r, \rho_t) = \mathrm{maj}_p(r,k) \binom{k}{r} \rho_t^r (1-\rho_t)^{|\Gamma|-r},$$

we obtain the probability $\mathbb{P}(X_{i,t+1} \mid \rho_t)$ for the random graph

$$p_{\mathrm{rg}}(\rho_t) = \sum_{k=0}^{n-1} \sum_{r=0}^{k} \mathrm{maj}_p(r,k) \binom{k}{r} \rho_t^r (1-\rho_t)^{k-r} \binom{n-1}{k} p_e^k (1-p_e)^{n-k-1}. \tag{3.6}$$

This is a rather lengthy expression and is not easy to work with. So we aim to approximate this expression and determine the error of approximation. Intuitively we would expect that we could replace the sum over all neighbourhood sizes k in (3.6) and take only the expected neighbourhood size $\lfloor p_e(n-1) \rfloor$ for each node, where $\lfloor a \rfloor$ is the integer part of a. Then we obtain a result for the average neighbourhood size, which should be reasonably close. This would result in a simpler formulation of the probability of a node to be active in a random graph and would make determining fixed points easier and simplify computation for large graphs considerably. We next show that such an approximation is a reasonable approach (the proof is in Appendix A).

Proposition 3 *(Probability on a random graph) If the neighbourhood in $G_{rg}(n, p_e)$ of each node in an SCA as defined above is fixed with the expected number of nodes under the random graph $\nu = \lfloor p_e(n-1) \rfloor$ such that the neighbourhood size $|\Gamma|$ is replaced by ν in the majority rule maj_p in p_{grid} in (3.2), then the probability p_{rg} can be approximated by*

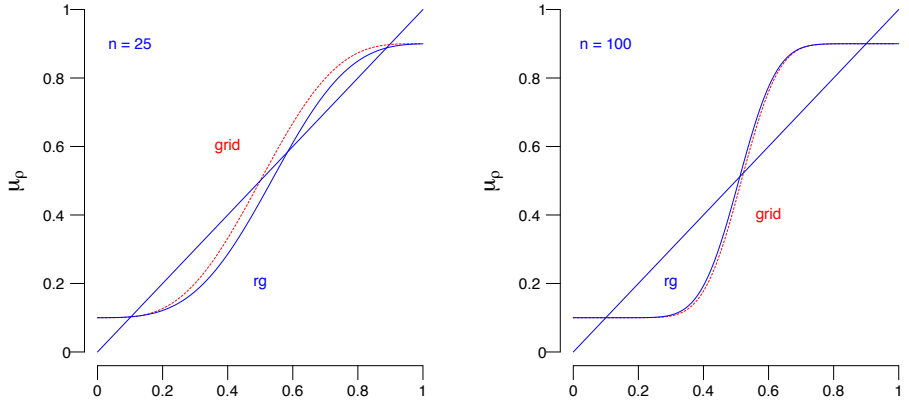


Figure 3.3: The expectation $\mu_{\text{rg}} = p_{\text{rg}}$ (blue, solid curve) of equation (3.6) and μ^{ν}_{grid} (red, dotted curve) of equation (3.7) with $p = 0.1$ and $p_e = 0.3$. Left panel shows the curves for a graph of size $n = 25$, showing a clear difference between the curves, and the right panel for graph size $n = 100$. Note that the difference between the curves at the crossings with the 45° line is small.

$$p^{\nu}_{grid}(\rho_t) = \sum_{r=0}^{\nu} maj_p(r,\nu)\binom{\nu}{r}\rho_t^r(1-\rho_t)^{\nu-r}. \tag{3.7}$$

The approximation error is

$$|p_{rg} - p^{\nu}_{grid}| \leq 2|p-1/2|\exp\left(-\frac{(n-1)\varepsilon^2}{p_e(1-p_e)} + \log(n)\right)$$

for any $\varepsilon > 0$ *and with* $0 < p_e < 1$.

This shows that the number of active nodes Y_t in the random graph with probability p_{rg} in (3.6) and the number of active nodes with probability p^{ν}_{grid} in (3.7) converge in probability with exponential rate with graph size n. Figure 3.3 illustrates the difference between p_{rg} and p^{ν}_{grid} for graph sizes $n = 25$ and $n = 100$.

Remark To retain a probability of an edge in p^{ν}_{grid} leads to a larger approximation error, i.e., using

$$p^{n-1}_{\text{grid}}(\rho_t) = \sum_{r=0}^{n-1} \text{maj}_p(r, n-1) \binom{n-1}{r} (\rho_t p_e)^r (1-\rho_t p_e)^{n-r-1} \tag{3.8}$$

leads to an error of at most $|p-1/2|$, which makes it non-ignorable (see Appendix A).

We now have expression (3.7) similar to (3.2) for a random graph with the probability of an active node at time t determined by both the density ρ_t and an edge being present p_e in the size of the neighbourhood. From (3.6) and Lemma 1 and Proposition 3 the evolution equation for the random graph can also be described by a binomial process with probability (3.7)

$$Y_{t+1} = B(n, p^{\nu}_{\text{grid}}(\rho_t)). \tag{3.9}$$

We immediately have that the probability p^{ν}_{grid} is close to the density ρ for each time point t for large graph size n. In fact, we find by the triangle inequality

$$|\rho - p^{\nu}_{\text{grid}}| \leq |\rho - p_{\text{rg}}| + |p_{\text{rg}} - p^{\nu}_{\text{grid}}|,$$

both terms on the right hand side converge to 0. The first term $|\rho - p_{\text{rg}}|$ converges to 0 by the Chernov bound (Lemma 2) with p_{rg}, and $|p_{\text{rg}} - p^{\nu}_{\text{grid}}|$ converges to 0 by Proposition 3.

Note that we require for obvious reasons that the random graph is connected. It follows that we need a minimum probability p_e such that the graph is connected. The probability that a random graph G_{rg} is connected is $\exp(-\exp(-\lambda))$, where $p_e = (\log n + \lambda + o(1))/n$ with λ fixed (Bollobás, 2001, Theorem 7.3). For instance, if we choose the probability of G_{rg} being connected to be 0.99 and we use $n = 50$, then we obtain $\lambda = 4.6$ and hence $p_e = 0.17$. We can therefore not go below 0.17 for a graph with $n = 50$ nodes.

3.3.3 Small-world graph

More interesting in many real-world applications is the small-world graph, and so we perform a similar analysis for the small-world graph as for the random graph. A small-world graph is one which has high average clustering and low average path length, relative

to a random graph with the same number of nodes and edges. These graphs have been shown to model realistic networks like those of working relations between actors and the nerve cells in the worm C. elegans (Watts & Strogatz, 1998), and subsequently the small-world has been shown to apply to many different networks, like the (parcellated) brain (Sporns & Honey, 2006). Most recently, the network of symptoms as defined by the diagnostic statistical manual (a compendium to diagnose patients) has been found to be a small-world. This finding is a possible explanation for the correlations between pairs of symptoms found in different sub-populations (Borsboom et al., 2011).

Here we use the modified Newman-Watts (NW) small-world of Newman & Watts (1999b), where for a given grid structure where each node has neighbourhood Γ, a set of $(n-1)p_w$ edges is on average independently added to the graph, where p_w is the probability of two nodes being wired. Such a graph is denoted by $G_{\mathrm{sw}}(n,\Gamma,p_w)$. The same idea as with the random graph, where the probability for an active node was corrected by the probability of the degree of a node, averaged over all possible neighbourhood sizes, can be used for the random part in the NW small-world. In the NW small-world we start with a grid with neighbourhood size $|\Gamma|$, which is fixed, and augment the graph randomly with edges according to a binomial variable with probability p_w. We then obtain

$$p_{\mathrm{sw}}(\rho_t) = \sum_{k=|\Gamma|}^{n-1} \sum_{r=0}^{k} \mathrm{maj}_p(r,k) \binom{k}{r} \rho_t^r (1-\rho_t)^{k-r} \binom{n-1}{k} p_w^k (1-p_w)^{n-k-1}. \tag{3.10}$$

We could define the small-world probability using this definition. But we can split up p_{sw} in two terms, one involving the fixed neighbourhood Γ of the grid, and one random neighbourhood consisting of the possible shortcuts. We therefore start with the probability in a grid p_{grid} corrected by the $(1-p_w)^{n-|\Gamma|}$ requiring that no possible randomly added edges are present, i.e., we obtain

$$p_{\mathrm{grid}}^{\mathrm{sw}} = p_{\mathrm{grid}}(1-p_w)^{n-|\Gamma|} \tag{3.11}$$

for the first part of the fixed grid with neighbourhood Γ. Then, in accordance with the random part of the NW small-world, a probability is added to emulate the possible addi-

tional neighbours in the random part of the graph, ignoring the first $|\Gamma|$ neighbours from the grid. We obtain this part directly from Proposition 3 for the remaining $n-|\Gamma|$ nodes. Define the probability

$$p_{\mathrm{rg}\setminus\Gamma}(\rho_t) = \sum_{k=|\Gamma|+1}^{n-1} \sum_{r=0}^{k} \mathrm{maj}_p(r,k) \binom{k}{r} \rho_t^r (1-\rho_t)^{k-r} \binom{n-1}{k} p_w^k (1-p_w)^{n-k-1}, \quad (3.12)$$

where the first $|\Gamma|$ neighbours are ignored since they were included already as neighbours in the grid structure in $p^{\mathrm{sw}}_{\mathrm{grid}}$. For the second part $p_{\mathrm{rg}\setminus\Gamma}$, however, we have the approximation as before from the random graph, leaving out the first Γ nodes from the grid. This leads to the simplification using the grid probability only

$$p^{\nu}_{\mathrm{grid}\setminus\Gamma}(\rho_t) = \sum_{r=|\Gamma|+1}^{\nu} \mathrm{maj}_p(r,\nu) \binom{\nu}{r} (\rho_t)^r (1-\rho_t)^{\nu-r}, \quad (3.13)$$

where $\nu = \lfloor p_w(n-|\Gamma|) \rfloor$. The error of approximation using $p^{\nu}_{\mathrm{grid}\setminus\Gamma}$ instead of $p_{\mathrm{rg}\setminus\Gamma}$ follows immediately from Proposition 3 for fixed grid neighbourhood Γ, except that the first $|\Gamma|$ nodes in the grid are taken out.

Corollary 5 *(Probability on a NW small-world) Let $G_{sw}(n,\Gamma,p_w)$ be the NW small-world graph of size n with $|\Gamma|$ nodes in the fixed neighbourhood for each node. Furthermore, let $0 < p_w < 1$ be the wiring probability and $\nu = \lfloor p_w(n-|\Gamma|) \rfloor$. Then the approximation error for the probability using the grid structure p^{ν}_{grid} in (3.13) in the random part is*

$$|p_{rg,\setminus\Gamma} - p^{\nu}_{grid,\setminus\Gamma}| \leq 2|p-1/2| \exp\left(-\frac{(n-|\Gamma|)\varepsilon^2}{p_w(1-p_w)} + \log(n-|\Gamma|+1) \right),$$

for any $\varepsilon > 0$.

Equations (3.10) to (3.13) and Corollary 5 prove that the equation for the evolution on an NW small-world is

$$Y_{t+1} = B(|\Gamma|, p^{\mathrm{sw}}_{\mathrm{grid}}(\rho_t)) + B(n-|\Gamma|, p^{\nu}_{\mathrm{grid}\setminus\Gamma}(\rho_t)) \quad (3.14)$$

We write

$$p^{\nu}_{\mathrm{sw}} = \frac{|\Gamma|}{n} p^{\mathrm{sw}}_{\mathrm{grid}} + \frac{n-|\Gamma|}{n} p^{\nu}_{\mathrm{grid}\setminus\Gamma} \quad (3.15)$$

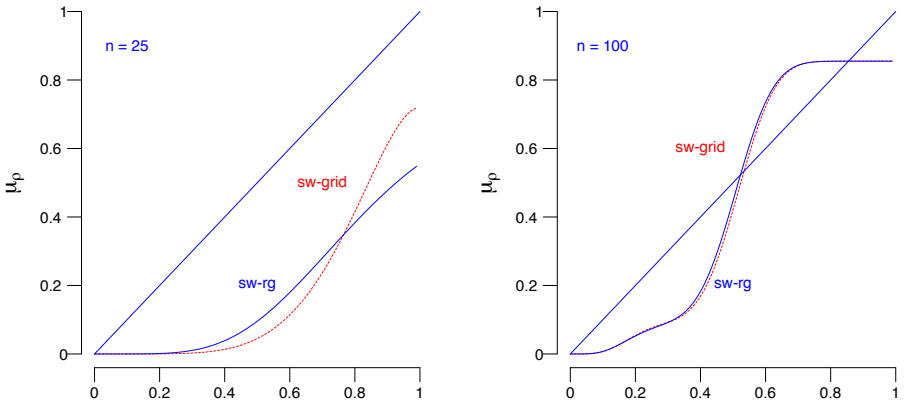


Figure 3.4: The expectation p_{sw} (blue, solid curve) of equation (3.10) and p_{sw}^{V} (red, dotted curve) of equation (3.13) with $p = 0.1$ and $p_w = 0.3$. The left panel shows the curves for a graph of size $n = 25$, showing a clear difference between the curves, and the right panel for graph size $n = 100$. Note that the difference between the curves at the crossings with the 45° line is small.

for the NW small-world probability based on the approximation with the grid. Figure 3.4 shows two examples of the approximation p_{sw}^{V} for the NW small-world. It is clear from Corollary 5 that convergence is a bit slow for small graphs since the difference of nodes in the fixed neighbourhood Γ and in the expected $p_w(n - |\Gamma|)$ neighbours in the random part, determines the rate. But again, we can use p_{sw}^{V} to determine the dynamics of the mean field for large graphs.

3.4 Dynamics properties

We are interested in the long run behaviour of p_{grid} because this mean field describes what the number of active nodes Y_{t+1} in (3.4) will be, which is a binomial process $B(n, p_{\mathrm{grid}}(\rho_t))$. We use the grid as a basis to generalise to both the random graph and the NW small-world graph, because the probability for a random and NW small-world

graph have been shown to be simple extensions of the probability on a grid.

For the process Y_t we know from that the transition probabilities are

$$\mathbb{P}(Y_{t+1} = r \mid Y_t = k) = \binom{n}{r} p_{\text{grid}}(\rho_t)^r (1 - p_{\text{grid}}(\rho_t))^{n-r} \tag{3.16}$$

where $\rho_t = k/n$. It is easily seen that this is a discrete time Markov process on a finite state space of size n, since the probability of Y_{t+1} depends only on Y_t. It is also clear that for each time point t the transition probability is different, i.e., we are dealing with an inhomogeneous Markov process. For graph G of size n we have $n+1$ states, and because we define the transition probability in (3.16) using the time dependent value $\rho_t = Y_t/n = \sum_{j=1}^{n} X_{j,t}/n$ in the function p_{grid}, we obtain a different transition probability at each t. Following Saloff-Coste & Zuninga (2010), we denote the transition probability in (3.16) by $K_t(k,r) = \mathbb{P}(Y_{t+1} = r \mid Y_t = k)$ (also called a Markov kernel), and denote the $(n+1) \times (n+1)$ matrix of transition probabilities by K_t.

By Lemma 2 we know that most of the proportions $\rho_t = Y_t/n$ will be around p_{grid}. Hence, if we have some form of stability of the long run distribution of the Markov process, we can gauge the probable locations by considering p_{grid}. Therefore, we investigate the stability properties of the process Y_t and then consider (qualitatively) the probability for the proportion of active nodes.

3.4.1 Stability

For a time-independent (homogeneous) transition Markov kernel $K(r,k) = \mathbb{P}(Y_{t+1} = r \mid Y_t = k)$, for any t, the Markov chain at time t is $\pi_t = \pi_0 K^t$, for initial distribution π_0 over the $n+1$ states. Recall that if K is irreducible, then π_t is essentially independent of the initial distribution, and there exists a unique invariant distribution π such that $\pi K = \pi$. If in addition K is aperiodic, then for large t the distribution π is well approximated by π_t (Norris, 1997; Levin et al., 2017). Similar (but weaker) results can be obtained with inhomogeneous Markov chains (Saloff-Coste & Zuninga, 2009). For an inhomogeneous Markov chain of finite state, we are concerned with a dissociation from the initial distri-

bution (merging) and whether at some time point the Markov chain becomes stable (the distribution remains within a certain range for each state). Note that probabilities $\pi_t(r)$ need not be stationary, and we do not always expect that such stability will be obtained (e.g., see Saloff-Coste & Zuninga, 2010, for examples).

First, we consider whether the distribution with the Markov chain obtained by the product

$$\pi_t = \pi_0 K_{0,t} = \pi_0 K_0 K_1 \cdots K_t$$

where π_0 is the initial distribution at time $t = 0$, will be the same for any other initial distribution π_0'. This is sometimes referred to as 'merging' because, if we consider the total variation distance $||P - G||_{\mathrm{TV}} = \frac{1}{2}\sum_x |P(r) - G(r)|$ for $r \in \{0, 1, 2, \ldots, n\}$ and two probability measures P and G, then

$$||\pi_0 K_{0,t} - \pi_0' K_{0,t}||_{\mathrm{TV}} \to 0 \quad \text{as} \quad t \to \infty$$

implies that two different initial distributions result in the same distribution π_t (Saloff-Coste & Zuninga, 2010); in other words, π_t is for large t independent of the initial distribution π_0. So, independently of where we start, π_0 or any other π_0', we will end up with the same (possibly time varying) distribution π_t (Levin et al., 2017). Merging is also sometimes called weak ergodicity (Mott & Schneider, 1957).

For merging (weak ergodicity) we require that $\pi_0 K_{0,t}$ is close to $\pi_0' K_{0,t}$ in total variation for large enough t. This is the case if the transition kernel $K_{0,t}$ is contracting in total variation distance. That is, if

$$||\pi_0 K_{0,t+s} - \pi_0' K_{0,t+s}||_{\mathrm{TV}} \leq ||\pi_0 K_{0,t} - \pi_0' K_{0,t}||_{\mathrm{TV}}$$

for any t and s. This is because for the single step transition $K_{0,1}$ we have by the triangle inequality

$$||\pi_0 K_{0,1} - \pi_0' K_{0,1}||_{\mathrm{TV}} \leq \frac{1}{2} \sum_{k \in V_0} \sum_{r \in V_0} K_{0,1}(k,r) |\pi_0(k) - \pi_0'(k)|$$

where $\pi_0(k)$ is the probability of state k at time point 0 and $V_0 = \{0,1,\ldots.n\}$. Because $K_{0,1}(k,r) = \mathbb{P}(X_{i,1} = r \mid X_{i,0} = k)$, we have that $\sum_{r\in V_0} K_{0,1}(k,r) = 1$, which gives the result when recursively applying the above inequality. It is clear that if the kernels $K_{t,t+1}(k,r)$ were too small, the inequality is not true, and we would not have a contracting Markov chain. Hence, this result from Saloff-Coste & Zuninga (2010, Thm. 4.3) about merging, assumes that each of the K_t in the product $K_0K_1\cdots K_t$ is irreducible. In fact, we require that for any time point t the elements of the transition kernel $K_{t,t+1}(k,r) = \mathbb{P}(Y_{t+1} = r \mid Y_t = k)$ in (3.16) are $> \eta_t$, for some $\eta_t > 0$ and states k and r (uniform irreducibility). The transition kernel $K_{t,t+1}(k,r)$ is 0 only if p_{grid} is 0. Hence, we obtain the fact that the Markov kernel $K_{t,t+1}$ on the finite state space $\{0,1,\ldots,n\}$ is irreducible if the orbit $p^t_{\text{grid}}(\rho_0)$ will not become 0 or 1 at some point. This requires that the parameter p in the majority rule maj_p in (3.2) is such that p_{grid} cannot be in the stable set $S(0) = \{\rho : p^t_{\text{grid}}(\rho_0) = 0 \text{ for any } t\}$ or similarly in $S(1) = \{\rho : p^t_{\text{grid}}(\rho_0) = 1 \text{ for any } t\}$.

Second, with the same irreducibility assumption, we obtain stability. We obtain that the probabilities are for all states r in $\pi_t(r)$ in an interval determined by the smallest and largest transition probabilities. This stability implies that the probabilities settle down to some specified interval, but not a specific value.

Lemma 6 *Let Y_{t+1} in (3.4) with mean field probability p_{grid} in (3.3) be the binomial process. Suppose that the parameter p in the function p_{grid} is not in the stable set $S(0)\cup S(1)$, and hence that for each t the matrices K_t are uniformly irreducible, i.e., for all states k and r, $K_t(k,r) > \eta_t$. Then we have the following.*

(1) The chain K_t for large t is merging, i.e., for any initial distribution the chain will approximately be π_t.

(2) The probabilities $\pi_t(r)$ are are stable, i.e., within the bounds

$$\min_{k,m} K_t(k,m) \leq \pi_t(r) \leq \max_{k,m} K_t(k,m).$$

This result tells us that from any initial distribution π_0 we will get to a distribution π_t, and these probabilities lie within a certain interval. This is true for some but not all values p in the majority rule maj_p. We must therefore investigate for which values of p in the majority rule the mean field process p_{grid} moves to either 0 or 1, in which case we know that the merging and stability results do not hold. We therefore investigate next the orbits of the mean field p_{grid}.

3.4.2 Dynamics of the mean field

The dynamics of the mean field in the grid G_{grid} from (3.2) have been described in Balister et al. (2006) and Kozma et al. (2005) for a neighbourhood size of $|\Gamma|=5$. The probability function $p_{\text{grid}} : [0,1] \to [0,1]$ defined in (3.2) is continuous. And so, since $[0,1]$ is closed and bounded, we find that p_{grid} has at least one fixed point in $[0,1]$ (Holmgren, 1996; Hirsch et al., 2004). A fixed point is one where we find $p_{\text{grid}}(\rho_t) = \rho_t$. Finding the fixed points for p_{grid} is generally not trivial. Balister et al. (2006) showed that if $|\Gamma|=5$ in the finite grid, then $p = 7/30 \approx 0.233$ is a critical (bifurcation) point, such that if p is in $[7/30, 1/2]$ then there is a stable fixed point at $\rho = 0.5$, but when $p < 7/30$ then $\rho = 0.5$ is unstable and there are two other stable fixed points. This can be seen in Figure 3.5, which shows two bifurcation plots, where for each value of $0 < p \leq 0.5$ the function $\mu_{\text{grid}} = p_{\text{grid}}$ is iteratively applied for 1000 steps, and only the last 50 are plotted. Figure 3.5 (left panel) shows that for $|\Gamma|=5$ neighbours the fixed point is at $7/30$, and shows bi-stability for $p \in [0, 7/30)$ and unistability for $p \in [7/30, 0.5]$, as predicted. Since p_{grid} is continuous, stability can be checked by considering the derivative $\partial p_{\text{grid}}/\partial\rho = \dot{p}_{\text{grid}}$. If $|\dot{p}_{\text{grid}}|$ is bounded by 1, then the fixed point ρ is attractive, otherwise it is repellent. The derivative with respect to ρ_t is

$$\dot{p}_{\text{grid}}(\rho_t) = \sum_{r=0}^{|\Gamma|} \text{maj}_p(r, |\Gamma|) \binom{|\Gamma|}{r} (r - \rho_t|\Gamma|) \rho_t^{r-1} (1-\rho_t)^{|\Gamma|-r-1}$$

For example, the derivative for $p = 0.15$ is not bounded by 1 for all values of ρ. The fixed point $\rho = 0.5$ is repellent since at this point $\dot{p}_{\text{grid}}(0.5) \approx 1.359$, and so iteration of

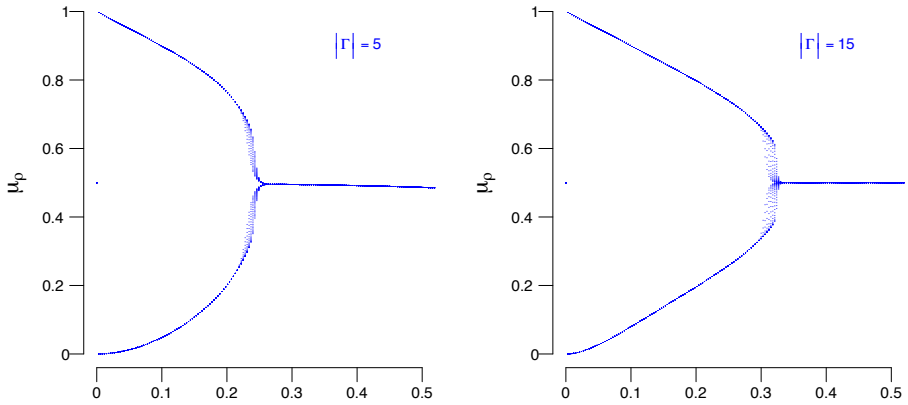


Figure 3.5: Bifurcation plots of μ_{grid} in a graph of size $n = 100$ for $|\Gamma| = 5$ (left) and $|\Gamma| = 15$ (right) neighbours.

p_{grid} will lead away from 0.5. The derivative for $p = 0.35$ is smaller than 1 (≈ 0.672) and so $\rho = 0.5$ is an attractive fixed point. It can be seen that for $|\Gamma| = 5$ the critical (bifurcation) point is at 0.233, as predicted by theory (Balister et al., 2006). It can also be seen that increasing the neighbourhood size to $|\Gamma| = 15$ (right panel of Figure 3.5) increases the critical point to about 0.32. This increase in critical point corresponds to the simulations in Kozma et al. (2005) where ('long range') edges were added to the nodes, which increased the neighbourhood size.

The dynamics of p^{ν}_{grid} in the random graph G_{rg} are similar to that of the grid. The main difference is that the critical point of the bifurcation is closer to $p = 0.5$. This follows from the previous section where we saw that the critical point increases when the neighbourhood size is increased. As is clear from the definition of p^{ν}_{grid} in (3.7), the only difference with that of the grid is the neighbourhood size which is increased to $\nu = \lfloor p_e(n-1) \rfloor$. Figure 3.6 shows the result for a graph with $n = 25$ nodes (left panel) and for a graph with $n = 100$ nodes (right panel). The approximation of p^{ν}_{grid} is quite accurate, also for the location of the critical point. With a graph of size $n = 100$ the accuracy is such that p_{rg} and p^{ν}_{grid} are nearly indistinguishable, which corresponds to the

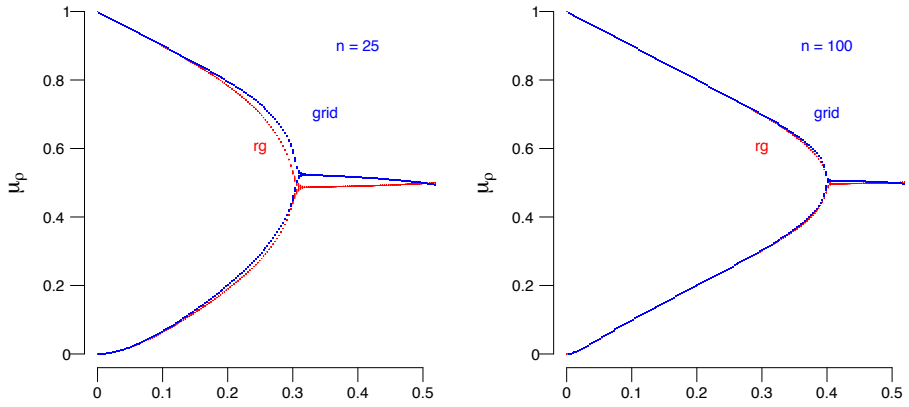


Figure 3.6: Plots of μ_{rg} (red) and μ^{v}_{grid} (blue) as a function of p. In the left panel the bifurcation plots are given for a random graph of size $n = 25$ and in the right panel for a random graph of size $n = 100$. All plots are obtained with edge probability $p_e = 0.4$ to guarantee connectedness of the graph.

result in Proposition 3.

The dynamic behaviour of p_{sw} is shown in Figure 3.7. Generally, the behaviour is similar to that on the random graph. In Figure 3.7 the left panel shows a bifurcation plot of p_{sw} and p^{v}_{sw} on $G_{\text{sw}}(49, 0.4)$. The accuracy of p^{v}_{sw} improves greatly for larger n, as seen in the right panel of Figure 3.7 for $G_{\text{sw}}(100, 0.4)$. For low values of new edges in the NW small-world p_w, the probability p^{v}_{sw} is smaller than in the grid. This is because the probability $p^{\text{sw}}_{\text{grid}} = p_{\text{grid}}(1 - p_w)^{n-|\Gamma|}$ is corrected by the number of edges not added to the graph.

In each case we see that for the interval $\rho_t \in (0, 0.5]$ the proportion is either stable or the proportion is bistable. This implies that we expect the binomial process (3.14) to have most probability for $p > p_c$ around a single stable fixed point or when $p < p_c$ most probability is around one of the two stable fixed points. As mentioned above, determining the critical value p_c is obtrusive.

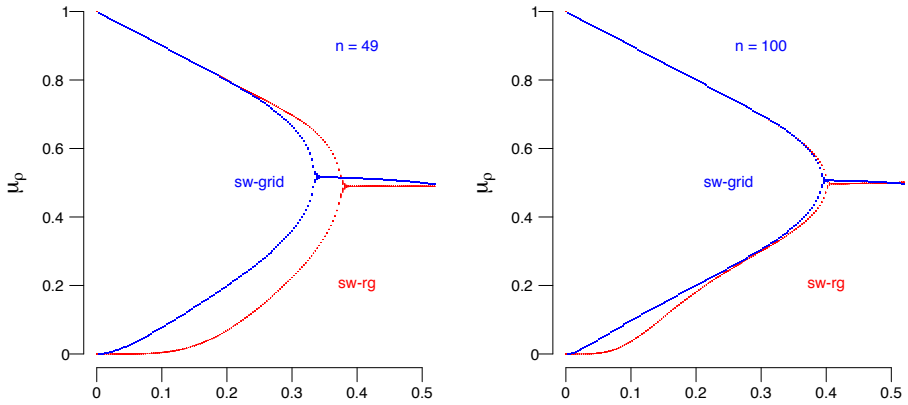


Figure 3.7: Bifurcation plots of the small-world mean field $\mu_{\text{sw}}^{v} = p_{\text{sw}}^{v}$ based on the approximation of the random part (in blue) and the mean field $\mu_{\text{sw}} = p_{\text{sw}}$ in (3.10) based on the random graph (in red). In the left panel a small-world of $n = 49$ nodes and in the right panel a graph of $n = 100$ nodes; all graphs are obtained with the probability of wiring (adding edges) in the NW small-world of $p_w = 0.4$.

3.5 Numerical evaluation of the mean field

To evaluate the accuracy of the predictions of the mean field in the grid, random, and small-world graph, we simulated networks of different sizes in the topology of a grid, a random graph, and an NW small-world graph. For each combination of parameters, 100 graphs were simulated. In combination with the majority rule the SCA for the grid, random graph or small-world was run for a certain duration T and the states of the last section of the time series were determined to see if it matches that of the predictions of the mean field. At time point $t = 0$, data in $\{0,1\}$ were generated for each one of the three types of graph, according to the Ising model using the R-package *IsingSampler* (Epskamp, 2015). Subsequent values $t > 0$ were obtained for all nodes by the majority rule given the value p for each type of graph. To determine the accuracy we used both 90% and 95% confidence intervals obtained from the central limit theorem (see Section

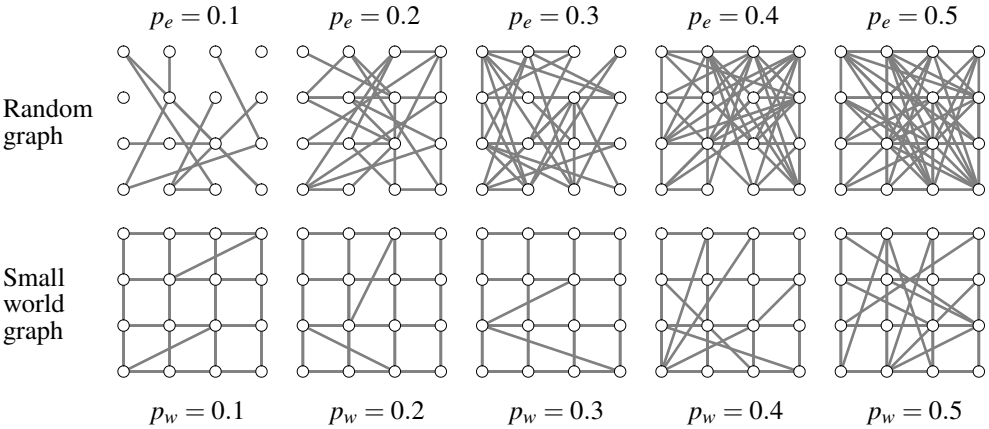


Figure 3.8: Examples of simulated random graphs (top row) and small-world graphs (bottom row) with network size $n = 16$. The probabilities of an edge p_e in the random graph and the probability of new edges in the small-world range between 0.1 and 0.5.

3.3.1) for each of the three different graphs. Figure 3.8 shows the topology of some simulated random graphs and small-world graphs arranged in a finite grid.

We varied the size of the network $n \in \{16, 25, 49, 100\}$, the number of time points $T \in \{50, 100, 200, 500, 5000\}$, and the probability of an active node in the majority rule $p \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$, see (3.1). We also varied the probability of an edge in the random graph $p_e \in \{0.1, 0.2, \ldots, 0.9\}$, and the probability of wiring in the small-world graph $p_w \in \{0.1, 0.2, \ldots, 0.9\}$. Figure 3.9 visualises the evolution of selected simulation conditions. All simulated data, figures, as well as the used R-code are publicly available at the Open Science Framework (Kossakowski & Waldorp, 2020).

Results in bifurcation diagrams with 90% and 95% confidence intervals are shown in Figure 3.10. The black lines are the bifurcation predictions from the mean field, the grey area above and below the mean field is the 90% confidence interval, and the grey dotted lines indicate the 95% confidence intervals. The red dots correspond to the last 100 points in the evolution of a particular kind of graph. It can be seen that for each of the different types of graph the densities of the simulated networks (red dots) are mostly within the 95% confidence interval. Except around the critical point (around 0.3) there

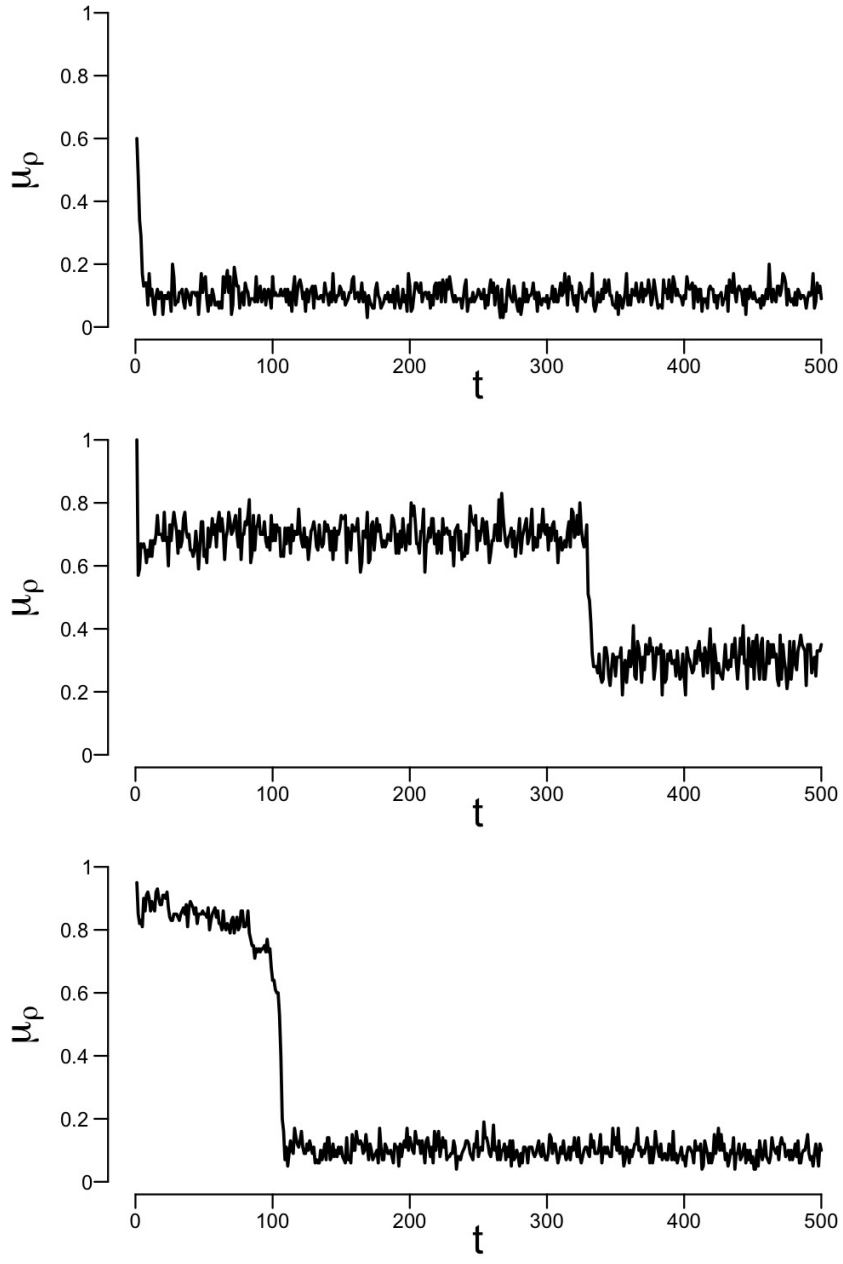


Figure 3.9: Examples of the evolution of a torus (upper panel; $p = 0.1$), a random graph (middle panel; $p = 0.3, p_e = 0.6$) and a small world graph (lower panel; $p = 0.1, p_w = 0.9$).

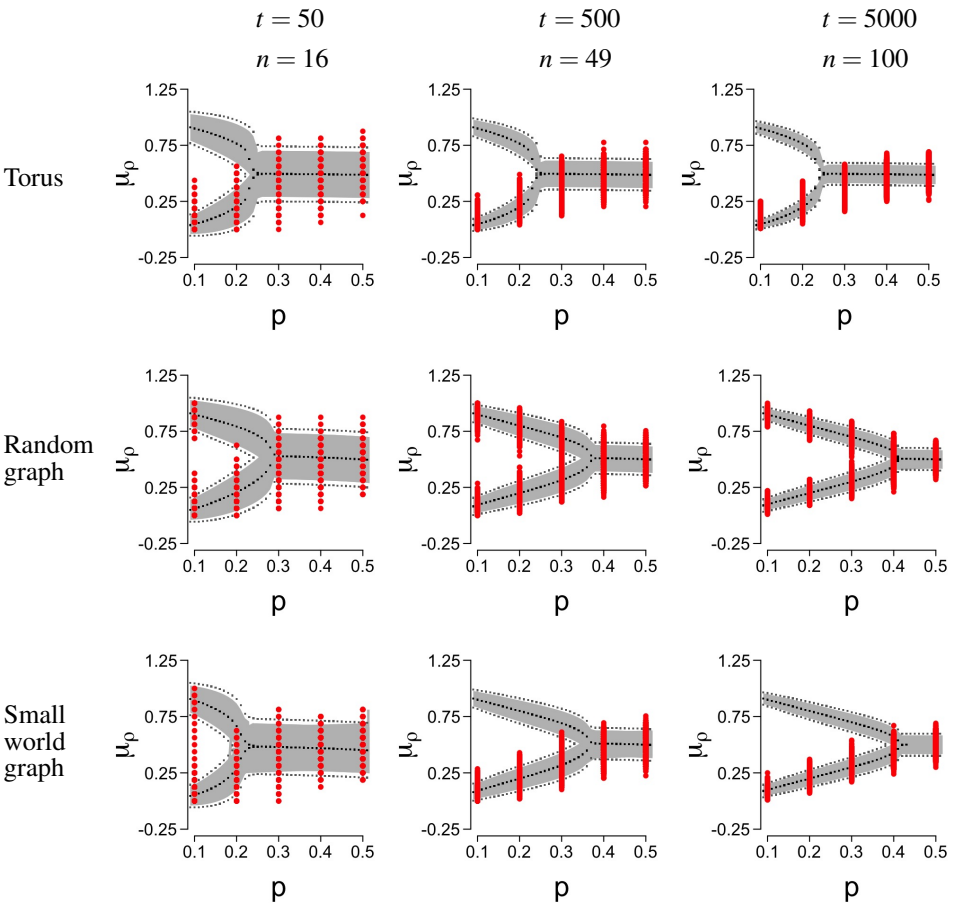


Figure 3.10: Bifurcation diagrams of a torus (upper panel), a random graph (middle panel; $p_e = 0.5$) and a small world graph (lower panel; $p_w = 0.6$). Grey solid area = 90% confidence interval around bifurcation. Dashed grey lines = 95% confidence interval around bifurcation. Red dots = last 100 points in the evolution at different values of p.

		$t = 50$					$t = 500$					$t = 5000$				
		$n = 16$					$n = 49$					$n = 100$				
	p	0.1	0.2	0.3	0.4	0.5	0.1	0.2	0.3	0.4	0.5	0.1	0.2	0.3	0.4	0.5
Torus	90%CI	0.75	0.99	0.92	0.94	0.96	0.29	0.97	0.63	0.75	0.81	0.12	1.00	0.50	0.75	0.75
	95%CI	0.80	0.99	0.99	0.99	0.99	0.67	0.99	0.75	0.83	0.87	0.55	1.00	0.50	0.83	0.75
Random graph	90%CI	0.97	1.00	0.55	0.78	0.85	1.00	1.00	0.98	0.92	0.94	1.00	1.00	0.87	0.98	0.99
	95%CI	0.97	1.00	0.76	0.88	0.92	1.00	1.00	0.98	0.97	0.98	1.00	1.00	0.91	1.00	1.00
Small-world graph	90%CI	0.34	0.20	0.38	0.54	0.63	1.00	1.00	1.00	0.98	0.99	1.00	1.00	0.99	1.00	1.00
	95%CI	0.76	0.43	0.59	0.69	0.75	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00	1.00

Table 3.1: Proportion of the densities of active nodes ending up in the confidence interval (90% or 95%). The setup is similar to Figure 3.10, where each row corresponds to a type of network and the columns correspond to the size of the network and length of the time series. In each cell the probability p of a 1 is varied, as in Figure 3.10.

is some difference between the density of the network and the mean field approximation, especially for smaller time series. Table 3.1, showing the proportions of the densities corresponding to Figure 3.10, confirms the lower proportions in the confidence intervals around the critical point in smaller time series. Note that we did not average the values obtained from the evolution of the graphs, and so the individual fluctuations are also represented by the red dots. All in all, results show that the mean field approximation also performs well when non-regular network structures are considered.

3.6 Conclusions and discussion

To model the complex dynamics of large-scale networks (graphs) is generally difficult. This is because there are many different 'agents' that operate within the graph. In particular, if the nodes in the graph represent symptoms and the edges represent their mutual influence, then the interacting symptoms show complex behaviour on a macroscopic scale, e.g., at the level of the number of active symptoms. Here we showed that the mean field model for a stochastic cellular automaton (SCA) with a majority rule, can serve as

an accurate approximation to such large-scale graphs, and can simplify the analysis of the dynamics of such systems. Specifically, we showed that averaging across the different possible node degrees for a random and small-world graph, results in approximations that lie with high probability close to the mean field. These theoretical results were confirmed by extensive simulations, showing that for smaller graphs the mean field lies within the 95% confidence interval.

Our approximation is based on the formulation of the grid (torus) where a relatively simple sum over possible active nodes determines the probability of a randomly selected node in the graph being active. We showed that for large graphs this approximation is accurate. This simplification could serve to obtain a more extensive analysis of the dynamics such as that presented in Janson et al. (2019). There the majority rule (the stochastic element) was removed from the model, to obtain exact fixed points for the model. Here we chose not to remove the stochastic element since we aim to introduce different rules for updates than the majority rule, like a conditional Ising probability.

Our initial motivation for these results was to obtain a model where we could assess the possibility of a person changing mood suddenly. Based on the estimate of the graph and the corresponding probability of an active node p and its corresponding bifurcation diagram, we would then be able to determine the possibility of that person 'jumping' from one state into another. This assessment might be useful in a clinical setting where a decision in a particular type of intervention is required. This idea is pursued chapter 4.

4

Applying a Dynamical Systems Model and Network Theory to Major Depressive Disorder

This chapter is published as: Kossakowski, J. J., Gordijn, M. C. M., Riese, H., & Waldorp, L. J. (2019). Applying a dynamical systems model and network theory to major depressive disorder. *Frontiers in Psychology: Quantitative Psychology and Measurement, 10*, 1-18. DOI: 10.3389/fpsyg.2019.01762.

Abstract

Mental disorders like major depressive disorder can be modelled as complex dynamical systems. In this study we investigate the dynamic behaviour of individuals to see whether or not we can expect a transition to another mood state. We introduce a mean field model to a binomial process, where we reduce a dynamic multidimensional system (stochastic cellular automaton) to a one-dimensional system to analyse the dynamics. Using maximum likelihood estimation, we can estimate the parameter of interest which, in combination with a bifurcation diagram, reflects the expectancy that someone has to transition to another mood state. After numerically illustrating the proposed method with simulated data, we apply this method to two empirical examples, where we show its use in a clinical sample consisting of patients diagnosed with major depressive disorder, and a general population sample. Results showed that the majority of the clinical sample was categorised as having an expectancy for a transition, while the majority of the general population sample did not have this expectancy. We conclude that the mean field model has great potential in assessing the expectancy for a transition between mood states. With some extensions it could, in the future, aid clinical therapists in the treatment of depressed patients.

4.1 Introduction

Major depressive disorder (MDD) is unfortunately not that uncommon: around 350 million people around the globe suffer from MDD (World Health Organization, 2012). While many studies have been conducted in the treatment of MDD, it remains unclear why certain people develop MDD and others do not; we do not know the exact circumstances of the person and its environment that may lead to MDD. There is some empirical evidence that people experience discrete mood states (Hosenfeld et al., 2015). This has led to the hypothesis that mood changes or (sudden) transitions to MDD may be related to dynamical systems theory (van de Leemput et al., 2014; Cramer et al., 2016; Wichers et al., 2016). In this chapter, we build on these ideas to assess the expectancy that a person has to develop MDD and embed such assessments more thoroughly in dynamical systems

theory and network theory in order to obtain a reasonable explanation of transitions to MDD.

Recently, the idea has been put forward that mental disorders, like MDD, can be considered as a system of interacting variables (Borsboom et al., 2011; Cramer et al., 2016; Guloksuz et al., 2017; Kossakowski & Cramer, 2019). Aspects of MDD, like loss of energy or feelings of worthlessness, can be seen as nodes in a network that interact with, and influence each other at later times and other symptoms of MDD (Cramer et al., 2012). This system of interacting emotions may change over time, making the system dynamic (Gulyás et al., 2013). Connections between various aspects of MDD can increase or decrease in strength over time, or aspects themselves may increase or decrease in strength as an individual develops MDD. We can measure these changes by means of the Experience Sampling Method (ESM; Csikszentmihalyi & Larson, 1987), where individual daily life experiences are measured several times a day for an extended period of time. At some point in time, when the system has surpassed some critical point (Scheffer et al., 2014), a discontinuous transition is made from a stable and healthy mood state to a stable and depressed mood state. Several studies have illustrated the bimodality of MDD, see for example van de Leemput et al. (2014); Wichers et al. (2016); Cramer et al. (2016) and Kossakowski & Cramer (2019). These sudden jumps, called *transitions* (Kuznetsov, 2013), are central to complex dynamical systems, and are the subject of the assessment that we will undertake in the present chapter. Please note here that we do not make any inferences about an individual's mental status before or after a transition has taken place. In this study we are mainly interested in the assessment of an individual's expectancy to transition between two mood states.

Attempts to anticipate a transition are often approached by so-called *early-warning signals* obtained from ESM studies (Kossakowski & Cramer, 2019). Dynamical systems leave 'breadcrumbs' behind in these time series that hint towards such a transition. These breadcrumbs occur before the transition, and after *critical slowing down* that may occur when the system finds it more difficult to return to the original equilibrium state (Scheffer

et al., 2014). Recently, it has been empirically shown that critical slowing down actually occurs prior to the transition (van de Leemput et al., 2014; Wichers et al., 2016). While critical slowing down is an important line of research, it is difficult to analyse critical slowing down in a system that has more than a handful of variables.

Hosenfeld et al. (2015) introduced a statistical measure to determine whether there are one or two stable mood states, based just on the distribution of the number of active symptoms per measurement. This statistical measure, called the *bimodality coefficient* (BC), only considers this distribution and determines whether there is evidence for one or two stable states. However, this approach offers no explanation of any kind of the phenomena observed in the distribution.

In this study we take a different approach and try to assess the expectancy of a transition between mood states. We investigate this expectancy by combining dynamical systems theory with network theory. More specifically, we use cellular automata as the framework for networks (cellular automata) and their stochastic counterparts to investigate dynamic behaviour. There are three reasons why we believe that the dynamics of a stochastic cellular automaton may be appropriate for psychopathology. First, there is some evidence that mood states are discrete, or at least they are experienced as such (i.e., see Hosenfeld et al., 2015), and mood can switch between these states. A cellular automaton such as the one we propose is able to have multiple stable states that are discrete, and the process can 'jump' between these states. The fact that the process can switch between states is important because we want to know the conditions under which such sudden changes can occur. Second, in line with network theory, we think that mood states and symptoms interact with each other and hence will influence each other (see Borsboom, 2017). A cellular automaton is a direct implementation of these ideas: it is a network and by definition each node affects its neighbours through an update rule, which can be specified based on the application. Third, because we always have uncertainty as to the correct specification of the variables in the network, we allow the updating process to be stochastic, accounting for unknown exogenous effects.

We will simplify the automaton by reducing the network to a single dynamic equation (given certain assumptions), and by characterising the possible states of this reduced system. We then have a process that may be an accurate description of what is going on with the changes in symptoms over time. We can, in turn, analyse these changes analytically and through simulations. We assume (intuitively) that the nodes in the network function roughly in the same manner and that each of the nodes affects the others in a similar way. The assumptions lead to a so-called *mean field model*. Using these assumptions, our focus becomes the proportion of active nodes in the system, which now forms a sequence of states ranging from 0 to 1. Since this sequence of states only depends on the proportion of active nodes at the previous time point, we obtain what is called a Markov chain and we can estimate the parameters by means of maximum likelihood estimation in a straight forward manner. Using this dynamical system allows us to determine whether it is possible for an individual to experience a transition or not.

As an example, we consider a time series of the proportion of active emotions for a single subject, shown in Figure 4.1 (left panel). We identify the possible states of this person with respect to the network of emotions, depending on the parameter of the process we assume underlies these observations. For this process we can obtain a so-called *bifurcation diagram* (Figure 4.1, right panel). This bifurcation diagram shows the possible (likely) states for this person given a value on the probability p of emotions changing from inactive to active. We assess from the time series of this person the parameters of our model and obtain an estimate of where in the bifurcation diagram this person is (represented by the vertical red line in the right panel of Figure 4.1). If the probability p is in the range of $[0.34, 0.50]$, where there is one point per value of p on the x-axis, then this person will remain stable. If the probability is lower than approximately 0.34, where there are two values for each value for p on the x-axis, then there are two stable states, one with a high proportion of active emotions and one with a low proportion of active emotions. The estimate of the probability p for this person is 0.192 (the vertical red line in the right panel of Figure 4.1). Based on this, we would classify this individual

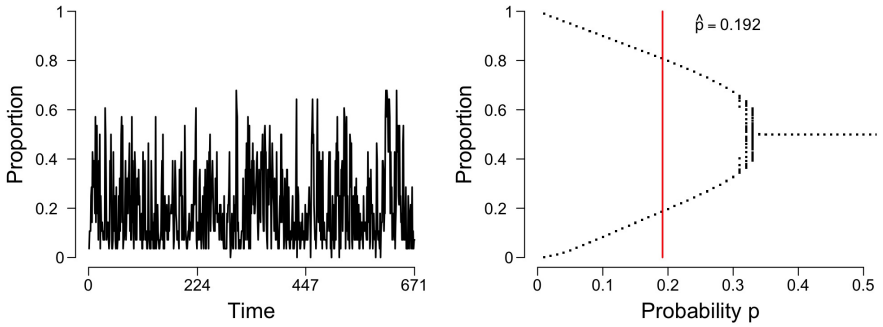


Figure 4.1: The evolution of the percentage of active nodes for each time point (left panel), and the accompanying bifurcation diagram (right panel). The red line in the bifurcation diagram in the right panel indicates the estimation of p. The x-axis in the left panel denotes the measurement occasions over time. The x-axis in the right panel denotes the probability p for a node to become active. The y-axis denotes the proportion of nodes in the system that are active.

as someone who may expect a (sudden) increase in the proportion of active emotions and thereby experience an episode of depression. And indeed, for this individual we know (from external evidence) that a depressive episode had taken place after the time series that we used to determine the state of the person (see Wichers et al., 2016; Kossakowski et al., 2017).

In the present chapter we obtain the maximum likelihood (ML) estimate for the model and the standard errors. We show, using simulations, that for many of the values of the parameter the estimate is reasonably close to the true value. Furthermore, we apply the proposed method to two real datasets, one with patients diagnosed with MDD, and one with subjects from the general population. This chapter is set up as follows. First, we will briefly explain the theory (Section 4.2) of the mean field model and the proposed method (Section 4.3 and 4.4). Then we present the simulation to show how the ML estimation performs in Section 4.5. Finally, in Section 4.6 we apply our method to two datasets to show how the method works in different contexts.

4.2 Stochastic cellular automata

To model interacting symptoms and emotions we use a particular kind of structure, a *stochastic cellular automaton* (SCA). Such automata are particular dynamical systems that show typical behaviour for stable and bistable behaviour depending on the settings (Kozma et al., 2004; Balister et al., 2006), which is what we assume to the case for MDD. For the interested reader, the books by Holmgren (1996); Hirsch et al. (2004); Hasselblatt & Katok (2003) and Golubitsky & Stewart (2003) provide background information on dynamical systems theory. A cellular automaton (CA) is a dynamical system where nodes are arranged in a fixed and finite grid, and where connected nodes determine the state of a node at each subsequent time point (Wolfram, 1984b; Sarkar, 2000). A node j that is directly connected to node i is called a neighbour. A grid is a graph $G_{\text{grid}}(n,\Gamma)$ with n nodes in the set $V = \{1,2,\ldots,n\}$ where each node i has the same number of neighbours in its neighbourhood $\Gamma = \{j \in V : j \text{ is connected to } i\} \cup \{i\}$ including itself. To ensure that all nodes have exactly the same number of neighbours, we impose the boundary condition such that a node at the boundary is connected to a node on the opposite end, making it a *torus*. An example of such a grid is shown in Figure 4.2 (left), where the middle node is directly connected to its four neighbours, marked in grey. We consider elementary CAs where each node can be in either of two states: 'active' (coded by 1) or 'inactive' (coded by 0). In a CA a deterministic, local update rule ϕ determines the state $x_{i,t}$ of each node $i \in V$ at the next time step based on which nodes are active in the neighbourhood of node $x_{i,t}$. An example of such an update rule is the *majority rule*, where each node becomes 1 (active) whenever more than half of the neighbours of node i at the previous time point are active, and 0 (inactive) otherwise. Although many other update rules are possible, we will focus on this particular rule in the present study. One of the reasons for choosing the majority rule is that it is stable. In other words, small changes in the number of active nodes will not affect the decision (O'Donnell, 2014). Repeated application of the update rule ϕ results in a vector of 0s and 1s, called an orbit: At any time point t the orbit $\phi^t(x_i) = \phi \circ \phi \circ \phi \cdots \circ \phi(x_{i,0})$ (initial value at $t = 0$), such that the same local rule is

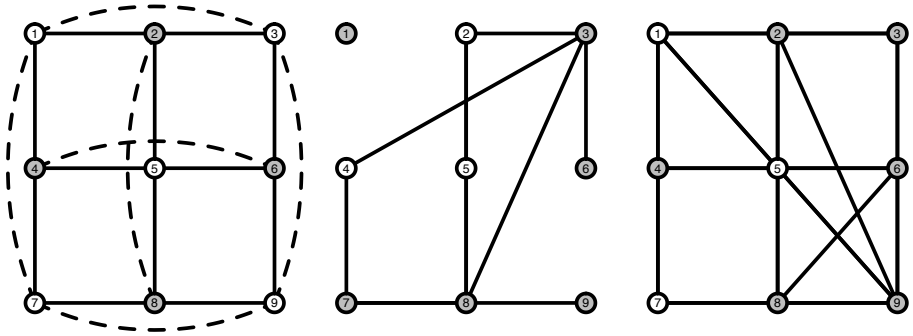


Figure 4.2: Visualisation of a grid structure (left panel), a random graph structure (middle panel), and a small-world structure (right panel). Grey nodes indicate the neighbours of the middle node in each graph. Solid lines indicate pairwise connections between nodes. Dashed lines are also pairwise connections, but have been curved and dashed as they are hidden behind other connections in a 2D-view.

applied to the result of the previous time point t times.

To illustrate, say that we have the network presented in Figure 4.2 (left), and we have the following orbit of active and inactive nodes $\phi^0(x_{1,0}) = 1$, and for the other 8 nodes 1, 0, 1, 0, 0, 1, 1, 0, as shown in Table 4.1. We can then determine how many active neighbours r each node has, by just counting the number of active nodes each node is connected to. As mentioned in Table 4.1, nodes 1, 5 and 7 have three active neighbours, nodes 2, 3, 4, 8 and 9 have two active neighbours, and node 6 has one active neighbour. For this example we will use the majority rule ϕ that is described earlier, which states that a node is activated ('1') when more than half of that node's neighbourhood is active. The majority rule uses $r > |\Gamma|/2$ to indicate whether the number of active neighbours is greater than half the size of the neighbourhood, where Γ denotes the size of a node's neighbourhood. In our example, $|\Gamma| = 5$: each node has exactly four neighbours, and the node itself at $t-1$ is the fifth addition to $|\Gamma|$. With the majority rule ϕ, the next time step becomes $\phi^1(x_{i,1}) = (1,0,0,0,1,0,1,0,0)$. We then use this sequence of active and inactive nodes to determine the number of active nodes r at $t = 1$, which is described in

Node	t_0	r_0	t_1	r_1	t_2
1	1	3	1	1	0
2	1	2	0	2	0
3	0	2	0	1	0
4	1	2	0	3	1
5	0	3	1	0	0
6	0	1	0	1	0
7	1	3	1	1	0
8	1	2	0	2	0
9	0	2	0	1	0

Table 4.1: Illustration of the majority rule as used for Figure 4.2. The columns t_0, t_1 and t_2 denote the sequence of active nodes at a specific time point. The columns r_0, r_1 and r_2 denote the number of active neighbours per node at time point t.

Table 4.1, column r_1. We can continue this process for a length T (not shown in Table 4.1), thus creating a $T \times n$ matrix that holds the orbit $\phi^t(x_{i,t})$ on the columns.

In the illustration above, the majority rule used to update the system was a deterministic one. In a stochastic cellular automaton (SCA), a probability is introduced to model uncertainty, based on the number of active neighbours (r). In our application to psychopathology, this uncertainty is required because we cannot predict the behaviour of emotions in our network exactly, and because we know that exogenous events influence these emotions that we cannot measure. By just counting the number of active neighbours that a node has, we can determine the probability for a node to become active. The probability $0 \leq p \leq 1$ determines whether or not a node becomes active at time point $t+1$. The majority rule combined with this probability equals the probability that we obtain for node $x_{i,t+1} = 1$, given that there are r active neighbours is

$$\mathbb{P}(X_{i,t+1} = 1 \mid r) = \begin{cases} p & \text{if } r \leq |\Gamma|/2 \\ 1-p & \text{if } r > |\Gamma|/2 \end{cases} \tag{4.1}$$

where $|\Gamma|$ is the size of the neighbourhood and r the number of active neighbours. The parameter p is determined a priori or is estimated from data (see below). Because $\mathbb{P}(x_{i,t+1} \mid r)$ depends on the behaviour of the majority of a node's neighbourhood, this update rule is also called the *majority rule*. In this SCA each node $i \in V$ is then updated according to the majority rule; all nodes are updated simultaneously (synchronous updating). The result for each node is a sequence (orbit) of 0s and 1s. From all $n = |V|$ nodes we can determine the total number of active nodes Y_t at time point t, for all time points up to time T. We are interested in the number of active nodes $Y_t = \sum_{i=1}^{n} X_{i,t}$ (where $X_{i,t}$ is the value of node x_i at time point t) and so we average over all nodes in the grid at each time point t, obtaining $\rho_t = Y_t/n$, which is often referred to in the literature as the *density*. An example of the density (proportion) is shown in Figure 4.1 (left panel).

4.3 Mean field model

It is rather difficult to infer the characteristics of what the system will do in the long run from an SCA (Lebowitz et al., 1990). We need to simplify the SCA in order to make it possible to derive the characteristics of the SCA. Here we use an approximation for the structure of the network, where we assume the average of the number of neighbours for each node $|\Gamma|$. We also assume that nodes can be in either of two states: active ('1') or inactive ('0'), and that the nodes behave in a similar manner. The latter assumption means that the majority rule, as presented in equation (4.1), is applied to all nodes in the network, and that all nodes become active or inactive in the same way. In the grid (Figure 4.2, left panel) it is easily seen that each node is similar to any other node since each node has the same number of neighbours, and becomes active or inactive in the same way by means of the majority rule that is based on the number of active neighbours. This allows us to

simplify an SCA to a single discrete time dynamical system, as in Kozma et al. (2005), Balister et al. (2006), and Waldorp & Kossakowski (2020).

In the *mean field* model we make use of the uniformity of the nodes in a grid. The change of a node from 0 to 1 or vice versa is based on the number of active neighbours in that node's neighbourhood (r) and the probability parameter p, following the majority rule defined in equation (4.1). In a grid each node has exactly the same number of neighbours and so the probability of a node changing value depends on the properties of the grid and not on the local activity. Therefore, as shown in Kozma et al. (2005) and Balister et al. (2006), we obtain at time point $t+1$ the number of active nodes in the grid Y_{t+1}, which is a random variable with a binomial distribution that has a success probability $\rho_t = Y_t/n$, the proportion of active nodes (density) at the previous time point t. The number of draws in the binomial probability is determined by the size of the neighbourhood $|\Gamma|$ particular to the graph. The majority rule in equation (4.1) determines for which number of active nodes we obtain p up until the number of active nodes $r \leq \lfloor |\Gamma|/2 \rfloor$, where $\lfloor a \rfloor$ is the integer part of a, or $1-p$ otherwise. To define the probability of the number of neighbours that are 1, we need to consider all possible configurations of $|\Gamma|$ in terms of active-inactive nodes in the graph. There are $\binom{|\Gamma|}{r}$ ways to choose r active nodes out of $|\Gamma|$ each with a success probability $\rho_t = Y_t/n$. We then obtain for the probability of r active nodes out of $|\Gamma|$

$$\mathbb{P}(r \mid \rho_t) = \binom{|\Gamma|}{r} \rho_t^r (1-\rho_t)^{|\Gamma|-r}. \tag{4.2}$$

Simultaneously, we require the probability p or $1-p$ from the majority rule in equation (4.1), which is assumed to be independent. We need to define the probability for any number of active nodes and therefore marginalise over the number of possible active nodes in the neighbourhood r. Putting equation (4.1) and equation (4.2) together, we obtain the joint probability $\mathbb{P}(X_{i,t+1} = 1 \mid r, \rho_t) = \mathbb{P}(X_{i,t+1} = 1 \mid r)\mathbb{P}(X_{i,t+1} \mid \rho_t)$. Hence we obtain the probability for any node in the graph to be 1 as

$$\rho_{t+1}^{\text{grid}} = \sum_{r=0}^{|\Gamma|} \mathbb{P}(X_{i,t+1} = 1 \mid r) \binom{|\Gamma|}{r} \rho_t^r (1-\rho_t)^{|\Gamma|-r}. \tag{4.3}$$

Because the evolution is binomial based on the proportion of active nodes at the previous time point (see equation (4.3)), it follows from the transition probability that the number of active nodes $X_{t+1} = x_{t+1}$, given that at t $X_t = x_t$ in the graph G_{grid}, is

$$\mathbb{P}(X_{t+1} = x_{t+1} \mid X_t = x_t) = \binom{n}{x_{t+1}} \rho_t^{\text{grid}}(x_t/n)^{x_{t+1}} (1 - \rho_t^{\text{grid}}(x_t/n))^{n-x_{t+1}}. \tag{4.4}$$

So, we know how in a grid with n nodes the proportion of active nodes ρ_t changes from time point t to time point $t+1$, for any t. The mean field model uses the mean of this binomial process and divides by n to obtain the proportion. We often denote the expected value of Y_{t+1}/n by $\mu_{\text{grid}} := \rho_t^{\text{grid}}$ to emphasise that we use the mean of the process in a grid. We know that the fluctuations around the mean are small (depending on the standard deviation and size of the grid, see Waldorp & Kossakowski, 2020), so the mean is a good approximation of the process itself.

As an illustration of the binomial process, in the left panel of Figure 4.3 we see a typical SCA process where it is clear that the fluctuations are around a particular mean (0.3) for time points before $t = 125$ approximately. After this point (tipping point) the fluctuations revolve around another mean (0.7) with a higher proportion of emotions. In the right panel of Figure 4.3 we see a plot of the expectation of the process, which is the mean field that predicts the values at which the mean of the process converges to eventually. It is this mean field function that we will use to represent the process and the network that evolves over time.

We now regard the mean field, the expectation of the binomial process $E(Y_t/n) = \mu_{\text{grid}}$, as the dynamical system that is a representation of the network. This dynamical system evolves by repeated application of μ_{grid} to its previous result. We analyse the dynamical properties of μ_{grid} by considering a so-called bifurcation diagram. Plugging in different values for the a priori set parameter p from equation (4.3) in the majority rule, in the interval $(0, 0.5]$, we obtain a *bifurcation diagram*, as shown in Figure 4.1 (right panel). In a bifurcation diagram the repeated application of μ_{grid} is applied to updated values of ρ_t^{grid} such that the last section of the orbit is displayed where the process is

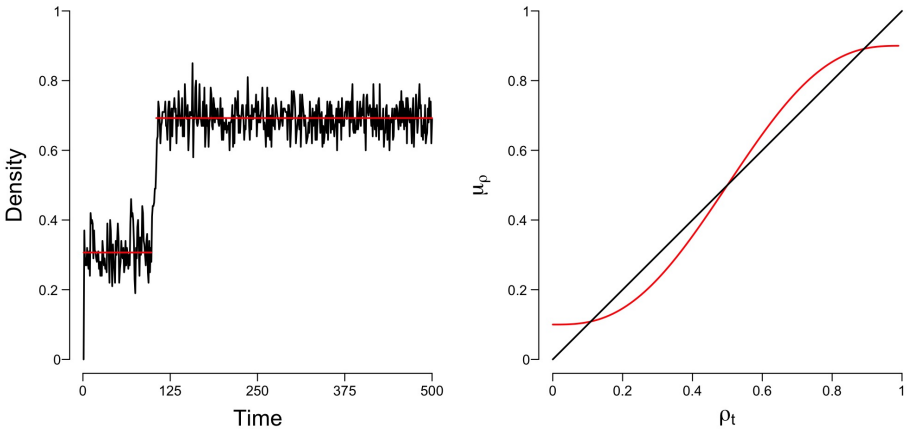


Figure 4.3: Example of a stochastic cellular automata process that includes a transition (left panel). Example of the corresponding mean field function (right panel). The red line indicates the expectation of equation (4.3) divided by the number of nodes.

in equilibrium (stable if stable fixed points exist; Hirsch et al., 2004). For each value of p, displayed on the x-axis in Figure 4.1, one sequence is generated, of which the last 50 are displayed in Figure 4.1. In some cases, the sequence will find two equilibria, and thus we draw two points at those two equilibria. In other cases, the sequence will converge to one equilibrium, and thus only one point will be drawn in the bifurcation diagram. Such diagrams show what kind of behaviour can be expected to be generated by the process. Here we see that there are two kinds of situations: (a) a stable situation when p is in the interval $(0.34, 0.50]$, where irrespective of the starting point, the process ends up at that stable fixed point, and (b) a bistable situation when p is in $[0, 0.34]$ where the process could (suddenly) switch between states (transition) to a low or high density. The parameter value p at which the process changes from a stable to a bistable situation is called the critical point. In Figure 4.1 the critical point lies at $p \approx 0.34$; the parameter area $0 < p \leq 0.34$ is bimodal where transitions can occur, whereas the parameter area $0.34 < p < 0.50$ represents a unimodal area where the mean field is stable. Thus, the parameter

p can be used to determine whether a process has two stable states, and therefore can transition between them, or one stable state, where no transition can occur.

The probability for the mean field in equation (4.3) is designed for a grid with a fixed neighbourhood size $|\Gamma|$. In the context of psychology and psychopathology, it is hard to come up with a graph representing the interactions between variables, that would take the form of a grid. We therefore also looked at the mean field model for a *random graph* and a *small-world graph*. A random graph $G_{\text{rg}}(n, p(e))$ is a graph structure with n nodes and a (constant) probability $p(e)$ for an edge to be present in the graph (Bollobás, 2001; Durett, 2007). In the mean field model of a random graph, the neighbourhood size $|\Gamma|$ is a random variable that is maximally $n-1$. Each node has a binomial number of neighbours with expected number of neighbours $p(e)(n-1)$. We extend the idea used for the grid, where we marginalise over all possible configurations of the number of active nodes for each neighbourhood of size $n-1$ for the random graph. One can approximate this probability accurately with a small modification of the probability used for a grid (Waldorp & Kossakowski, 2020). The difference with the probability on the grid is in the size of the neighbourhood (see Figure 4.2, left and middle panel), where in the grid the neighbourhood size is fixed to $|\Gamma|$. In the mean field model for the random graph, we fix this to the expected number of neighbours $p(e)(n-1)$. Let $\nu = \lfloor p(e)(n-1) \rfloor$ be the integer part of the expected number of neighbours. For the random graph G_{rg} the neighbourhood size is no longer $|\Gamma|$ (like it is for the grid), but ν. The probability in a random graph for a node to become active given the graph's density at time point t (ρ_t) and the edge probability then becomes (Waldorp & Kossakowski, 2020):

$$\rho_{t+1}^{\text{rg}} = \sum_{r=0}^{\nu} \mathbb{P}(X_{i,t+1} = 1 \mid r) \binom{\nu}{r} \rho_t^r (1-\rho_t)^{\nu-r}. \tag{4.5}$$

A small-world graph is in between a grid and a random graph where, compared to a random graph, the average clustering is high and the average path length is low (Watts & Strogatz, 1998). A modified version of the small world is the Newman-Watts (NW) small-world (Newman & Watts, 1999a). In the NW small-world $G_{\text{sw}}(n, \Gamma, p(w))$ the n nodes

each have a neighbourhood Γ as in the grid and edges are added to the network following a (constant) wiring probability $p(w)$ (Newman & Watts, 1999a). We can then split up the probability in a part associated with the grid and a part associated with the random graph. The part for the grid is adjusted such that it corresponds to no other edges being present, i.e., we obtain $\rho_t^{\text{grid}}(1-p(w))^{n-|\Gamma|}$, where the product $(1-p(w))^{n-|\Gamma|}$ represents the probability that no other edges are present for $n-|\Gamma|$ nodes. For the random part we obtain the probability as in equation (4.5) but the first $|\Gamma|$ edges left out, because they have already been accounted for by the grid part. We denote this probability by $\rho_t^{\text{rg},\Gamma}$, which denotes the probability as in equation (4.5) but with $\rho_t^{\text{rg},\Gamma}$ starting at $|\Gamma|+1$ instead of 0. Then the probability for a node to become active given the graph's density at time point t (ρ_t) and the wiring probability $p(w)$ in the small-world graph G_{sw} is

$$\rho_{t+1}^{\text{sw}} = \frac{|\Gamma|}{n}\rho_t^{\text{grid}}(1-p(w))^{n-|\Gamma|} + \frac{n-|\Gamma|}{n}\rho_t^{\text{rg},\Gamma}. \tag{4.6}$$

The small-world probability is therefore a combination of the probability from the grid and from a random graph, proportionately weighted.

4.4 Estimation of probability p and graph parameters

Our objective is to derive an estimate of the probability parameter p from a time series to determine whether an individual can expect a transition between two mood states. One way of obtaining such an assessment is to determine where in a bifurcation diagram a person is located with respect to the parameter p in the majority rule; is this in the stable area, where no transition can occur, or is it in the bistable area where a transition can occur. In order to do this, we need to estimate the parameter p that is essential in the majority rule in equation (4.1). Here we use maximum likelihood (ML) to obtain an estimate of p (Rajarshi, 2012).

If we take a closer look at equation (4.3), it can be noticed that all parameters are known prior to the analysis, with the exception of the probability parameter p. To obtain p, we can estimate it from the data using ML estimation. We then obtain the maximum of

the log-likelihood for the probability parameter p that exists in equation (4.1). We write the transition probability in going from state x_t to state x_{t+1} (number of active nodes in the graph) in equation (4.4) from t to $t+1$ as $\mathbb{P}(X_{t+1} = x_{t+1} \mid X_t = x_t)$. The log of the joint probability function (loglikelihood) for the number of active nodes is then

$$\log \mathbb{P}(X_t, t \geq 0) = \sum_{t=0}^{T-1} \log \mathbb{P}(X_{t+1} = x_{t+1} \mid X_t = x_t), \tag{4.7}$$

where T denotes the total duration of the sequence in time points. The transition probability $\mathbb{P}$ is as in equation (4.4). The data that are plugged into this equation is a vector of length T that holds the number of active nodes for each time point t. At each time point the number of active nodes is given as input to the probability in the binomial process $\rho_t = (Y_t/n)$, where x_t is the number of observed active nodes at time point t. The data are plugged in the transition probabilities, where we recognise in the SCA that we can relatively easily find the transition probability to go from x_t to x_{t+1} active nodes. We can find these transition probabilities because of the fact that we have, for each of the graphs G_{grid}, G_{rg}, and G_{sw}, a binomial process with a probability of success particular to each type of graph. The parameter ρ_t for the random graph G_{rg} and the small-world graph G_{sw} are similar except that we change the probability of success to ρ_t^{rg} or ρ_t^{sw}, respectively.

The process is ergodic whenever the probability ρ_t^{grid} is not in the basin of attraction of 0 and 1 (see Waldorp & Kossakowski, 2020). In other words, a process is ergodic when the process is stationary and when all nodes in the graph follow the same dynamics (Molenaar, 2007). In those cases we could simplify equation (4.7) using only the transition probabilities that do not depend on time. In general, however, we do not know where the probabilities are, and therefore we do not assume ergodicity and cannot simplify the log-likelihood to terms consisting only of the states and not on time (Fleming & Harrington, 1978). We maximise the log-likelihood function in equation (4.7) with respect to p to obtain its estimate from an empirical time series, making it possible to place that person on the bifurcation diagram and assess the expectancy of possible switching.

In both the random and small-world graph we have additional graph parameters: in the random graph we have the probability of an edge $p(e)$, and in the small-world graph we

have the probability of re-wiring $p(w)$. Both parameters are obtained by maximising the log-likelihood with respect to $p(e)$ and $p(w)$ respectively. Equations (4.3), (4.5) and (4.6) each show how we calculate the density (ρ) in a grid, a random graph, or a small-world graph, respectively. One only needs to plug in a value for p (and the graph parameters $p(e)$ or $p(w)$ in the case of a random graph or a small-world graph) into the equation and let it run for some time T (often 1000 is enough), to find out at what density it will end up, or between which two values it may transition in the case of two stable states. By varying the value for p, one can create a bifurcation diagram, of which examples are shown in Figure 4.4. Each dot represents a separate run of the mean field equation. Equation (4.3) is reflected in the top panel of Figure 4.4, equation (4.5) in the middle panel, and equation (4.6) in the bottom panel of Figure 4.4.

Taking the top panel of Figure 4.4 as an example, if we run equation (4.3) with $p = 0.1$, it can be seen that the binomial process ends up in either 0.1 or 0.9 approximately, and could switch between these states. Our mean field model says that if we estimate the probability parameter p for an individual to be $\hat{p} \approx 0.1$, then this person could experience a transition between the two states, which could be related to a depressive episode. When we increase the value of the probability parameter $\hat{p} \approx 0.3$, the binomial process no longer has the possibility of a transition between states, but will remain around 0.5 approximately. The critical point, the point where the system changes from having two stable states to one stable state, differs depending on the size of the graph and the type of graph; for the random graph and the small-world graph the location of the critical point also depends on the graph parameters $p(e)$ or $p(w)$, as seen in Figure 4.4. To summarise, in order to categorise individuals, we need to determine the most appropriate the size of the graph, the most appropriate graph structure and its associated graph parameter to find the critical point in this personalised mean field model. Using this model, we can then determine an individual's position in terms of the probability parameter p.

Uncertainty can be quantified by the standard error of the estimate $\hat{p}$. For the grid we only have the estimate of p and for the random graph and the small-world graph

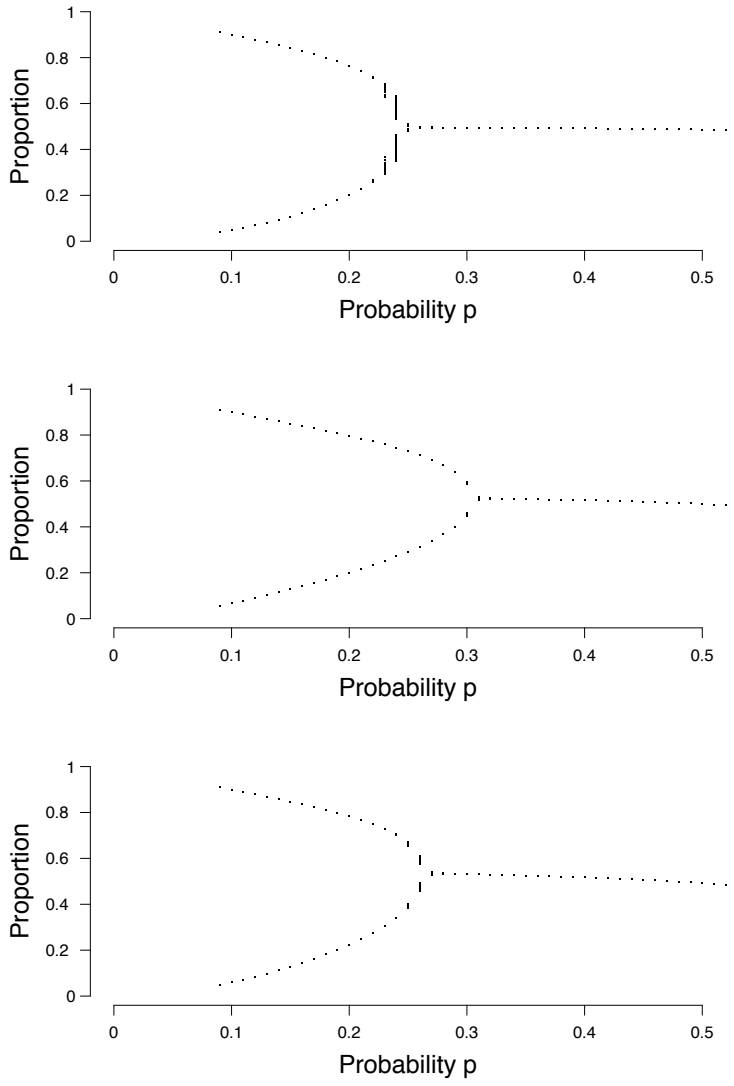


Figure 4.4: Examples of bifurcation diagrams for a grid (upper panel), a random graph (middle panel; $p(e) = 0.1$) and a small-world graph (lower panel; $p(w) = 0.1$). The x-axis denotes the probability p for a node to become active. The y-axis denotes the proportion of nodes in the system that are active.

we have the edge probability $p(e)$ and $p(w)$, respectively. Standard errors are obtained from the second order derivatives (Hessian) of the log-likelihood (Rajarshi, 2012). The inverse (matrix) of the Hessian and scaled by $1/T$ results in the variance of the parameter estimates. The square root of the diagonal elements are the standard errors, i.e., $\text{SE}(\hat{p}) = \sqrt{\frac{1}{T}h^{11}}$ is the standard error for $\hat{p}$, $\text{SE}(\hat{p}(e)) = \sqrt{\frac{1}{T}h^{22}}$ is the standard error for the edge probability in the random graph, and $\text{SE}(\hat{p}(w)) = \sqrt{\frac{1}{T}h^{22}}$ is the standard error for the rewiring probability in the small-world graph, where h^{ij} is the ijth element of the inverse Hessian.

4.5 Numerical illustration of probability p and graph parameters

Before we apply the mean field model to empirical data, we want to know how well the mean field model can estimate the probability parameter p in simulated data. We simulated 100 networks for each topology of a grid, a random graph, and a small-world graph. We varied the size of the network $V \in \{ 16, 25, 49, 100 \}$, the number of time points $T \in \{ 50, 100, 200, 500, 5000 \}$, and the probability $p \in \{ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 \}$ (Equation 4.3). We also varied the probability for an edge in the random graph $p(e) \in \{ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 \}$ (Equation 4.5) and the probability for an edge to be rewired in the small-world graph $p(w) \in \{ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 \}$ (Equation 4.6). For $t = 0$, a random number of nodes was set to active by using the R-package *IsingSampler* version 0.2 (Epskamp, 2015).

For each of the 100 simulation runs, we used the $T \times n$ set of active and inactive nodes to estimate the probability parameter p and the graph parameters $p(e)$ and $p(w)$. All simulated data, figures, and the used R-code are publicly available (OSF; Kossakowski, 2019). For clarity of presentation, figures are only presented for $T = 50$, as results for the other number of time points were nearly identical. We also only present the results for p, $p(e)$ and $p(w) \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ as the simulation results for these parameters $>$ 0.5 hardly occur in empirical data, and are therefore for this chapter less interesting. These and other results can be found online (Kossakowski, 2019). For each simulation run, we

calculated the absolute difference between the probability parameter p, under which the data were simulated, and $\hat{p}$, which we estimated from the data using ML estimation. We denoted this absolute difference by $\Delta(p)$, after which we take its mean ($\overline{\Delta}(p)$). This mean is determined for each replication, and can be interpreted as an error rate. The lower this value, the closer the estimate $\hat{p}$ is to the original value p. The same procedure was performed to determine the accuracy for graph parameters $p(e)$ ($\overline{\Delta}(p(e))$) and $p(w)$ ($\overline{\Delta}(p(w))$). A complete overview of all results across all conditions can be found in Table S1 the supplementary files online.

Figure 4.5 shows a visual representation of the mean absolute difference ($\overline{\Delta}(p)$) between the true probability parameter p, and the estimated probability parameter $\hat{p}$. It can be seen that the error rate $\overline{\Delta}(p)$ for p is low for all different network structures. Supplementary file S1 shows the mean estimate of p and its associated standard deviation for all conditions. The standard deviation for $\hat{p}$ is pretty low across conditions and never exceeds 0.04. The mean error rate $\overline{\Delta}(p)$ did not exceed 0.08 for the grid (for $T = 5000$, $n = 100$, $p = 0.4$), 0.06 for the random graph (for $T = 50$, $n = 25$, $p = 0.2$, $p(e) = 0.1$), and 0.04 for the small world graph (for $T = 50$, $n = 16$, $p = 0.5$, $p(w) = 0.4$). The error rate ranged between $[0.006 - 0.12]$ for the grid, $[0.0009 - 0.15]$ for the random graph, and $[0.008 - 0.16]$ for the small world graph. A small increase in the error rate can be noticed for the grid around the values $p = 0.3$ and $p = 0.4$. We think that a possible explanation is that the mean field model has some issues with estimating p around the critical value, the point where the system either has one stable state, or two stable states. Because of fluctuations in the process, the exact critical point is difficult to estimate.

The same conclusion cannot be drawn for graph parameters $p(e)$ and $p(w)$, as seen in Figure 4.6. For a random graph, $\overline{\Delta}(p(e))$ is high when $p(e)$ is low, and decreases as $p(e)$ is increased. This shows that the graph parameter $\hat{p}(e)$ is most accurate when $p(e)$ is high. A possible explanation for this finding could be found in the connectedness of random graphs. When $p(e)$ is small, the probability that not all nodes are connected increases, resulting in isolated nodes. When we look at the minimum probability $p(e)$,

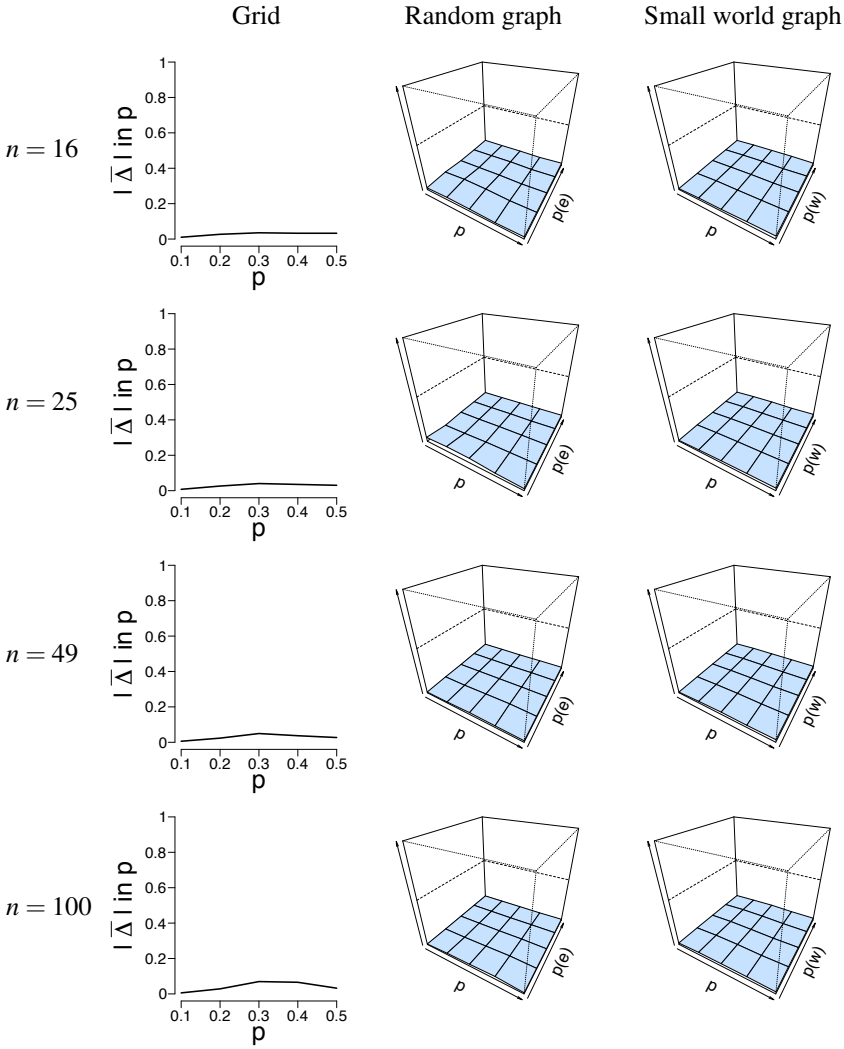


Figure 4.5: Visualisation of the mean error rate between p and $\hat{p}$. Mean absolute difference ($|\Delta|$) is shown for a grid (left column), a random graph (middle column), and a small-world graph (right column) at $T = 50$. For the left column, the x-axis denotes the parameter p for which we simulated data, and the y-axis the mean absolute difference $|\Delta|$ between p and $\hat{p}$. For the middle and right column, the x-axis denotes the parameter p for which we simulated data, the z-axis the graph parameter that was used to simulate data ($p(e)$ or $p(w)$), and the y-axis the mean absolute difference $|\Delta|$ between p and $\hat{p}$. The mean absolute difference ranges between 0 and 1, where a lower value indicates a smaller difference between p and $\hat{p}$.

such that the graph is connected for different network sizes, we see that $p(e)$ must be at least 0.46 when the network size is 16, 0.31 when the network size is 25, 0.17 when the network size is 49 and 0.09 when the network size is 100. Thus, as $p(e)$ increases, the probability for the network to be connected increases, and as a result of this, the error $\overline{\Delta}(p(e))$ decreases. The reverse is true for the small-world graph, where $\overline{\Delta}(p(w))$ is high when $p(w)$ is high and p is low, and decreases when $p(w)$ also decreases. This shows that the graph parameter $\hat{p}(w)$ is most accurate when $p(w)$ is low and when p is high.

To investigate the standard errors, we calculated the mean standard error (SE) and its associated standard deviation for all conditions using the Hessian matrix provided by the ML estimation procedure. Table S1 in the supplementary files depict the mean SE and its standard deviation for all conditions. It can be seen that the mean SE is extremely low across all conditions, indicating good accuracy of the estimates. We calculated the absolute difference between the standard deviation of $\hat{p}$ and the SE of $\hat{p}$. The difference ranged from 0.0003 to 0.18, and in 98.9% of all conditions, the difference between the standard deviation and the SE is smaller than 0.05.

Next to the SE, we calculated the error rate $\Delta(p)$ when the network structure is misspecified, and thus used the incorrect model to estimate $\hat{p}$ from the data. We used two datasets that represent the best and worst case scenario in terms of data structure. The worst case is the data with $n = 100$ nodes and $T = 50$ time points. The best case is the data with $n = 16$ nodes and $T = 5000$ time points. By taking the least and most ideal combination of n nodes and T time points, we obtain results where all other combinations will most likely lie in. With these properties in mind, we selected the data simulated for all three network structures, and applied all three models to estimate $\hat{p}$. Figure 4.7 depicts the error rate $\overline{\Delta}(p)$ for $n = 100$ nodes and $T = 50$ time points. We chose not to present the results for the data with $n = 16$ nodes and $T = 5000$ nodes for clarity of presentation. We estimated $\hat{p}$ for all three network structures for all datasets. It can be seen in Figure 4.7 that $\overline{\Delta}(p)$ is generally low, regardless of the network structure that was used to simulate the data.

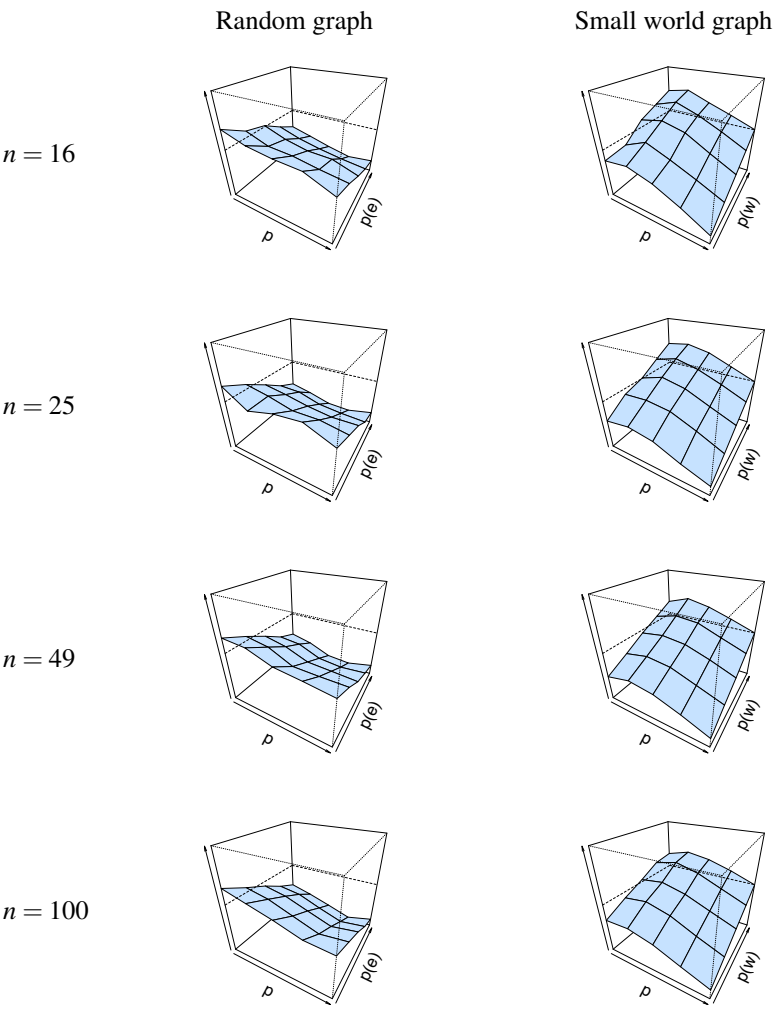


Figure 4.6: Visualisation of the mean absolute differences between $p(e)$ and $\hat{p}(e)$ and $p(w)$ and $\hat{p}(w)$ at $T = 50$. Mean absolute difference is shown for a random graph (left column) and a small-world graph (right column). The x-axis denotes the parameter p for which we simulated data, the z-axis the graph parameter that was used to simulate data ($p(e)$ or $p(w)$), and the y-axis the mean absolute difference $|\Delta|$ between $p(e)/p(w)$ and $\hat{p}(e)/\hat{p}(w)$. The mean absolute difference ranges between 0 and 1, where a lower value indicates a smaller difference between $p(e)/p(w)$ and $\hat{p}(e)/\hat{p}(w)$.

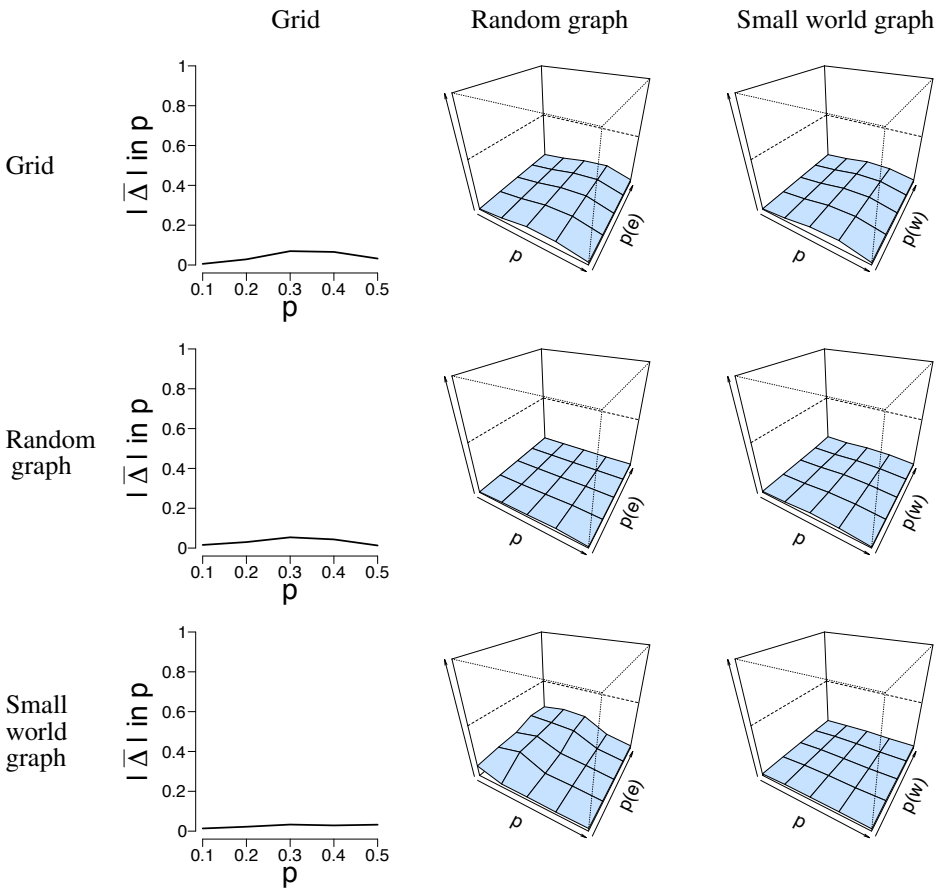


Figure 4.7: Visualisation of the mean absolute differences between p and $\hat{p}$ that resulted from the misspecification analysis with $n = 100$ nodes and $T = 50$ time points. The rows denote the structure for which the data were simulated. The columns denote the structure for which $\hat{p}$ was estimated. For the left columm , the x-axis denotes p for which we simulated data, and the y-axis the mean absolute difference between p and $\hat{p}$. For the middle and right column, the x-axis denotes the parameter p for which we simulated data, the z-axis the graph parameter that was used to simulate data, and the y-axis the mean absolute difference between p and $\hat{p}$ that ranges between 0 and 1. The mean absolute difference ranges between 0 and 1, where a lower value indicates a smaller difference between p and $\hat{p}$.

For each estimation we calculated the *Bayesian Information Criterion* (BIC) and compared it to the BIC of the other two network structures. The BIC is used for model selection, where the model with the lowest BIC is most preferred (Wit et al., 2012). We chose the BIC here over other information criteria because of its dependency on the sample size (Vrieze, 2012). As the sample size increases, the penalty of the BIC increases as well. We calculated the *Akaike Information Criterion* as well and obtained similar results. We therefore decided to only present the results for the BIC. Results showed that the grid structure was never the preferred network structure. The random graph is often selected (63.8% of the cases across conditions) as the preferred network structure when the data are simulated under a random graph. The small-world graph is preferred over the random graph at $p(e) = 0.1$ or $p(e) = 0.2$. A possible explanation for this is that, at this value for $p(e)$, the network is very sparse and it may be difficult to distinguish between the network structures. For data simulated under the small-world graph, the small-world graph itself is most often selected based on the BIC (69.3% of the cases across conditions). There are no conditions in which the random graph is preferred over the small-world graph. It is worthy to note that, as p increases, the difference in mean BIC between the network structures decreases, and more often the "incorrect" model is selected. This is also shown in Figure 4.4, where there is little difference between the bifurcation diagrams, especially when p is high.

As a last measure to study the robustness of our ML estimation, we performed a subset analysis, taking either 50% or 75% of the simulated time points to estimate $\hat{p}$. Similar to the misspecification analysis that we described previously, we looked at data with $n = 100$ nodes and $T = 50$ time points, and data with $n = 16$ nodes and $T = 5000$ time points. For each simulation condition, we randomly selected one simulation and selected a subset of the data, which we repeated 100 times. Figure 4.8 shows the mean error rate $\overline{\Delta}(p)$ between p and $\hat{p}$ for $n = 100$ nodes and $T = 50$ time points. It can be seen that the mean error rate is generally low for all conditions and network structures. This means that, even when we take a subset of the data, the mean field model is able to correctly estimate p

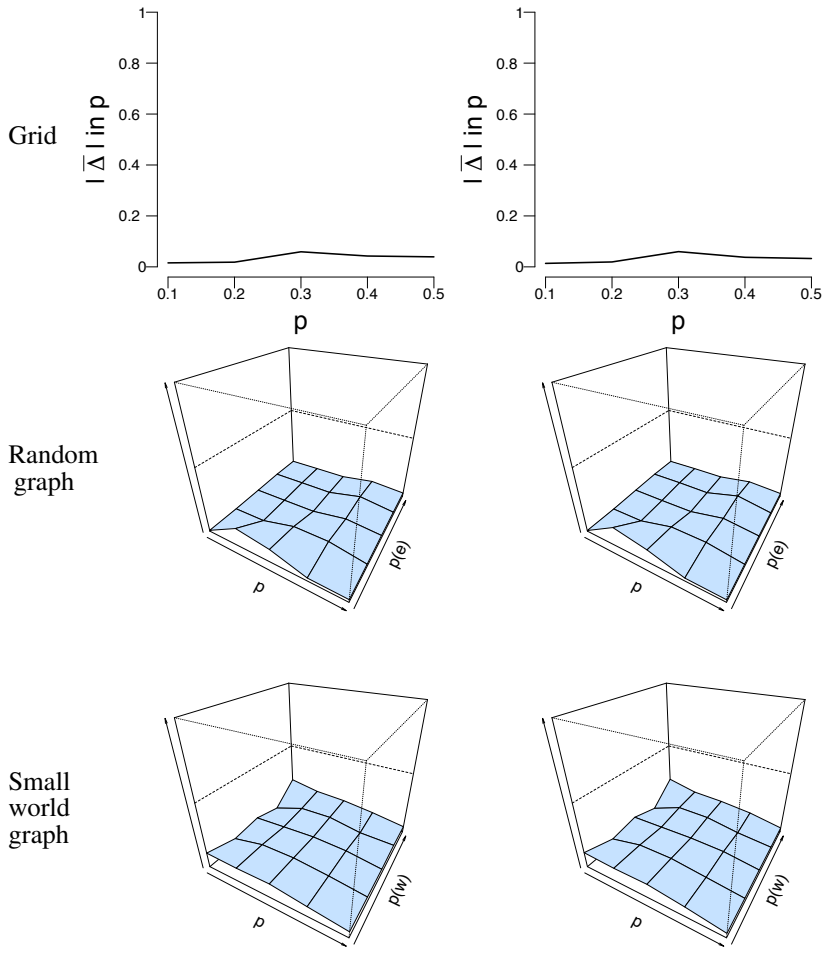


Figure 4.8: Visualisation of the mean absolute differences between p and $\hat{p}$ that resulted from the subset analysis with $n = 100$ nodes and $T = 50$ time points. The rows denote the structure under which the data were simulated and analysed. The left columns shows the result for the subset analysis with 50% of the data retained, while the right column shows the results with 75% of the data retained. The x-axis denotes the parameter p for which we simulated data, the z-axis the graph parameter that was used to simulate data (in case of the 3D figures), and the y-axis the mean absolute difference between p and $\hat{p}$. The mean absolute difference ranges between 0 and 1, where a lower value indicates a smaller difference between p and $\hat{p}$.

from the data that we used.

In sum, the mean field model estimates p well from the data; the graph parameters $p(e)$ and $p(w)$ could not be estimated as accurately. For the random graph parameter $p(e)$, this could potentially be solved by taking the ratio of edges present in the graph, and the total number edges possible in the graph. Alas, there is no similar solution for the small-world graph parameter $p(w)$. In the application of the mean field model, we assume all graphs to be random graphs. As estimating $p(e)$ from the data and extracting it from the graph resulted in nearly identical results, we decided to use the former option.

4.6 Application to empirical time-series data

Here, we will demonstrate how the probability p of an emotion to be active is estimated from empirical data. In the following sections, we will show two empirical examples and demonstrate how the proposed method works in each of these examples. By showing the application of our proposed method on two different kinds of data, we aim to show how our proposed method works for different participants, and different types of data. The first example is a dataset of patients who were admitted as patients to a closed, psychiatric ward of an academic hospital (Gordijn et al., 1994, 1998). The second example is a dataset of healthy participants who were originally recruited in a nation-wide study (van der Krieke et al., 2015).

The data in these examples are time-series data. When collecting these types of data, participants are asked to complete a questionnaire several times a day. These questionnaires often contain items regarding a participant's current mood state, but can also hold items regarding a participant's physical condition, for example. In both examples, participants received a 'beep' on fixed times during the day and were asked to complete the questionnaire. These beeps, in turn, correspond to the time points in time-series data. For example, when a participant completed twenty questionnaires, the data contains $T = 20$ time points. All analyses were performed using the R statistical software 3.4.4 (R Core Team, 2016).

Next to the estimation of the probability parameter p, we calculate the *Bimodality Coefficient* (BC; Hosenfeld et al., 2015) for each participant in both datasets, and compare the outcome of the two measures. The BC only takes information from the distribution of the proportion of active nodes (density) to determine whether there is evidence for one or two stable states. The BC is calculated as follows:

$$BC = \frac{s^2 + 1}{k + C} \tag{4.8}$$

where s is the skewness of the distribution, k the kurtosis of the distribution, and C a correction factor that depends on the number of variables: $C = \frac{3(n-1)^2}{(n-2)(n-3)}$. The BC obtains values between 0 and 1 and considers values > 0.55 to mean there is evidence for two states (Hosenfeld et al., 2015). We only use the BC for comparison, the BC uses no specific information or assumptions about the process, only distributional properties are involved. We assume that the process is essential for the assessment of a possible bimodal system, and as the BC and our proposed method are not mathematically interchangeable, we believe that these two methods should not necessarily correspond. The mean field model that we propose here takes the process that generated the data into account, which is an advantage in comparison to the BC.

4.6.1 Example 1: Clinical sample

This example involves a secondary analysis of data that were originally gathered for a study in patients diagnosed with MDD, who were admitted to a closed, psychiatric ward of an academic hospital (Gordijn et al., 1994, 1998). The data have been described in detail in previous papers (Gordijn et al., 1994, 1998). The study was approved by the medical ethical committee of the university hospital of Groningen, the Netherlands. Patients gave their written informed consent. Patients in this study completed the Dutch version of the Adjective Mood Scale (AMS; von Zerssen, 1986) twice a day at fixed time points for a period of six weeks, resulting in a maximum of 84 measurements per patient. Patients had to indicate on this 28-item questionnaire which of two given emotions (or

neither) corresponded most closely to the patient's emotion at that moment in time. A detailed description of the items of the AMS can be found in Table 4.2.

We dichotomized the data by collapsing the 'neither' condition with the positive mood state per individual item. We coded the positive mood state as '0' and the negative mood state as '1'. We also collapsed the 'neither' condition with the negative mood state and ran the analyses with these data, but as these results were very similar to the ones we present, we left it out of this study. After dichotomising the data, we replaced any missing measurements with the previous measurement. We also considered removing the missing measurements entirely, but as we found nearly identical results, we chose not to present these results.

A total of 82 patients were initially included in the study. Thirty three patients were excluded from the analyses due to either a too low number of measurements ($< 5; N = 4$), or a lack of variance in the response categories (smallest response category must contain at least 5% of the responses; $N = 29$). This resulted in 49 patients that were included in the analyses. Excluded patients (mean age = 48.79 years, SD = 14.09 years, 72.73% women) missed on average 28.10% of the measurements, and completed on average 60.39 measurements (SD = 30.33). These patients were admitted between 1988 and 1994, and were admitted on average for 209.45 days (SD = 119.59 days, min = 53 days, max = 536 days). Excluded patients completed significantly less measurements than included patients ($W = 587.5$, $p = 0.036$). Included patients had a mean age of 47.92 ($SD = 13.13$ years) at the time of admission to the closed ward, with 71.43% women. These patients missed on average 9.86% of the measurements, and registered on average 75.71 measurements (SD = 11.29). Patients were admitted between 1988 and 1994, and were admitted on average for 179.35 days (SD = 129.75 days, min = 49 days, max = 572 days). Mann-Whitney tests revealed that the excluded and included patients did not significantly differ in age ($W = 831$, $p = 0.835$), and admission period ($W = 755.5$, $p = 0.131$). Under the *EU General Data Protection Regulation*, we are not allowed to publish raw results. Result figures for all patients can be found online (Kossakowski, 2019).

	More	Dutch/English	More	Dutch/English	Neither/nor
1	0	Openhartig/openly	2	Geremd/inhibited	1
2	0	Welgemoed/good mood	2	Droefgeestig/bad mood	1
3	2	Inactief/passive	0	Bedrijvig/active	1
4	2	Ziekelijk/sickly	0	Kiplekker/healthy	1
5	0	Doelbewust/purposefully	2	Doelloos/aimlessly	1
6	2	Ernstig/serious	0	Geestig/humorous	1
7	2	Fantasieloos/unimaginative	0	Fantasierijk/imaginative	1
8	0	Gevoelig/sensitive	2	Gevoelloos/numb	1
9	2	Pessimistisch/pessimistic	0	Optimistisch/optimistic	1
10	0	Zorgeloos/carefree	2	Tobberig/worried	1
11	2	Gebroken/broken	0	Monter/cheerful	1
12	0	Liefderijk/lovingly	2	Liefdeloos/loveless	1
13	2	Schuldig/guilty	0	Onschuldig/innocent	1
14	2	Uitgeput/tired	0	Uitgerust/rested	1
15	2	Levensmoe/life-tired	0	Levenslustig/lively	1
16	0	Goed/good	2	Slecht/bad	1
17	0	Vrolijk/cheerful	2	Treurig/tearful	1
18	0	Bemind/loved	2	Onbemind/unloved	1
19	2	Lui/lacking in energy	0	Actief/energetic	1
20	2	Gesloten/withdrawn	0	Open/sociable	1
21	0	Levendig/lively	2	Levenloos/sluggish	1
22	0	Temperamentvol/temperamentfull	2	Futloos/lifeless	1
23	0	Oplettend/watchful	2	Verstrooid/absent	1
24	2	Wanhopig/desperate	0	Hoopvol/hopeful	1
25	0	Tevreden/satisfied	2	Ontevreden/dissatisfied	1
26	2	Angstig/anxious	0	Strijdlustig/combative	1
27	0	Krachtig/powerful	2	Krachteloos/powerless	1
28	0	Evenwichtig/balanced	2	Gejaagd/agitated	1

Table 4.2: Items of the Adjective Mood Scale (AMS) and their assigned labels. Items marked with a * have a reversed response scale. The English translation may differ from the original AMS scale, as well as the order of the items.

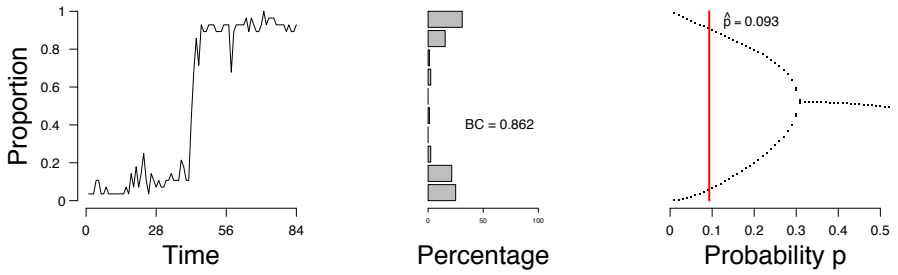


Figure 4.9: Proportion of active emotions (left panel), distribution of the proportion of active emotions (middle panel) and bifurcation diagram (right panel) of one participant from the the Groningen data. BC = bimodality coefficient. Red line indicates the estimate $\hat{p}$.

Figure 4.9 shows the evolution of the density (left panel), a distribution of the density ρ_t (frequency of the number of active nodes; middle panel), and the estimate of $\hat{p}$ in the bifurcation diagram (right panel) of a single patient. Figures of all patients are available online. According to the mean field model 87.8% of the patients had an expectancy for a transition. This is not surprising given that the sample is from a population of patients in a psychiatric ward. To compare, we calculated the bimodality coefficient (BC), which uses a function of the skewness and kurtosis from the distribution of the time series of the proportion of symptoms (see Hosenfeld et al., 2015 for details). The BC classified 59.2% of the cases as being bimodal. When we compare the results from the mean field model to the BC, we see that the methods agree in 55.1% of the cases, with $\kappa = 0.26$ (maximal $\kappa = 1$). In the case of the patient whose results are depicted in Figure 4.9, the BC is very high (0.86), which is reflected in the shape of the distribution of the density and corresponds to the result of the MFA.

We investigated the robustness of the mean field model in an empirical setting by running a subset analysis. This analysis is similar to the one we conducted with simulated data that is described earlier. We randomly selected either 50% or 75% of the time points per patient and used ML estimation to estimate $\hat{p}$. Results showed that in 96.3% of the

participants, taking a subset of the data resulted in the same conclusion according to the mean field model. For the BC, we found that, taking a subset of the data resulted in the same conclusion in 86.6% of the patients. This shows that the mean field model is fairly robust when one does not use all the data available.

4.6.2 Example 2: General sample

Participants were originally recruited in a nation-wide study called HoeGekIsNL (in English: HowNutsAreTheDutch) and have been described in detail in a previous paper (van der Krieke et al., 2015). The original study was approved by the medical ethical committee of the university medical center Groningen, the Netherlands. Informed consent was digitally obtained during the primary data collection. Participants in this study filled out a 43-item questionnaire that consisted of new items created for this study, and items from existing and validated questionnaires. Participants completed this questionnaire three times a day with a six-hour interval between the time points, for a period of 31 days, resulting in a maximum of 93 measurements per participant (van der Krieke et al., 2015).

From the original questionnaire, we selected items that pertained to mood states (21 items), appetite (one item) and laughter (one item), ending up with 23 items. Table 4.3 shows a detailed description of the included items. We recoded 10 positive items so that high scores indicate a more negative affect on all items. All included items were measured on a 0-100 scale. We dichotomized the data using a median split. This means that we calculated the median for each item for each participant, and split the data accordingly. We coded all the responses below the median as '0', and everything above the median as '1'. We also considered using a k-means clustering to dichotomise the data, but as these results were very similar to the results that we present, we chose not to include these results here. We replaced any missing measurements with the previous measurement. We also considered removing the missing measurements entirely, but as we found nearly identical results, we chose not to present these results.

Item	Meaning	Range
1	I feel relaxed	not at all (0) – very much (100)
2	I feel gloomy	not at all (0) – very much (100)
3	I feel energetic	not at all (0) – very much (100)
4	I feel anxious	not at all (0) – very much (100)
5	I feel enthusiastic	not at all (0) – very much (100)
6	I feel nervous	not at all (0) – very much (100)
7	I feel content	not at all (0) – very much (100)
8	I feel irritable	not at all (0) – very much (100)
9	I feel calm	not at all (0) – very much (100)
10	I feel dull	not at all (0) – very much (100)
11	I feel cheerful	not at all (0) – very much (100)
12	I feel tired	not at all (0) – very much (100)
13	I feel valued	not at all (0) – very much (100)
14	I feel lonely	not at all (0) – very much (100)
15	I feel I fall short	not at all (0) – very much (100)
16	I feel confident	not at all (0) – very much (100)
17	I worry a lot	not at all (0) – very much (100)
18	I am easily distracted	not at all (0) – very much (100)
19	I feel my life is worth living	not at all (0) – very much (100)
20	I am unbalanced	not at all (0) – very much (100)
21	I am in the here and now	not at all (0) – very much (100)
22	My appetite is..	much small than usual (0) – much larger than usual (100)
23	Since the last measurement I had a laugh	not at all (0) – very much (100)

Table 4.3: Items that were included in the analysis, the meaning of each item, and the response range in word and number.

A total of 974 participants participated in this study. We excluded 182 participants from the analyses due to a too low number of measurements (< 5), resulting in 792 participants that were included in the remainder of this section. Excluded participants (mean age = 41.17 years, SD = 13.56 years, 84.06% women) missed on average 88.57% of the measurements, and completed an average of 1.38 measurements (SD = 1.35). Excluded participants completed significantly less measurements than included participants ($W = 0$, $p < 0.001$). Included participants had a mean age of 40.21 (SD = 13.48 years) at the start of the data collection, with 82.49% women. These participants missed on average 35.81% of the measurements and registered on average 58.67 measurements (SD = 36.37). Mann-Whitney tests revealed that excluded and included participants did not significantly differ in age ($W = 75248$, $p = 0.353$). We also looked at the mean scores of the *Depression and Anxiety Stress Scale* (DASS; S. H. Lovibond & Lovibond, 1995; P. F. Lovibond & Lovibond, 1995), the *Quick Inventory of Depressive Symptomatology* (QIDS; Rush et al., 2003, 2006), and the *Positive Affect Negative Affect Scale* (PANAS; Peeters et al., 1996; Raes et al., 2009). Mann-Whitney tests revealed that excluded and included participants did not significantly differ on the DASS ($W = 10093$, $p = 0.194$), the QIDS ($W = 74275$, $p = 0.127$), the positive items of the PANAS ($W = 67314$, $p = 0.557$) or the negative items of the PANAS ($W = 72253$, $p = 0.366$). Under the *EU General Data Protection Regulation*, we are not allowed to publish raw results. Result figures for all participants can be found online (Kossakowski, 2019).

Figure 4.10 shows the evolution of the density (left panel), a distribution of the density (frequency of the number of active nodes; middle panel), and the estimate of $\hat{p}$ in the bifurcation diagram (right panel) of a single participant. Figures of all participants are available online. According to the mean field model 20.8% of the participants have an expectancy for a transition. This is not surprising given that the sample is from the general population. To compare, we calculated the bimodality coefficient (BC), which uses a function of the skewness and kurtosis from the distribution of the time series of the proportion of symptoms (see Hosenfeld et al. (2015) for more details). The BC classified

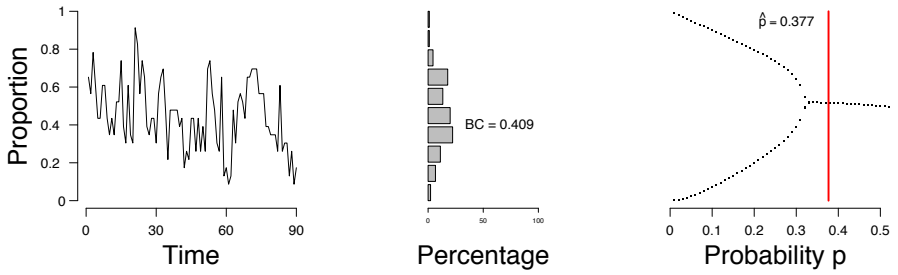


Figure 4.10: Proportion of active emotions (left panel), distribution of the proportion of active emotions (middle panel) and bifurcation diagram (right panel) of one participant from the HowNutsAreTheDutch data. Red line indicates the estimate $\hat{p}$.

31.9% of the participants as being bimodal. When we compare the results from the mean field model to the BC, we see that the methods agree in 61.1% of the cases, with $\kappa = 0.39$ (maximal $\kappa = 1$). In the case of the participant whose results are depicted in Figure 4.10, the BC is not that high (0.409); this is reflected in the shape of the distribution of the density, which has a unimodal shape. This corresponds to the MFA result which indicates stability.

We investigated the robustness of the mean field model in an empirical setting by running a subset analysis. This analysis is similar to the one we conducted with simulated data that is described earlier. We randomly selected either 50% or 75% of the time points per participant and used ML estimation to estimate $\hat{p}$. Results showed that in 85.5% of the participants, taking a subset of the data resulted in the same conclusion according to the mean field model. For the BC, we found that, taking a subset of the data resulted in the same conclusion in 81.6% of the participants. This shows that the mean field model is fairly robust when one does not use all the data available.

4.7 Discussion

The present study combined dynamical systems theory and network theory to assess the expectancy for a transition, a sudden jump between two stable mood states, using a mean field model. We provided a numerical illustration that shows that a mean field model can accurately identify (simulated) individuals who may expect to experience a transition. We then applied the mean field model to two different empirical examples: data from patients admitted to a closed ward, and data from a general sample from a nation-wide study. Results from these applications show how our proposed method works in practice.

A big question remains to be answered after this study: did the participants who were expected to transition actually had a transition between mood states? The analyses that we ran are of a probabilistic nature; expecting that a participant makes a transition between mood states does not mean that a participant actually makes this transition. Unfortunately, we do not have any follow-up measures to investigate whether or not these transitions actually occurred. From the patients in the clinical sample we know that they were eventually released from the closed ward. It can thus be hypothesised that this is an indication of a transition occurring in these patients. Future research could shed some light on this hypothesis by collecting data after patients are released from a (closed) ward. It would also be interesting to follow-up on the participants from the HowNutsAreTheDutch study to investigate if participants transitioned between mood states.

The data used in this study were collected in different decades. Data from the clinical sample were collected between 1988 and 1994, whereas data from the general sample were collected between 2014 and 2016. Between the time of data collection and the current time, the general view towards mental disorders like MDD has changed, and questionnaires and methods for collecting data have adapted with it. Even though the data from the clinical sample was collected thirty years ago, their approach to collecting the data (intensive data or time-series data collection) is an approach that is still used today, and is becoming more and more a common practice (e.g., Janssens et al., 2018). Also, the questionnaire used to collect these data has a different design than the questionnaire

used in the general sample, as shown in Tables 4.2 and 4.3. Nevertheless, we believe that we can draw similar conclusions from the results that were obtained with these datasets, as both questionnaires enquired about various aspects of a participant's mood in a similar way.

When collecting time-series data, participants are requested multiple times a day to fill out a questionnaire for a certain period of time. This type of data collection demands time and effort of the participants. It thus makes sense that participants sometimes forget to complete a questionnaire, or are simply not up for it at that specific moment, for whatever reason. In the data that we analysed, we came across different ratios of missing data and completed measurements, ranging from no missing measurements to almost as much as 90%. Since we assumed a Markov model and so, the item responses should not change much and thus, we replaced missing measurements with the previous measurement. Adopting this approach for handling missing data decreases the variance that individual items may have, thereby increasing the probability that a participant may be expected to experience a transition. Although we did not find evidence that our analysis differed much if we removed these measurements altogether, at this point in time, there is no clear picture of the effects of missing data in the current analysis. Future research should focus on mapping the effects that different types of missing data have on the current analysis, and what the effect of various imputation methods have on the analysis.

The current study only allows for binary and non-missing data. We applied different techniques for dichotomising the data and handling missing data. Even though these different approaches did not lead to different conclusions, the current approach may not be ideal. Data are often imperfect: low variance within item scores, as well as missing data occurs recurrently in time-series data. More importantly, it can be argued that MDD symptoms may not be binary, but categorical or even continuous. One can imagine that there exists a scale on which individual MDD symptoms lie. For example, two participants may experience insomnia (one of the MDD symptoms as listed in the Diagnostic and Statistical Manual of Mental Disorders American Psychiatric Association, 2013), but

the severity of this symptom may differ greatly between individuals. In the future we aim to expand the mean field model so that it allows continuous data as well as items with low variance.

In the clinical sample, we had to exclude quite a few patients as a result of data having too little variance. Too little variance here means that the response category with the smallest number of responses (either 0 or 1) included less than 5% of the total number of responses. We had to exclude these patients because the mean field model that we proposed in this study needs a minimal amount of variance to assess an individual's expectancy for a transition. The question can be raised here what it empirically means when an individual shows little variation in their item responses. Does it indicate that an individual can expect to transition between mood states, or that an individual varies so little in their emotions that a transition cannot be expected. Unfortunately the mean field model in its current state cannot solve this issue. More research is needed to theoretically decide what no variance in item responses means in a clinical setting, and to expand the mean field model so that it may account for little to no variance in the data.

The mean field model that we used in this chapter has three assumptions: (1) we assume that each node in a graph has the same neighbourhood size, (2) nodes can only be in one of two states (active/inactive) and (3), we assume that all nodes in a graph show equal behaviour. Waldorp & Kossakowski (2020) showed one can deviate from the first assumption, whilst maintaining a high accuracy in estimating the probability p. We discussed the second assumption in more depth previously, which leaves us with the final assumption of the mean field model. In the current study, we operationalised the third assumption by fixating the probability p to be equal for all nodes in the graph. However, it is unlikely that all symptoms of psychological disorders like MDD behave in a similar manner. For example, some individuals can handle sleep deprivation better than others. In this case, the "sleep problems" node would less easily be activated in individuals that can handle sleep deprivation in comparison to individuals that cannot handle sleep deprivation that well. A possible extension of the mean field model as is used in this chapter is to vary

the probability value p, which appears in the majority rule, for every node in the graph. In the example of sleep deprivation, we could operationalise the sensitivity difference by using different values for p between nodes.

In the current study, we estimated the probability p per individual for the entire time-series. This means that p cannot change between time points. One may wonder if this value is supposed to be static, or that it could change between time points. The advantage of a static probability value is that it is easy to estimate. However, a static probability value may not reflect an individual's expectancy for a transition accurately. By allowing the probability p to vary over time, one could gain more insight into how an individual moves throughout time with respect to p. One possible method to accomplish this is to work with a moving window, in which one uses a window to select a snippet within the time series to estimate p, and let that window move throughout the time series. In this situation, we can estimate p several times on different segments of the time series; the size of the window will determine how many values are estimated. In the future we hope to expand the mean field model and allow for the probability p to vary.

In the numerical illustration section of our current study we only looked at values for p, $p(e)$ and $p(w)$ between 0.1 and 0.5. Since these parameters are probabilities, their theoretical range lies between 0 and 1. Although we did run the numerical illustration for values up to 0.9, we chose not to present them as values rarely occur in empirical data. Also, at higher values for $p(e)$ and $p(w)$, the clustering within the network structures increases and can create some strange behaviours that are beyond the scope of this chapter. A possible solution when dealing with high clustering values within a network is to switch to a so-called *scale-free degree distribution*.

Based on the simulated and empirical examples provided in this study, we believe that the mean field model is a promising method. We do emphasise that the predictions of our proposed model have not been verified using empirical evidence. We surely must investigate further to what extent the proposed method could be useful in clinical practice, but depending on the possible adjustments of the probability or majority rule in the model,

the validity of the method could be high and therefore useful.

Part II

Causal Systems in Psychopathology

5

The Search for Causality: A Comparison of Different Techniques for Causal Inference Graphs

This chapter is submitted as: Kossakowski, J.J., Waldorp, L. J., & van der Maas, H. L. J. (2020). The Search for Causality: A Comparison of Different Techniques for Causal Inference Graphs. *Under review at Psychological Methods*.

Abstract

Many are interested in causal relations between two or more variables. Establishing a causal relation between two variables can help us in answering that big question of why something happens. However, using solely observational data is insufficient to get the complete causal picture. The combination of observational and experimental data may give enough information to properly estimate causal relations. In this study, we consider the conditions where estimating causal relations might work and we show how well five different algorithms determine causal relations in a simulation study. Results showed that two algorithms, based on the principle of invariant causal prediction, perform best in most simulation conditions. We believe that the combination of the two algorithms may be suitable to be used in future research.

5.1 Introduction

> Some tens of thousands of years ago, humans began to realise that certain things cause other things and that tinkering with the former can change the latter. No other species grasps this, certainly not to the extent that we do. From this discovery came organised societies, then towns and cities, and eventually the science- and technology-based civilisation we enjoy today. All because we asked a simple question: Why? (Pearl & Mackenzie 2018)

The quest for causality is one that people have been striving for for decades. Establishing a causal relation between two phenomena or variables can help us in answering that big question of why something happens. In psychology, we study the (possible) causal relation between psychological constructs, like sleep, concentration, or feelings of guilt. For example, does sleep deprivation lead to concentration problems? And could sleep deprivation be caused by increased feelings of guilt? Knowing what the cause is of something so intrusive as sleep problems may in turn lead to finding the solution to help

an individual who experiences sleep problems. If we know what causes a problem, we can help to solve it.

What is a causal relation? Although much can be said and has been said about causal relations, we confine ourselves here to an interventional (or, equivalently, a counterfactual) definition of causality. The idea is that, if one changes (perturbs) variable X, then this should only have an effect on variables with which X has a causal relation. For instance, if we consider the structure $X \rightarrow Y \leftarrow Z$, then we expect that changing X will change Y, but this change in X will not change Z. Therefore, we use the following definition of a causal relation: "a relation between two variables ($X \rightarrow Y$) where, when one changes one variable (X), one observes a change in the other variable (Y)". In general, when more than three variables are involved, we require that, conditioned on all other variables, a direct cause X is one that changes the distribution of Y. If we go back to the example of sleep problems, concentration problems and feelings of guilt, our definition of a causal relation states that, when there is an change in sleep problems, there should be a change in the level of concentration as well, or any other aspect of the distribution of the effect variable. The definition of a causal relation that we use here also implies a counterfactual relationship (Pearl, 2009). If there is a causal relation from X to Y, our definition implies that when we do not observe a change in Y, no change has occurred in X. It may be argued that our definition is not sufficient to capture all aspects of a causal relation. For instance, we ignore the question of what kind of events could have a causal relation, thereby interpreting the causal relations themselves. We confine our definition to a rather practical version which allows one to statistically determine directional effects.

The type of data that is most often used to estimate causal relations between variables are *observational data*. These are (empirical) data in which no perturbations have taken place. Observational data includes cross-sectional data that one collects with questionnaires, for example. The most widely used technique to estimate causal relations with observational data is the algorithm developed by Pearl (2009) and Spirtes et al. (2000). Pearl uses the notion of *(conditional)* dependence and independence between sets of three

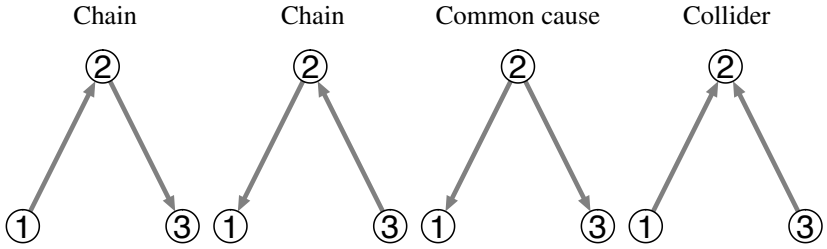


Figure 5.1: The different causal structures that can be detected using conditional (in)dependence. Arrows indicate causal relations. The chain structures (first and second panel from the left) and the common cause structure (third panel from the left) are statistically equivalent, whereas the collider structure (fourth panel) is statistically unique.

variables to determine a causal relation. The ideas from Pearl & Verma (1991) and Spirtes et al. (2000) indicate that, if one were to solely use multivariate normal observational data, we can infer causal relations using the notion of conditional (in)dependence. Based on the raw (simple, Pearson) and partial correlations, four different causal structures can be obtained, as shown in Figure 5.1. In the first three situations (the two chain structures and the common cause structure), nodes 1 and 3 have a nonzero correlation, but their partial correlation is zero when conditioning on node 2. In the fourth structure (collider structure), nodes 1 and 3 have a zero correlation, but a nonzero partial correlation when conditioning on node 2 (See Appendix D.1 for more details on this).

As the rules for conditional independence are equal for the first three causal structures, they are statistically equivalent, and therefore one cannot distinguish them from one another. It is only possible to identify the fourth (collider) structure from the other three (Pearl, 2009, but see Mooij et al., 2016, for some interesting cases). These ideas have been used in different methods to obtain causal relations. Tetrad (Glymour & Scheines, 1986) applied a conditional independence test to each possible alternative path, while Gaussian Graphical Models (GGMs; Drton & Richardson, 2004) use a likelihood based method for a complete set of conditional independencies. Temporal ordering has also been used (Hamaker et al., 2015; Usami et al., 2019; Zyphur et al., 2019). However,

using observational data exclusively will not resolve all causal relations.

This led Granger (1980) to state that an "observed relationship does not allow one to say anything about causation between the variables", and Holland (1986) argued that there can be "no causation without manipulation". Although one can use observational data to estimate some causal relations, this alone is not enough to properly estimate all relations between variables. As implied by our definition of a causal relation, one needs to perturb one variable and observe its effect in order to establish causal relations between variables. This means that we also need so-called *experimental* data to estimate causal relations. These are (empirical) data where some perturbation has taken place. Examples include psychological experiments where a perturbation takes place by assigning participants to different conditions, or questionnaire research where participants' attitude are changed with hypothetical scenarios (see Hoekstra et al., 2018, for an empirical example). The combination of observational and experimental data may give us a complete picture of the causal relations between variables, which in turn may be used by professionals to set up a treatment plan where the causes of constructs like concentration problems are attacked, instead of just the effect. We need observational data to see what the relation between two variables is without any perturbations, and experimental data to see how variables change as an effect of perturbing one or more variables.

The goal of this chapter is two-fold: (1), we want to provide an overview of a set of algorithms that stem from different fields, describing and illustrating each algorithm. The second goal is to investigate how well each of these algorithms can estimate causal relations by means of a simulation study. First, we will describe the algorithms that can be used to estimate causal relations. For each algorithm we use a simulated dataset as an illustration. Then we will describe the simulation study that we have set up to test not only the performance of these techniques individually, but also in comparison to one another.

5.2 Methods of causal inference

The goal of this study is to compare different algorithms for inferring causal graphs. We focus here on *Directed Acyclic Graphs* (DAGs). These are graphs that only show causal (directed) relations between variables, and that do not contain any feedback loops (acyclic). Furthermore, we assume that all (conditional) independencies are obtained in the graph and vice versa (the Markov and faithfulness conditions, see Appendix D.1 for more details).

Identifying causal relations is not an easy task. Take for example the causal graph shown in Figure 5.2, where one can see that there is no direct relation between variables 2 and 5. There is a correlation between variables 2 and 5 due to the chain structures $2 \longrightarrow 3 \longrightarrow 5$ and $2 \longrightarrow 4 \longrightarrow 5$. So, there are three possible paths. The trick is then to remove the path $2 \longrightarrow 5$ in this case. Two methods are generally used. The first is called *transitive reduction* (Klamt et al., 2010; Pinna et al., 2013). Here, a causal graph is set up, and direct connections are removed if there is enough evidence to suggest that two variables are not directly connected. When there is a direct causal relation between two variables, any alternative path between these variables should be removed with transitive reduction. However, when a direct causal relation is small, algorithms that use transitive reduction will erroneously remove the direct connection in favour of the alternative paths. Transitive reduction thus may not also work in practice. We expand on this more in Appendix D.2.

The second method is by conditioning on the remaining variables, both by using observational and experimental data (Meinshausen et al., 2016; Peters et al., 2017). In our example this is guaranteed to work, since conditioning on both 3 and 4 will remove the correlation between variables 2 and 5. Here, we investigate the accuracy of various algorithms for causal graphs. To explain these algorithms, we will use one simulated causal graph and associated dataset that contain five variables, visualised in Figure 5.2. For illustration purposes we simulated data for 1000 measurements. We will compare five algorithms: the *Peter and Clark* algorithm (PC; Kalisch & Buhlmann, 2007), the *Down-*

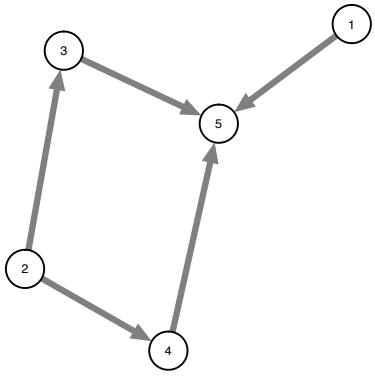


Figure 5.2: Visualisation of the causal graph that we use to illustrate the different algorithms. Arrows represent causal relations between individual variables, which are depicted as circles.

Ranking of Feed-Forward Loops algorithm (DR-FFL; Pinna et al., 2013), the *Transitive Reduction for Weighted Signed Digraphs* algorithm (TRANSWESD; Klamt et al., 2010), the *Invariant Causal Prediction* algorithm (ICP; Meinshausen et al., 2016) and the *Hidden Invariant Causal Prediction* algorithm (HICP; Peters et al., 2017). We chose to include the PC-algorithm, even though it only uses observational data, to compare its results to algorithms that include experimental data next to observational data. Also note that there are more algorithms that can be used to estimate causal relations. We chose to restrict our study to these five algorithms for clarity of presentation.

The data used to illustrate the different algorithms are publicly available, so that the reader may use the data to replicate our examples. Throughout this section, we will use p to denote the number of nodes that exist in the causal graph, n to denote the number of measurements, and e_{ij} to denote the directed edge $i \to j$. Symbols that are associated with specific algorithms will be explained when we introduce the symbol for the first time. At the end of this section, we provide a summary table (Table 5.1) that gives an overview of the algorithms that are discussed here, and their properties.

5.2.1 Peter and Clark Algorithm

The PC-algorithm (Spirtes et al., 2000) has a two-step process that solely uses observational data (in contrast to the other algorithms that will be discussed later). We used the *R*-package pcalg (version 2.6-2; Kalisch et al., 2012) to run the PC-algorithm. The first step in the PC-algorithm is to find the *skeleton* of the causal graph. A skeleton is an undirected graph that shows all edges that are possible causal relations. For each node individually, we look at every possible relation with every other remaining node in the graph. The raw correlation between each pair of nodes is calculated, after which partial correlations are calculated between every pair of nodes, conditioning on subsets of the remaining variables, increasing in size of the subsets. All possible partial correlations are calculated until either the algorithm has calculated the partial correlation for all possible subsets, or if a partial correlation returns zero when conditioning on a specific subset.

In the second step of the PC-algorithm, the direction of the relation is determined by considering *collider* structures (fourth panel, Figure 5.1). A collider structure is a causal structure where a node is caused by two other nodes. Because the correlation pattern for a collider (a nonzero partial correlation between nodes 1 and 3 when conditioning on node 2) is different from the chain and common cause structure (zero partial correlation between nodes 1 and 3 when conditioning on node 2), the collider structure can be distinguished, and hence gives information about the direction of the causal relations. By looking for collider structures first, the PC-algorithm can already determine the direction of specific causal relations. This facilitates the search for the chain or common cause structure. If the direction of specific edges is already determined, the PC-algorithm can eliminate possible causal structures that are no longer optional.

To illustrate the PC-algorithm, we use the illustration data that is based on Figure 5.2. This is a $n \times p$ matrix with $n = 1000$ and $p = 5$. The PC-algorithm needs the correlation and partial correlation matrices between the nodes in the graph. As the data are simulated from a normal distribution, we calculated Pearson's r correlation between every pair of nodes, as well as the partial correlations (conditioning on all remaining variables):

$$r = \begin{pmatrix} 1.000 & & & & \\ -0.028 & 1.000 & & & \\ -0.019 & 0.711 & 1.000 & & \\ -0.073 & 0.717 & 0.517 & 1.000 & \\ 0.298 & 0.723 & 0.759 & 0.748 & 1.000 \end{pmatrix} \quad r_p = \begin{pmatrix} 1.000 & & & & \\ -0.006 & 1.000 & & & \\ -0.468 & 0.389 & 1.000 & & \\ -0.515 & 0.392 & -0.467 & 1.000 & \\ 0.686 & 0.038 & 0.685 & 0.693 & 1.000 \end{pmatrix} \tag{5.1}$$

In the first step of the PC-algorithm, we determine the skeleton of the causal graph. Figure 5.3 (left panel) shows the skeleton based on our illustration data, using a significance level of 0.05. It can be seen that most edges are already in the correct location, when we compare it to the true causal graph (Figure 5.2). The only exception is e_{14}. It is possible that small correlations (like the raw correlation between variables 1 and 4) become significant faster with a larger sample size, even though this should theoretically not be the case if we look at the true causal graph (Figure 5.2). Only when we condition on node 5 should the correlation between nodes 1 and 4 be significantly different from zero.

In the next step of the PC-algorithm, we determine the direction of the relation between those nodes for which an edge was found in the skeleton graph. The algorithm first determines the collider structures, which exist as node 5 is the effect of nodes 1, 3 and 4. After determining the collider structures, the PC-algorithm continues with determining the direction of the remaining edges. Figure 5.3 (right panel) shows the final result of our illustration. Four out of six edges that are present are correctly identified. What is interesting in this example is the edge e_{23}. This edge is undirected, as the PC-algorithm could not determine a direction. This makes sense when we look at the causal structure that is formed between nodes 2, 3 and 5, and between nodes 3, 2 and 4. No matter the direction of the edge between nodes 2 and 3, the causal structures between nodes 2, 3 and 5, and between nodes 3, 2 and 4 will remain statistically equivalent. It is therefore impossible for the PC-algorithm to determine a direction. This illustration shows one of the prime disadvantages of the PC-algorithm, in that it sometimes cannot determine the causal direction of an edge at all. The other edge that stands out is e_{14}. We already saw

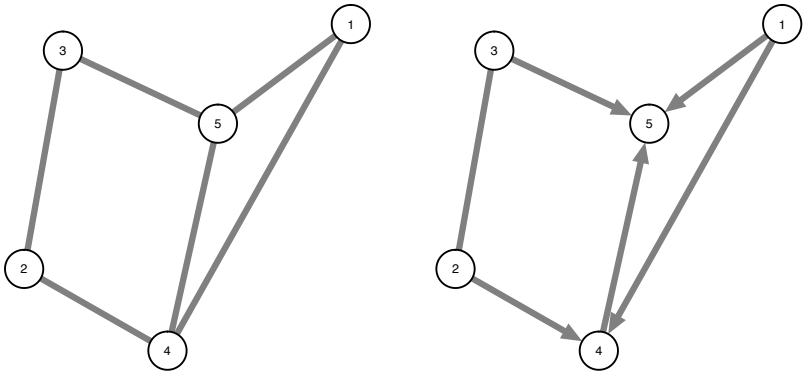


Figure 5.3: Visualisation of the skeleton (left panel), and the causal graph (right panel) estimated with the PC-algorithm (Kalisch et al., 2012) available in *R*.

this (spurious) edge in the skeleton graph. In the final causal graph, one can see that the edge creates a collider structure with e_{24}. This indicates that the algorithm found a significant partial correlation between nodes 1 and 2 when conditioning on node 4.

Overall, in this illustration the PC-algorithm performs reasonably well, only one edge is incorrectly estimated, and one edge is left undirected. Most of the edges that are present in the true causal graph are correctly estimated.

5.2.2 Down-Ranking of Feed-Forward Loops Algorithm

The PC-algorithm is solely suitable for observational data. The Downward Ranking of Feed-Forward Loops algorithm (DR-FFL; Pinna et al., 2013) has an advantage over the PC-algorithm in that it uses both observational data and experimental data to estimate a causal graph. The DR-FFL algorithm (Pinna et al., 2013) originates from the field of gene biology and estimates unweighted, unsigned causal graphs for single subjects and single measurements (where each node was perturbed once). Unweighted here means that the algorithm will not indicate the strength of the causal relation between two nodes, and unsigned means that the algorithm will not indicate whether the causal relation between

two nodes is positive or negative. The DR-FFL algorithm employs a two-step process. In the first step, the algorithm compares the effect of perturbing a node to the average effect that includes the observational data as well to create a *perturbation graph* (PG). The DR-FFL algorithm then applies *transitive reduction* to remove direct causal relations from the perturbation graph where indirect effects are in order. See Appendix D.2 for an extended description of the notion of transitive reduction.

The DR-FFL algorithm needs two components to infer a causal graph: observational data for each of the nodes (G^{wt}; also known as wild-type data) and experimental data (G^{ko}; also called knock-out data) where each node in the data is perturbed. The experimental data is a $p \times p$ matrix where the rows correspond to the node that is perturbed, and the columns to all the nodes in the graph and thus the nodes that the perturbed node may have an effect on. For example, row 1 of such a knock-out matrix will depict the new values that the nodes in the graph have after perturbing node 1.

The first step in the DR-FFL algorithm is the generation of the PG. This graph will show only those causal relations between nodes for which the perturbation effect was strong enough to suggest a causal relation. In order to determine how strong a perturbation effect is, we first calculate absolute z-scores for each possible edge in the graph:

$$|z_{ij}| = \left| \frac{G^{ko}_{ij} - \mu_j}{\sigma_j} \right|, \tag{5.2}$$

where μ_j is the mean of all perturbation effects on node j, and σ_j the standard deviation of node j across different perturbations. Both μ_j and σ_j are calculated using both the observation for node j and the perturbation effects of all nodes on node j. The PG is then generated by selecting those edges whose $|z|$-score is larger than a pre-specified threshold β.

The second step of the DR-FFL algorithm is transitive reduction. In this step, the algorithm narrows its search to edges that connect strongly connected components. A strongly connected component is a (sub)set of nodes where any node can be reached (i.e., there must be a directed path) from any other node in the component. The DR-

FFL algorithm only focuses on edges between strongly connected components because cycles exist between the nodes within a strongly connected component. For each edge e_{ij} that connects two strongly connected components, the DR-FFL algorithm searches for alternative paths, and removes the direct edge e_{ij} if the alternative path satisfies two criteria (1), edge e_{ij} can only be removed when e_{ij} connects different strongly connected components in the PG and (2), edge e_{ij} can only be removed when there is an alternative route from node i to node j without using e_{ij}.

To illustrate the DR-FFL algorithm, we again use the data simulated using the causal graph shown in Figure 5.2. For this illustration, we averaged over the sample size ($N = 1000$), thereby creating set of observational data G^{wt} and experimental data G^{ko}. We chose to deviate from the single-subject objective of this algorithm to highlight the differences between the DR-FFL algorithm and the TRANSWESD algorithm that is described in the next section. Below you find the observational and experimental data that is created to illustrate the DR-FFL algorithm.

$$G^{wt} = (-0.015, 0.025, -0.001, 0.013, 0.015) \tag{5.3}$$

$$G^{ko} = \begin{pmatrix} -0.018 & 0.115 & -0.003 & 0.064 & 0.072 \\ -0.070 & -0.012 & -0.004 & 0.063 & 0.073 \\ -0.079 & 0.128 & 0.134 & 0.067 & 0.075 \\ -0.073 & 0.108 & -0.004 & 0.082 & 0.080 \\ -0.093 & 0.105 & -0.003 & 0.079 & -0.032 \end{pmatrix} \tag{5.4}$$

In this knock-out matrix, element G^{ko}_{25} represents the new value that node 5 has after perturbing node 2. We first calculate the $|z|$-scores that indicate the existence of a causal relation. Using the data that we described earlier, we obtain the following $|z|$-scores:

$$|z| = \begin{pmatrix} 0.000 & 0.651 & 0.405 & 0.090 & 0.542 \\ 0.370 & 0.000 & 0.425 & 0.072 & 0.567 \\ 0.641 & 0.872 & 0.000 & 0.233 & 0.611 \\ 0.452 & 0.513 & 0.423 & 0.000 & 0.718 \\ 1.047 & 0.463 & 0.416 & 0.693 & 0.000 \end{pmatrix} \tag{5.5}$$

The PG, shown in the left panel of Figure 5.4, is created using a threshold $\beta = 0.60$, an arbitrary value. In this illustration, four strongly connected components exist (see the middle panel of Figure 5.4): nodes 4 and 5 form a strongly connected component (component A), and nodes 1, 2 and 3 each from their own individual component (components B, C and D, respectively). There are only five edges that connect these strongly connected components, shown by the middle panel of Figure 5.4.

For each of these five edges, the DR-FFL algorithm determines whether an alternative path exists to connect these two components. There are no alternative paths between components A and B, components A and C, and components A and D. There is an alternative path between components D and B ($D \rightarrow A \rightarrow B$) and components D and C ($D \rightarrow A \rightarrow B \rightarrow C$). Thus, the edges that directly connect components D and B (e_{31}) and components D and C (e_{32}) are removed from the causal graph, resulting in the graph shown in Figure 5.4 (right panel). The resulting causal graph is an unweighted and unsigned graph, meaning that the edges do not indicate any strength between the nodes, nor does it indicate whether this causal relation is positive or negative.

Overall, the DR-FFL algorithm does not perform well in this illustration. Only two edges that exist in the true causal graph (Figure 5.2) are also estimated here. Two edges are estimated in the wrong direction, one edge is incorrectly estimated and one edge is incorrectly absent from the graph.

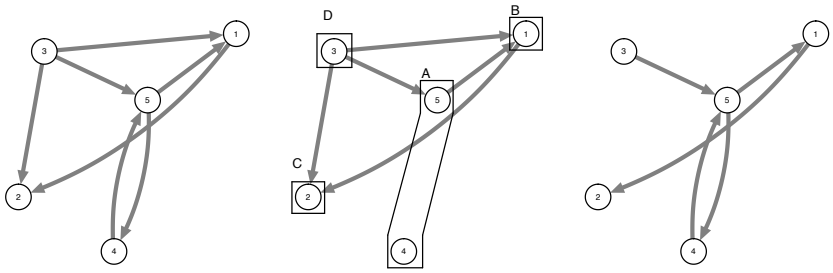


Figure 5.4: Visualisation of the process of the DR-FFL algorithm. The left panel denotes the perturbation graph in which present edges represent potential causal relations whose effect where strong enough (using a threshold value $\beta = 0.60$). The middle panel depicts the edges between strongly connected components that can be removed in the transitive reduction step of the DR-FFL algorithm. The black boxes around the nodes in the middle panel indicate the strongly connected components. The right panel depicts the final causal graph that results from the DR-FFL algorithm.

5.2.3 Transitive reduction for weighted signed digraphs

A disadvantage of the DR-FFL algorithm is that its resulting causal graph is unweighted and unsigned. This means that the algorithm does not add any weight to the causal relations, nor does it indicate if the causal relation is positive or negative. The TRANSitive reduction for WEighted Signed Digraphs (TRANSWESD; Klamt et al., 2010; Pinna et al., 2013), which has a similar approach as the DR-FFL algorithm, does return a causal graph with weighted edges that indicate a positive or a negative relationship, while applying transitive reduction to estimate a causal graph at the same time (See Appendix D.2 for an extended description of the notion of transitive reduction). Furthermore, where the DR-FFL mostly handles single-subject data (it can handle between-subjects data by averaging over measurements, although the effect of doing this is unknown), the TRANSWESD algorithm can be solely applied to between-subjects data. Like the DR-FFL, the TRANSWESD algorithm uses observational and experimental data to estimate the causal graph.

As a first step, we generate the PG. Like the DR-FFL algorithm, we calculate $|z|$-scores. In addition to the $|z|$-score, we calculate an absolute change score $|d|$ between G^{wt} and G^{ko} that shows the absolute effect of perturbing a node. Edges are retained in the *PG* when their associated $|d|$-scores exceed a pre-specified threshold γ. Each edge in the PG gets a sign s_{ij} that reflects the direction of the change that node j has made after node i was perturbed: if the change score is positive, then the edge will be blue, and when the change score is negative, the edge will be red. Each edge also has a weight w_{ij} that reflects the certainty of the causal relation, where a higher weight indicates a lower certainty. The weight w_{ij} is determined by $1-|\rho_{ij}|$, where ρ_{ij} is the *conditional correlation* (Rice et al., 2005). A conditional correlation is a correlation between two nodes i and j, given that node i was perturbed, and is calculated as follows:

$$\rho_{ij} = \frac{\sum_{a=1}^{2n}(x_{i,a}-\bar{x}_i)(x_{j,a}-\bar{x}_j)}{\left(\sum_{a=1}^{2k}[x_{i,a}-\bar{x}_i]^2\right)^{1/2}\left(\sum_{a=1}^{2n}[x_{j,a}-\bar{x}_j]^2\right)^{1/2}}. \tag{5.6}$$

For nodes i and j equation (5.6) only uses the observational data and the experimental data where node i was perturbed. This gives us two vectors, $x_{i,a}$ and $x_{j,a}$, each of length $2n$, assuming the same number of data points for observational and experimental data. Parameters $\bar{x}_i$ and $\bar{x}_j$ represent the means of these two vectors.

In the second step of the TRANSWESD algorithm, transitive reduction is applied, just like in the DR-FFL algorithm. In this step, the algorithm removes any edge e_{ij} when there is an alternative path between nodes i and j and when that alternative path satisfies the following four conditions: (1) the alternative path must not contain a cycle, (2) the alternative path cannot contain the edge e_{ij} that is under consideration, (3) the overall sign of the alternative path must be equal to that of the edge e_{ij} under consideration (obtained by multiplying the signs of all edges on the alternative path) and (4), the maximum weight of all edges on the alternative path must be lower than the weight of the edge e_{ij} under consideration multiplied by a pre-specified threshold α. For all analyses, we set $\alpha = 0.95$, the default value used by Klamt et al. (2010). All edges that exist in the PG are sorted based on their edge weight. The transitive reduction starts with the edge that has the

highest weight (and thus the lowest certainty).

To illustrate the TRANSWESD algorithm, we use the simulated data that is used throughout this overview. Equation (5.4) depicts the observational and experimental data. Equation (5.7) depicts the absolute change scores between the observational data G^{wt} and the experimental data G^{ko}.

$$|d| = \begin{pmatrix} 0.000 & 0.131 & 0.013 & 0.079 & 0.087 \\ 0.096 & 0.000 & 0.029 & 0.038 & 0.048 \\ 0.079 & 0.129 & 0.000 & 0.068 & 0.076 \\ 0.086 & 0.094 & 0.017 & 0.000 & 0.067 \\ 0.108 & 0.090 & 0.018 & 0.064 & 0.000 \end{pmatrix} \tag{5.7}$$

The PG, shown in Figure 5.5, is created using thresholds $\beta = 0.5$ and $\gamma = 0.05$, which are arbitrary values. The conditional correlations calculated with equation (5.6) are used to determine the edge weights. The conditional correlation matrix and the resulting edge weights are listed below:

$$\rho = \begin{pmatrix} 0.00 & -0.015 & -0.034 & -0.008 & 0.032 \\ -0.039 & 0.000 & 0.068 & 0.108 & 0.087 \\ -0.029 & 0.965 & 0.000 & 0.695 & 0.707 \\ -0.034 & 0.953 & 0.676 & 0.000 & 0.700 \\ 0.344 & 0.755 & 0.788 & 0.804 & 0.000 \end{pmatrix} \tag{5.8}$$

Note that, due to the design of the conditional correlation, the resulting matrix is not symmetric. Each edge is assigned a weight w_{ij}. The lower w_{ij} (the closer it is to 0), the more certainty there is about the existence of the edge. For example, $w_{41} = -0.034$ has a very low weight, and thus a high certainty. On the other hand, $w_{32} = 0.965$ is relatively high, so there is low certainty when it comes to that specific edge.

All edges in Figure 5.5 are sorted based on their edge weight. The transitive reduction starts with the edge that has the highest weight (and thus the lowest certainty). In the PG in Figure 5.5, five edges (e_{12}, e_{31}, e_{32}, e_{35}, and e_{42}) contain a cycle on their alternative

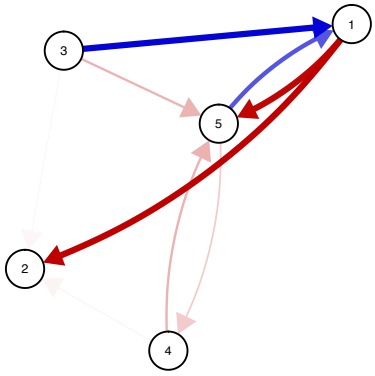


Figure 5.5: Visualisation of the perturbation graph generation. Blue edges indicate positive causal relations, and red edges denote negative causal relations. The thickness and saturation of the edge colour indicate the strength of the causal relation.

paths (condition 1), and the same five edges are the only edges that have an alternative path (condition 2). This means that no edges are removed from the causal graph, and that the perturbation graph in Figure 5.5 will not change after the transitive reduction step. If we were to look at the third and fourth condition, we see that, of these five edges, four (e_{12}, e_{32}, e_{35}, and e_{42}) satisfied the third condition (the product of the signs of the alternative path must match the sign of the edge e_{ij} that is under consideration). All of these edges did not meet the final requirement that states that the maximum weight of all edges on the alternative path cannot exceed the weight of the edge e_{ij} under consideration multiplied by α.

Similar to the DR-FFL algorithm, the performance of the TRANSWESD algorithm seems subpar. Three edges that exist in the true causal graph (Figure 5.2) are correctly estimated, two edges are estimated in the wrong direction and four edges are incorrectly estimated.

5.2.4 Invariant causal prediction

The ICP-algorithm (Meinshausen et al., 2016) combines both the advantage of the PC-algorithm in that it considers a multivariate system (and not a two-step procedure like the DR-FFL and the TRANSWESD algorithms), and uses both observational and experimental data in a single analysis. Another advantage of the ICP-algorithm is that the perturbations inflicted on the data do not have to be node-specific: perturbations can be non-specific and generic for subsets of nodes. The core assumption of the ICP-algorithm is that the conditional distribution of an individual node, controlling for its direct causes, does not change across perturbations (Peters et al., 2016). In other words, a causal relation between two nodes only exists when the residuals do not change when a perturbation has taken place on any node, except for the receiving node, called the *target node* here.

The ICP-algorithm needs two components: the raw data (both observational and experimental data), and an identifier variable ε. We use ε to distinguish between different perturbations, which we call *environments*, following Peters et al. (2016). This is similar to the experimental data matrix used in the DR-FFL and TRANSWESD algorithms, where the rows indicate each separate perturbation. Another example of a situation where multiple environments exist is in datasets where every participant is measured on two or more time points. Every time point is then a unique environment. The minimal requirement is that the data must have at least 2 environments. Typically, one environment consists of observational data.

The ICP-algorithm works node-wise, each time by selecting a *target node*. We then use the remaining nodes to identify all possible subsets. Subsets can range from an empty subset (where the target node had no cause) to a subset that contains all remaining nodes. After we determined the possible subsets, the ICP-algorithm regresses the target node onto each subset, and obtains its associated residual distribution. The ICP-algorithm then splits the residuals up according to the environment it belongs to, and it will compare the residuals of one environment against the residuals of all remaining environments using a Kolmogorov-Smirnov test. Subsets whose residual distribution are equal across envi-

ronments are called "invariant"; only those nodes that are part of each invariant subset (intersection) are said to be causes of the target node, and edges are drawn from those nodes to the target node. The ICP-algorithm is then repeated for each node in the data, ensuring that every node is the target node once. It is important to note that the direct causes of a target node and its associated residuals must be independent of each other. The following section will describe a solution when this assumption cannot be satisfied.

To illustrate the ICP-algorithm, we use the causal graph that we used in previous sections, shown in the top row in Figure 5.6, and the data that we used to describe the other algorithms. As our target node, we select node 5. We then proceed by identifying all possible causal relations that can exist with our target node at the receiving end of that causal relation. Figure 5.6 shows all possibilities for the target node. In our illustration, the subset enclosed by the black square depicts the true subset; the causal relations that exist in our original causal graph. We will give two examples of subsets to show how the ICP-algorithm works: the subset where no node has a causal relation with the target node (the empty subset), and the subset that contains the nodes that have a causal relation with the target node (the true subset). The residual distribution for the empty subset is shown in the left panel of Figure 5.7. The residual data for the two environments is separated by a dashed vertical line. It is obvious that the residual distributions are not equal, which is also confirmed by the Kolmogorov-Smirnov test (using a significance level of 0.05). Based on this result, we conclude that the empty subset does not hold the optimal subset of nodes that may have a causal relation with the target node.

Let's take a look at the true subset. The residual distribution for the true subset is shown in the right panel of Figure 5.7. The residual data for the two environments is again separated by a dashed vertical line. In contrast to the empty subset, the residual distributions of the true subset do not show a visual difference between the two environments. The Kolmogorov-Smirnov test confirms this (using a significance level of 0.05). Based on this result, we conclude that this subset is "invariant" across environments, and is accepted as the set that may hold the causal relations with the target node. In the sit-

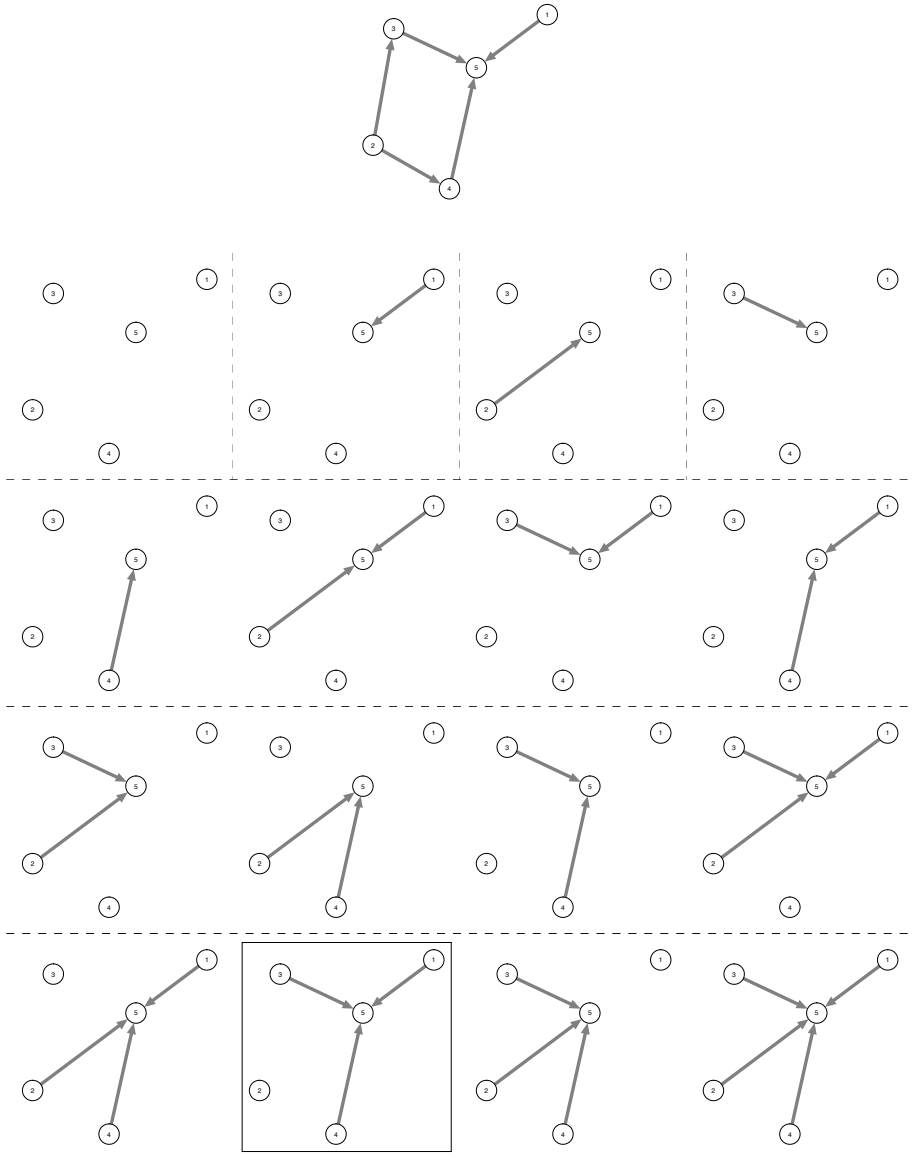


Figure 5.6: The true causal graph (most upper panel), and all possible subsets that may potentially cause the target node 5. The set in the black square indicates the subset that captures the true causal relation with the target node.

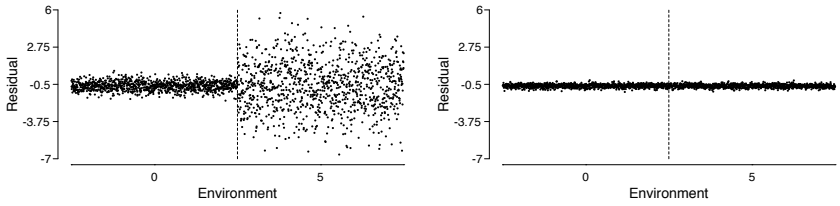


Figure 5.7: Visualisation of the residual distribution for the empty subset (left panel), and the true subset (right panel). Vertical dashed lines indicate the partition of the residuals according to the different environments that we used for this illustration.

uation where more than one subset is accepted, the ICP-algorithm will only select those nodes that appear in all accepted subsets (the intersection of the subsets) and will return that set as the set of nodes that have a causal relation for which the target node is on the receiving end. When we repeat this process for each node in the data, we end up with the original causal graph shown in the top row of Figure 5.6. The ICP-algorithm can return a value for the causal relations, and thus it can return a weighted and signed causal graph. In our simulation study we chose not to vary the strength of the causal relations in the simulated graphs, therefore, we do not use the weights of the causal relations that are returned by the ICP-algorithm.

The data we used to illustrate the ICP-algorithm contains six unique environments: one environment that contains the observational data, and five environments that match the number of nodes in the causal graph. For each target node, we select the environment that holds the observational data, and the environment in which we perturbed all nodes but node 5 (see the section on data simulation for a more detailed description). We only select these two environments because one of the main assumptions of the ICP-algorithm states that perturbations can take place at all nodes but the target node (Peters et al., 2016). We use the *R*-package InvariantCausalPrediction (version 0.7-2, Meinshausen 2018) to run the ICP-algorithm. We programmed a wrapper function (available at https://osf.io/n8gxh/) so that the ICP-algorithm is repeated for every variable in the data,

and then combined into a single adjacency matrix.

Overall, the ICP-algorithm performs exceptionally well in this illustration. All edges that exist in the true causal graph (Figure 5.2) are correctly estimated, and there are no edges that are incorrectly estimated or determined to be absent from the graph.

5.2.5 Hidden Invariant Causal Prediction

The HICP-algorithm (Peters et al., 2017) is similar to the ICP-algorithm discussed previously. The major difference between the two algorithms is that the HICP-algorithm controls for *hidden variables*, variables that are unobserved, but may affect the observed variables. Where the ICP-algorithm assumes that a target node's direct causes and its residuals are independent, this assumption can no longer be satisfied when hidden variables exist. To illustrate, see Figure 5.8, copied from Peters et al. (2017). Here, Y denotes the target node, X its direct cause, H the hidden variable, and Z the instrumental variable. As seen in Figure 5.8, the hidden variable affects both the target node and its direct cause. If one were to use the ICP-algorithm, where hidden variables are not accounted for, the correlation between the target node Y and its direct cause X will be inflated due to the influence of the hidden variable H. It is therefore impossible to infer the unique influence between X and Y in this illustration. Consider the setup in Figure 5.8. For explanatory purposes, the HICP-algorithm implements an instrumental variable Z in order to remove the effect of the hidden variable on $X \longrightarrow Y$. This instrumental variable cannot directly influence the target node Y, as shown in Figure 5.8. Here, the environmental variable ε is used as the instrumental variable, as the division of the data into two separate environments does not directly influence the target node Y. By using the environmental variable ε as the instrumental variable Z, the regression of the target node onto the remaining variables will be split for the different time points, and the difference between these time points is used to estimate the causal effect.

For explanatory purposes, we name the causal effect from X to Y to be α (as shown in Figure 5.8 and as described by Peters et al. 2017). The variables X and Y are defined

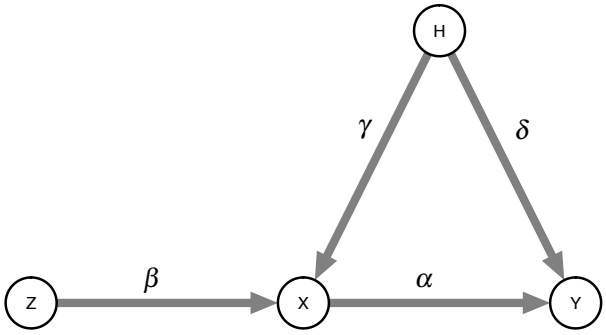


Figure 5.8: Illustration of the HICP-algorithm. Figure is adapted from Peters et al. (2017)

as follows:

$$X = \beta Z + \gamma H + N_x \tag{5.9}$$
$$Y = \alpha X + \delta H + N_y,$$

which follows directly from Figure 5.8. The terms N_x and N_y denote the error terms (here we assume normally distributed variables with mean 0 and variance σ). The estimate of the causal effect from X to Y, $\hat{\alpha}$ is defined as follows:

$$\hat{\alpha} = \frac{\text{cov}[X,Y]}{\text{var}[X]} = \alpha + \frac{\delta\gamma\,\text{var}[H]}{\text{var}[X]} \tag{5.10}$$

where α denotes the causal effect from X to Y, and $\frac{\delta\gamma\,\text{var}[H]}{\text{var}[X]}$ the bias term to account for the hidden variable. When there are no hidden variables, the variance of H is 0, and the bias term will disappear as a result of this. When hidden variables exist but not accounted for, one ends up with a biased estimate for the causal effect from X to Y. This shows that, in this situation, the estimate for the causal effect is not representative of the true causal effect. The HICP-algorithm follows a two step regression to estimate α. It first regresses X on Z to estimate β, where the estimate is denoted by $\hat{\beta}$. Then, this coefficient is used to estimate α:

$$\hat{\alpha} = \frac{\text{cov}[\hat{\beta}Z, Y]}{\beta^2 \text{ var}[Z]} = \frac{\alpha\hat{\beta}^2 \text{ var}[Z]}{\beta^2 \text{ var}[Z]} \tag{5.11}$$

When the sample size becomes large, $\hat{\beta}$ and β will be arbitrarily close. Equation (5.11) shows that, in the limit, the estimate for α ($\hat{\alpha}$) will be equal to the true causal effect. The steps that Peters et al. (2017) took to obtain equation (5.10) are straightforward, see Appendix D.3 for a more detailed description. It is important to note that the HICP-algorithm assumes that the hidden variable H and the instrumental variable Z (the environmental variable ε) are independent of one another. The ICP-algorithm has to satisfy the assumption that the causes of a target node and its associated residuals are uncorrelated. In contrast, the HICP-algorithm frees up this assumption. Another difference between the HICP and the ICP-algorithm is that the HICP-algorithm does not create subsets of the set of nodes that remain after selecting a target node. To speed up computations, all variables are simultaneously tested to see if they are a cause of the target node.

We use the data that we used earlier to illustrate the HICP-algorithm. Figure 5.6 shows the correct subset of direct causes when we take node 5 to be the target node. The causal coefficients α are estimated as follows. Let X be a $n \times (p-1)$ matrix that contains the raw data (both observational and experimental data) for all variables but the target node:

$$\hat{\alpha} = (X'X)^{-1}X'y = \begin{pmatrix} -8.204 & 0.171 & 0.194 & 0.866 \\ 0.171 & -4.056 & -5.138 & -5.955 \\ 0.194 & -5.138 & -12.874 & -7.624 \\ 0.866 & -5.955 & -7.624 & -16.888 \end{pmatrix}^{-1} \begin{pmatrix} -7.124 \\ -10.961 \\ -20.295 \\ -23.772 \end{pmatrix} \tag{5.12}$$

$$\hat{\alpha} = \begin{pmatrix} 0.998 \\ 0.009 \\ 0.991 \\ 1.009 \end{pmatrix} \tag{5.13}$$

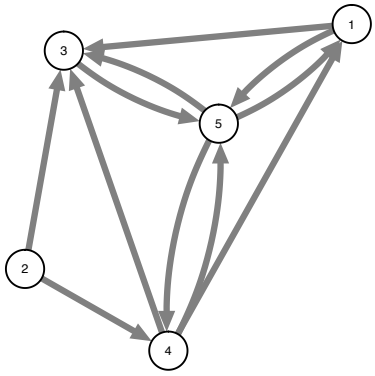


Figure 5.9: Illustration of the HICP-algorithm.

One can immediately see that nodes 1, 3 and 4 have a causal coefficient that is very high, and that node 2 has a almost non-existent causal coefficient. The causal coefficients in α are then tested for significance. When we repeat the HICP-algorithm for each node in our illustration data (that does not contain any hidden variables), we end up with the causal graph depicted in Figure 5.9. It is noticeable that, next to the edges that are present in the true causal graph (shown in Figure 5.2), many spurious edges exist. Since the HICP-algorithm tests all variables in the data simultaneously, spurious edges can arise as a result of partialling out the effect of the hidden variables. With a large sample size (like the sample size of the illustration data), these spurious edges are easier deemed as significant causal relations.

We have added a detailed description of the entire estimation part as it occurs in the *R*-package in the supplementary materials for the interested reader in Appendix D.3. We use the *R*-package InvariantCausalPrediction (version 0.7-2, Meinshausen 2018) to run the HICP-algorithm. We programmed a wrapper function (available at https://osf.io/n8gxh/) so that the HICP-algorithm is repeated for every variable in the data, and then combined into a single adjacency matrix.

Overall, the HICP-algorithm does not seem to perform too well in this illustration.

All edges that exist in the true causal graph (Figure 5.2) are correctly estimated, but there are many edges (6) that are incorrectly estimated.

Algorithm	Observational data	Experimental data	$N = 1$	$N > 1$	Within-subjects	Between-subjects	Limitations	Sensitivity	Specificity
PC	$\surd$	$-$	$-$	$\surd$	$\surd$	$\surd$	Uses only observational data	Partially high	High
DR-FFL	$\surd$	$\surd$	$\surd$	$-$	$\surd$	$-$	Resulting graph is unweighted and unsigned	Low	High
TRANSWESD	$\surd$	$\surd$	$-$	$\surd$	$\surd$	$-$	Employs arbitrary threshold	Low	High
ICP	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	Computationally slow with many variables	Partially high	High
HICP	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	Contains spurious relations	High	Partially high

Table 5.1: Overview of the algorithms. $\surd =$ algorithm can handle that specific property. $- =$ algorithm cannot handle that specific property. Partially high means that the sensitivity/specificity is high depending on certain conditions.

5.3 Data simulation

In order to study the accuracy of the algorithms that we described in the previous section, we simulate data according to a DAG, and apply each of the five algorithms to estimate a causal graph. Based on the adjacency matrix of the simulated DAG (see Appendix D.4 for more detail on how we created DAGs) that we call B, we simulate data in which some sort of perturbation has taken place. The rows in B reflect the nodes where the causal relations are outgoing, and the columns the nodes where the causal relations are incoming. We start out by creating a $n \times p$ matrix, called X that we fill with numbers drawn from a normal distribution with a mean of 0 and a standard deviation of 0.5. Here, p reflects the number of nodes in the graph, and n the number of participants or observations for which we want to simulate data. We then select a node (called the target node), and we create n error terms called e, all drawn from a normal distribution with a mean of 0 and a standard deviation of 0.5. We then create observational data in a following manner:

$$y = bX_b + e \tag{5.14}$$

where b denotes the row in B of length $1 \times p$ that corresponds with the target node, and X_b the column in X corresponding with the target node. The result is a $n \times 1$ vector y that holds the observational data. We repeat this process for all nodes in the DAG, and combine them in a $n \times p$ matrix Y. The next step is to create perturbation data. For each node independently, we select all data from our original matrix X, excluding the target node. This $n \times (p-1)$ matrix is then multiplied with $p-1$ values (called a) drawn from a normal distribution with a predetermined mean ($\overline{m}$) and standard deviation (sd), creating a new matrix X_{per}. We then create experimental data similar to the previous step:

$$y_{per} = b(aX_b) + e \tag{5.15}$$

We repeat this process for all nodes in the DAG, and combine them in a $n \times p$ matrix Y_{per}.

We also simulated data in which we added hidden variables. The equations for the simulation of data with hidden variables are similar to the regular data simulation (shown in (5.14) and (5.15)) with the following addition. We create hidden variables by drawing n values from a normal distribution (with a mean of 1 and a standard deviation of 1). These n values are then multiplied by a parameter h that we have set to be 5. We then take the outer product of h and a vector of 1s of length p, and add this to our matrix X. This matrix is then used in a similar manner as described above.

5.4 Numerical evaluation of causal inference algorithms

In this simulation study we evaluated the performance of five methods of causal inference, namely the PC, DR-FFL, TRANSWESD, ICP and the HICP-algorithm. We simulated six DAGs in total: three with $p = 5$ nodes, and three with $p = 10$ nodes. We varied the density of the graphs (the proportion of edges present in the graph) $d \in \{0.1, 0.25, 0.5\}$. Figure 5.10 depicts all causal graphs that were created for this simulation study. We varied the number of participants $n \in \{50, 100, 500, 1000, 5000\}$ and the mean of the perturbation distribution $\overline{m} \in \{1, 5\}$. These values correspond to a small and large perturbation effect. The standard deviation of the perturbation distribution was also varied $sd \in \{0.5, 5\}$. These values correspond to an effective and a noisy perturbation. For the DR-FFL and the TRANSWESD algorithms, we varied $\beta \in \{0.5, 1, 1.64, 1.96, 2.58\}$, and γ was set to 0. We simulated data with and without hidden variables to see how the addition of hidden variables affected the performance of the algorithms. For each combination of parameters, we ran the simulation 100. We set the significance level for the PC, ICP and HICP-algorithm to be 0.05. All simulated data, as well as the used R-code are publicly available at https://osf.io/n8gxh/.

In the numerical evaluation we focus on *Matthew's correlation coefficient* (MCC; Powers, 2011). The MCC takes both true and false positives and negatives into account, and therefore gives a good overview of the overall performance of the different algorithms. The MCC is calculated as follows:

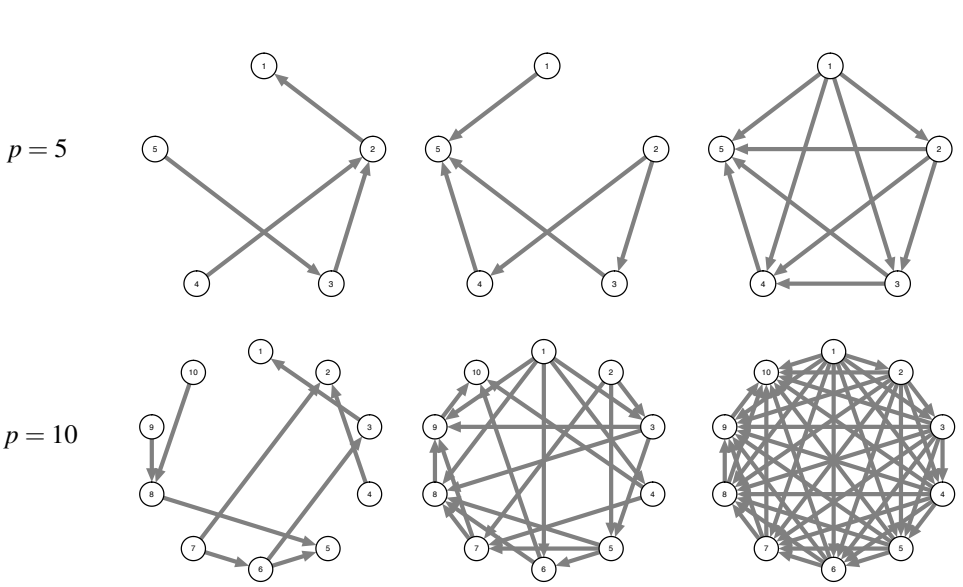


Figure 5.10: DAGs that were used to simulate data. p = number of nodes in the graph. d = the percentage of edges present in the graph.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{5.16}$$

where TP represents the number of true positives, TN the number of true negatives, FP the number of false positives and FN the number of false negatives. The MCC can be interpreted the same way as a regular correlation coefficient (Matthews, 1975). The more positive the MCC, the better the correspondence between simulated and estimated edges. We also calculated other metrics (e.g., positive/negative predictive rate, false negative/-positive rate), but we chose not to present these here. Results for all metrics can be found online at https://osf.io/n8gxh/.

Figure 5.11 shows the MCC for the different algorithms. Overall, the ICP and HICP-algorithm have the highest MCC. The MCC of the PC-algorithm is generally low. For clarity of presentation, we only show results for $p = 10$ nodes, with a graph density $d = 0.25$. All other results can be found online. The PC-algorithm seems to benefit from

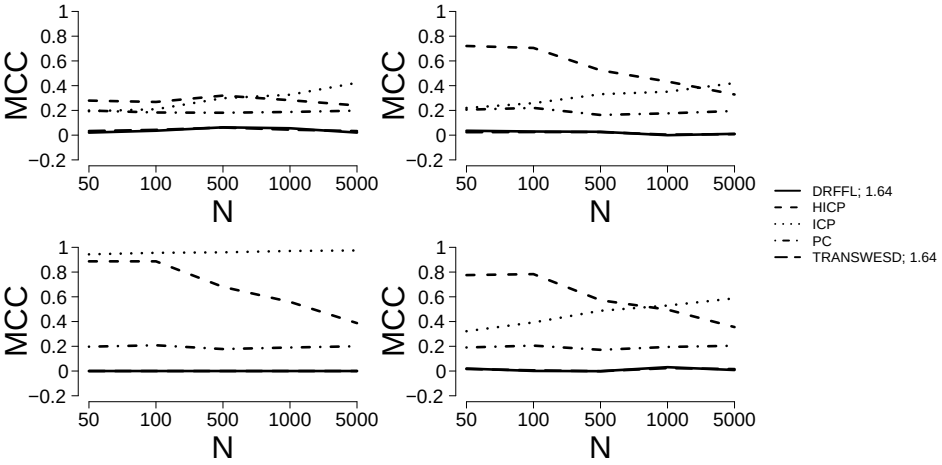


Figure 5.11: Matthew's correlation coefficient (MCC) for $p = 10$ nodes with a network density of $d = 0.25$ when no hidden variables were simulated. Top left = $\overline{m} = 1, sd = 0.5$, top right = $\overline{m} = 1, sd = 5$, bottom left = $\overline{m} = 5, sd = 0.5$, bottom right = $\overline{m} = 5, sd = 5$.

a density that is not too high. The MCC increases when the graph density d increases from 0.1 to 0.25, but decreases again when d is increased to 0.5. Also, the size of the graph (reflected by p) has an impact on the MCC: when p is increased from 5 to 10, the MCC decreases from on average 0.55 to 0.23. The PC-algorithm can have issues determining the direction of directed edges. In around 18.75% of all the simulations, the PC-algorithm returned an undirected graph. It turns out that the MCC is almost always higher when we do not take the direction of the edges into account. This effect is especially present when the density of the graphs is low. To illustrate, when $d = 0.1$, the average MCC increases from 0.47 to 0.87 when we do not take the direction of the edges into account, but when $d = 0.5$, the difference in average MCC is only 0.05 on average. These results indicate that the PC-algorithm can be useful in determining the pairs of variables between which a causal relation exists, but that it may not be the appropriate algorithm to determine the direction of these causal relations.

In general, both the DR-FFL and the TRANSWESD algorithm perform badly, as seen in Figure 5.11. Both the mean of the perturbation distribution ($\overline{m}$) and its standard

deviation (sd) do not have an effect on their performances. When $p = 5$, both algorithms seem to perform a little better, but the difference in MCC is only about 0.1 between the two network sizes. The root cause of this seems to lie in the first step of the transitive reduction scheme that the two algorithms apply. Where on average each simulated graph had about 13.45 edges, both the DR-FFL and the TRANSWESD algorithm returned on average 2.95 edges. When there are only a few edges present in the graph, transitive reduction works sub optimally, as even fewer edges can be removed from the graph. Our findings indicate that transitive reduction may not be the best way to go when estimating a causal graph. We tried a different approach in the data simulation, but we found similar results. We also simulated data with an extremely large perturbation mean ($\overline{m} = 100$), but the MCC hardly improved.

The ICP-algorithm seems to do a better job at correctly estimating the causal graphs in some cases. Overall, the ICP-algorithm works best at $d = 0.25$. The MCC is much lower when the network density is lower (or higher). In general, when $p = 5$, the ICP-algorithm performs best (at $d = 0.25$). In that situation, the mean and standard deviation of the perturbation distribution do not have an effect on the MCC. In contrast, when $p = 10$, the influence of these two parameters is much bigger. As shown in Figure 5.11, only when $\overline{m} = 5$ and $sd = 0.5$ is the MCC high; in all other cases it is mediocre at best. The ICP-algorithm can be conservative: of all the edges that it finds, the algorithm will only take the intersection as the (sub)set of causes of a target node. The results shown here suggest that the ICP-algorithm can estimate causal graphs pretty accurately when there are not too few or too many edges in the graph, and when the perturbation effect is strong and precise enough.

Lastly, the HICP-algorithm displays a mixed performance. When $p = 5$, we see a similar pattern as with the ICP-algorithm, albeit it less extreme. With a graph density $d = 0.25$, the HICP-algorithm has a high MCC, but for the other two graph densities, the MCC is smaller. The declining effect that we observe in Figure 5.11 returns for other graph sizes and densities as well: when the sample size n increases, the MCC

decreases. The cause of the decrease can be found in the number of false positives. Like we saw in the illustration in the previous section, the causal graph that is the result of the HICP-algorithm often contains spurious edges. This is likely due to the fact that the HICP-algorithm only investigates the entire set of remaining nodes in the graph to determine the causes of the target node, in contrast to the ICP-algorithm that investigates each possible subset separately. Because of this set-up, spurious edges can occur, which become significant more easily with a larger sample size. This indicates that the HICP-algorithm does not need a large sample to estimate a causal graph.

Figure 5.12 gives a more detailed insight into the strengths and weaknesses of each algorithm. To increase readability of the causal graphs, we chose to display the results for $p = 5$ and $d = 0.25$. These graphs are publicly available for all simulation conditions. The number of false positives (red edges) for the DR-FFL, TRANSWESD and HICP-algorithm immediately stand out. The HICP-algorithm stands out from the DR-FFL and the TRANSWESD algorithm because it also has a high number of true positives, indicated by the thickness and saturation of the blue edges. The conservativeness of the ICP-algorithm is less visible, but present nonetheless. Where the HICP and the PC-algorithm correctly identified the present edges (as shown in the true graph) in either $99-100\%$ of the simulations, the ICP-algorithm's rate lies around $90-96\%$, which is still very high. The PC-algorithm's struggle with determining the direction of the edges is also depicted in Figure 5.12. The direction of two edges here (e_{42} and e_{32}) are just as often correctly as incorrectly identified.

We also ran the simulation study using data that contained hidden variables. The results are very similar to the results using data without hidden variables. See Appendix D.5 for the results using data with hidden variables.

5.5 Discussion

The present study compared five different algorithms that are used for causal inference. We provided a simulation study in which we showed how well each algorithm is able to

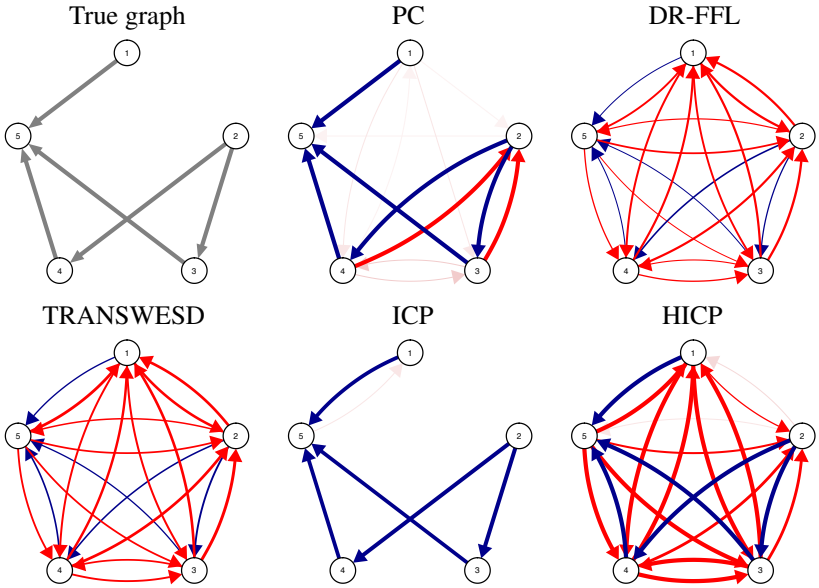


Figure 5.12: Visualisation of the number of true positives and false positives for $p = 5$, $d = 0.25$, $n = 5000$, $\overline{m} = 5$, $sd = 0.5$ and $\beta = 0.5$ without the addition of hidden variables. Blue edges indicate true positives, and red edges indicate false negatives. The saturation and thickness of the edge represents how often that edge was (in)correctly estimated. Upper left = true graph, upper middle = PC, upper right = DR-FFL, lower left = TRANSWESD, lower middle = ICP, lower right = HICP.

estimate the causal graph under which the data were simulated. We simulated data from causal graphs with different properties to assess the effect of the number of nodes and the density of the graph on the estimation of the graph itself. The results that we showed did not present us with a clear winner: only under specific circumstances did each algorithm perform well. The exception to this are the DR-FFL and TRANSWESD algorithm that never showed a good performance, contrary to earlier studies on these algorithms (Pinna et al., 2013).

Every algorithm that is discussed here had its own advantages and disadvantages. The PC-algorithm only uses observational data to estimate the causal relations between

variables. Although this implies that less data is needed to run the algorithm, we did see that the MCC of the PC-algorithm is never higher than 0.8. Furthermore, we found that the PC-algorithm often has the location of a causal relation correctly identified, but not its direction. This may suggest that the PC-algorithm is useful to find the location of the causal relations, but that some other algorithm is needed to identify the direction of that relation.

It is quite noticeable that the DR-FFL and the TRANSWESD algorithm do not perform as well as the others do. In nearly 75% of all simulations 10% or less of the number of possible edges survived the first step of these algorithms. This is why the MCC for both the DR-FFL and the TRANSWESD algorithm is so low across simulation conditions. We performed the same simulation study, but in contrast to the one described here, we only select 1 node (instead of $p-1$) that we perturbed and used to create our experimental data. However, results from this design are similar to the results presented here. It is possible that a different data simulation design might give different results, but this remains to be investigated.

The ICP-algorithm has proven to be able to correctly identify edges that are present in the true graphs. It also often makes the correct decision when edges are absent in the true graphs. However, the ICP-algorithm will only perform well when specific conditions are satisfied with respect to the data. For instance, the high MCC that we found in this study was only obtained when the mean of the perturbation distribution was high, and its associated standard deviation small. In all other conditions, the MCC was mediocre and in some cases poor. This suggests that the ICP-algorithm is a conservative one. A conservative attitude is not necessarily a disadvantage of an algorithm, but when it is applied to empirical data, the resulting graph may be sparser than one may hope for. Also, the ICP-algorithm investigates every possible subset of the variables that remain after selecting a target variable. This step leads to computational issues when graphs with a large number of nodes are studied. When a graph has $p = 5$ nodes, the number of subsets per target variable is 16. When a graph has $p = 10$ nodes, the number of subsets

grows to 516, and for $p = 15$ nodes, the number of subsets equals 16384. In the future we hope to develop an adaptation for the ICP-algorithm where a subset selection is made in such a way that the computation time decreases while maintaining similar specificity and sensitivity values.

Lastly, the HICP-algorithm outperforms the other algorithms in terms of the MCC in many simulation conditions. However, as is clearly shown in Figures 5.12 and D.4, many edges are incorrectly seen as present (false positives; red edges) next to the correctly identified edges (true positives; blue edges). This phenomena most likely occurs because not every possible subset is investigated separately, as is the case in the ICP-algorithm.

Next to the disadvantages that we discussed previously, we made some arbitrary decisions for the algorithms in this study. The PC, ICP and HICP-algorithm all require a significance level that we set to be 0.05. The DR-FFL and the TRANSWESD algorithm have one or two threshold parameters that need to be set prior to the analysis. We chose to use different values to evaluate the effect of these parameters. Results of our simulation study showed that the value of these thresholds impact the MCC of the algorithm: the higher the threshold, the less edges are returned after the first step of these two algorithms, and therefore the less edges can be reduced from the graph. Choosing a value for these threshold parameters is thus not trivial. Ideally, one would want to set these parameters in such a way that enough edges are returned after the first step of the algorithms, but not too many that none of them can be removed due to cycles, for example. It remains that, with the DR-FFL and the TRANSWESD algorithm, setting the threshold parameter(s) is no trivial matter and confounds the results tremendously. Future research could look into the possibility of using maximum likelihood estimation to obtain a reasonable threshold parameter based on the data.

We simulated data without and with the addition of hidden variables. Although they are present, it is hard to find differences between the results for data without, and data with hidden variables. We followed the steps shown in the description of the HICP function of the associated *R*-package. Due to the high number of simulation conditions that we

already have, we chose not to add any by varying the strength of the hidden variable. It is possible that a hidden variable with a stronger effect may result in larger differences in specificity and sensitivity. In a future extension of this study, one can vary the hidden variable to investigate the effect of a hidden variable on the results.

As every algorithm has its own advantages and disadvantages, a possible combination of two or more algorithms may be the solution. For instance, one can use the PC-algorithm to determine the skeleton of a causal graph, and use that input for the ICP-algorithm. In this combination, the number of subsets decreases substantially. Another option would be to copy the subset design of the ICP-algorithm, and use in with the HICP-algorithm. Investigating multiple subsets may result in a lower number of false positives, and thus a more accurate depiction of the true causal graph.

To our knowledge, this is the first study that compared different algorithms for causal inference based on experimental data. Based on the simulation results, we gain more insight into the accuracy of each algorithm, and how suited they are for empirical (psychological) data. The ICP and HICP-algorithm are the top candidates to be used in psychological research. As hidden variables are a common problem in psychological research, a possible combination of the ICP and HICP algorithm may be the best plan of attack to estimate causal relations between psychological variables.

6

Introducing the Causal Graph Approach to Psychopathology: An Illustration in Patients with Obsessive-Compulsive Disorder

This chapter is submitted as: Kossakowski, J.J., van Oudheusden, L. J. B., McNally, R. J., Waldorp, L. J., Riemann, B. C., & van der Maas, H. L. J. (2020). Introducing the Causal Graph Approach to Psychopathology: An Illustration in Patients with Obsessive-Compulsive Disorder. *Under review at Clinical Psychological Science*.

Abstract

Clinicians aim to discern causal relations among features of mental disorders. Yet observational data are insufficient to confirm these relations. In the present study, we describe a causal graph approach to illustrate how one can discern causal relations among aspects of obsessive-compulsive disorder (OCD) by applying the Invariant Causal Prediction (ICP) algorithm and the Hidden Invariant Causal Prediction (HICP) algorithm, supplemented by findings from a literature review. Applying these methods to symptom severity data from four time points (before, during [twice], and after behavioural treatment) we estimated a causal graph for OCD. The resulting causal graph reveals multiple cycles between aspects of OCD that may play a role in the maintenance of the disorder. We conclude that the proposed design has great potential and could benefit clinicians with some adjustments and extensions.

6.1 Introduction

The network perspective is an emerging view that postulates that mental disorders form a system of interactive symptoms. Instead of a latent variable – the "underlying disorder" – causing symptoms that presumably reflect its presence, symptoms and their interactions constitute the disorder, rendering the latent variable entirely redundant (Borsboom, 2017; Borsboom & Cramer, 2013). Figure 6.1 illustrates a causal graph: a network where the symptoms (here X, Y and Z.) are depicted as nodes (the circles), and the causal relations between symptoms are represented by an arrow (directed edge).

Clinical researchers aim to identify causal relations among symptoms. Unfortunately, their aspirations are often stymied by mere correlational data that can suggest, but never confirm, causal relations. Indeed, experimental manipulations designed to underwrite causal conclusions are frequently infeasible in clinical settings. However, new methods have emerged that enable such inferences (Kossakowski et al., 2020). The purpose of our chapter is to introduce these methods by applying them to patients with obsessive-compulsive disorder (OCD).

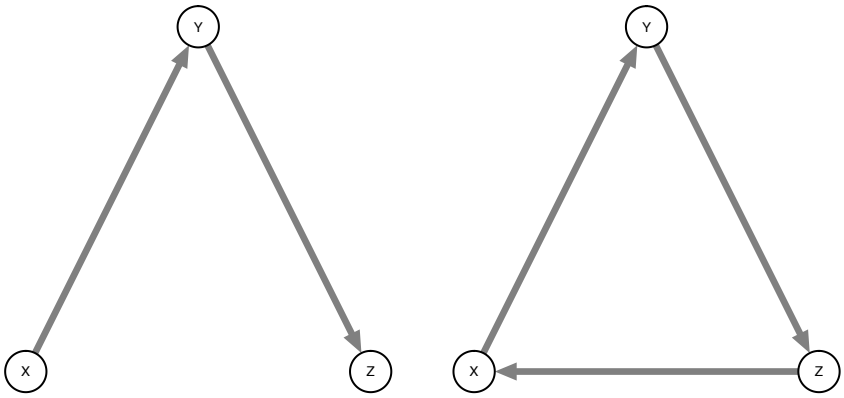


Figure 6.1: Illustration of causal relations among three variables X, Y, and Z. The left panel depicts a Directed Acyclic Graph (DAG), whereas the right panel shows a Directed Cyclic Graph.

In this study, we define a *causal relation* as a relation between two variables $X \longrightarrow Y$ where, when we manipulate (or intervene on) variable X, we observe a change in variable Y (conditioned on all other variables; Kossakowski et al., 2020). This definition also means that, if we do not observe any changes in Y, no intervention has taken place on X. When it comes to the treatment of patients, it is important to know whether symptom X causes symptom Y, or perhaps vice versa. Figure 6.1 (left panel) illustrates a simple example where we have three variables, $X \longrightarrow Y \longrightarrow Z$. If the causal relations among these three variables are unknown, a therapist may first therapeutically target variable Z. However, as shown in this illustration, it may be more effective to target variables X and Y, as both of these have an (in)direct effect on variable Z.

Many studies use *observational* data to estimate causal relations between variables. Observational data are data on which no interventions have taken place. However, observational data alone do not allow one to estimate causal relations accurately, one needs to manipulate variables to estimate true causal relations (Holland, 1986). Thus, we need interventional data upon which some manipulation or intervention has taken place. Inter-

vention here indicates that some manipulation is performed compared to the observational data. Examples include starting treatment or reducing one's medication. By using both observational and interventional data, we could potentially estimate true causal relations between symptoms (Meinshausen et al., 2016).

In this study, we describe a causal graph approach in which we combine observational and interventional data to estimate a causal graph. This approach is illustrated with an application to data of patients diagnosed with OCD who were measured at four time points: before, twice during, and after treatment. OCD is characterised by intrusive, persistent thoughts, urges or images (obsessions), and repetitive behaviours or mental acts (compulsions) performed to relieve the distress that accompanies the obsessions (American Psychiatric Association [APA], 2013). US national and global life-time prevalence rates for OCD range from between 1.6% to 2.3% (e.g., Kessler et al., 2005; Fontenelle et al., 2006; Ruscio et al., 2010). Two network studies on OCD estimated Directed Acyclic Graphs (DAGs) involving features of OCD and symptoms of depression (McNally et al., 2017; Jones et al., 2018). A DAG illustrates potential predictive causal relations among symptoms that together depict a causal system (Pearl & Mackenzie, 2018), but only if two stringent assumptions are met. The causal graph must contain all important variables, and there cannot be any cycles among the nodes. A cycle is shown in Figure 6.1 (right panel), where variable Z is the cause of variable X, thereby creating a cycle.

Ideally, to estimate causal relations, we would want to manipulate one symptom at a time, and observe its effects on other symptoms. This approach would mimic an approach in gene biology whereby researchers "knock-out" or silence a specific gene and observe the subsequent effects (Serra et al., 2004; Pinna et al., 2010). Using a "slim finger", geneticists can deactivate a gene while leaving all other genes untouched. Unfortunately, clinical psychologists are saddled with "fat fingers" that render it difficult to deactivate a single symptom without simultaneously affecting other symptoms. For example, it would be nearly impossible to administer a benzodiazepine for insomnia without also reducing anxiety at the same time. Ironically, targeting one symptom (e.g., insomnia) will not

only affect other symptoms in serial fashion (e.g., less fatigue the following day), but will also produce effects on additional symptoms in parallel fashion (e.g., anxiety), thereby complicating causal analysis. Indeed, algorithms designed for gene knock-out experiments may not work as accurately when applied to psychological data (Kossakowski et al., 2020).

In the causal graph approach, we applied the *Invariant Causal Prediction* (ICP) algorithm (Meinshausen et al., 2016) and the *Hidden Invariant Causal Prediction* (HICP) algorithm (Peters et al., 2016, 2017). These algorithms represent a significant advance over DAGs as the daunting assumptions of the latter are difficult to meet in clinical science. The ICP-algorithm can handle interventional data where multiple symptoms are simultaneously manipulated. The HICP-algorithm complements the ICP-algorithm because it can accommodate *hidden variables* – unmeasured variables that may affect the measured variables nonetheless (Peters et al., 2016, 2017). Both algorithms have shown great potential for psychological applications (Kossakowski et al., 2020). Lastly, following Meinshausen et al.'s (2016) approach, we also searched the literature for evidence of causal relations between features of OCD. We combined results from the ICP-algorithm, the HICP-algorithm and the literature review into a single causal graph that may give an overview of the causal structure of OCD.

6.2 Methods

6.2.1 Participants

We used archival clinical data from patients with a primary OCD diagnosis. A total of 3474 patients entered the study at one of the ten locations of the Rogers Behavioral Health centers across the United States treated between June 2012 and June 2018. Patients received treatment either in an intensive outpatient ($n = 1320$), partial hospitalisation ($n = 1247$), or in a residential unit ($n = 906$). The treatment setting for one patient was unknown. The primary intervention was exposure and response prevention therapy.

We included only those patients who completed the relevant questionnaires prior to treatment, at the first and second progress assessments, and the post-treatment assessment ($t = 4$ time points). We ended up with 903 patients that were included for all analyses. Patients gave their written informed consent to use de-identified data for research purposes. Of this sample, 282 patients received treatment in an intensive outpatient unit, 371 patients received treatment in a partial hospitalisation unit, and 250 patients received treatment in a residential unit. One patient switched from an intensive outpatient unit to a partial hospitalisation unit after the initial measurement. The average number of days between admission and discharge was 56.39 ($SD = 20.29$ days). The average time between measurements ranged from 15.81 to 23.25 days, with a standard deviation ranging from 7.43 to 17.67 days. The sample consisted of 478 (52.93%) females and 424 (46.95%) males, whose age ranged from 18 to 76 ($M = 29.69$, $SD = 11.42$). The gender of one participant was unknown. Based on the recommendations presented by Mataix-Cols et al. (2016), who argue that a treatment response is shown by a reduction in Y-BOCS score by at least 35% compared to their score at baseline, 538 patients in our sample experienced a clinically meaningful reduction in symptoms.

6.2.2 Yale-Brown Obsessive-Compulsive Scale Self-Report

Patients completed a test battery that contained, among other questionnaires, the Yale-Brown Obsessive-Compulsive Scale Self-Report (Y-BOCS-SR; Steketee et al., 1996), the self-report version of the Y-BOCS (Goodman, Price, Rasmussen, Mazure, Fleischmann, et al., 1989). Patients completed the Y-BOCS-SR at four time points before, during, and after treatment. The Y-BOCS-SR is a 10-item questionnaire that evaluates features of OCD without taking the content of a patient's obsessions and compulsions into account (e.g., harming versus contamination obsessions; checking versus washing compulsions). The Y-BOCS-SR has a symmetric structure: the questionnaire evaluates five aspects of obsessions and of compulsions, respectively. Table 6.1 gives a more specific overview of the individual items, including the means and standard deviations of each item per

measurement. All items were measured on a five-point Likert scale, ranging from 0 (no symptoms) to 4 (extreme). The Y-BOCS-SR shows excellent psychometric properties (Goodman et al., 1989, Goodman et al., 1989, Steketee et al., 1996).

6.3 Causal Graph Approach

The proposed causal graph approach combines evidence for specific directed edges from three different sources: the ICP-algorithm (Meinshausen et al., 2016), the HICP-algorithm (Peters et al., 2017), and a literature study.

6.3.1 Invariant Causal Prediction

The ICP-algorithm (Meinshausen et al., 2016; Peters et al., 2016) is designed to handle the combination of observational and interventional data. The basic idea is that the conditional distribution of a single node (called the *target node*), given its direct causes, does not change if some manipulation takes place on variables other than the direct causes of a target node. Hence, for each node, the ICP-algorithm ascertains the subset of remaining nodes that cause the target node. The algorithm starts by selecting one node, which is the *target node*. The target node is then regressed on each possible subset of remaining nodes, ranging from an empty subset, to all remaining nodes in the graph. For each of these regressions, we obtain the residual distribution. This is then split according to an instrumental variable. This variable indicates which score (i.e., residual) belongs to which measurement time point. The algorithm then compares the residual distributions for every pair of measurements (4 in our illustration: before, twice during, and after treatment). If the residual distributions of a pair of measurements are deemed equal, we conclude that the causal relation between the subset of nodes and the target node is "invariant." After all subsets are examined, we draw directed edges from the intersection of all subsets to the target node. If, for a given subset of nodes, the residuals are similar across all measurements, then conditional on the direct causes of the target node, any manipulation will not change the distribution of the residuals of the target node when one conditions on the

Item	Item label	Mean (SD) Baseline	Mean (SD) Progress1	Mean (SD) Progress2	Mean (SD) Post
Time occupied by obsessive thoughts	O-Time	2.66 (1.15)	2.34 (1.04)	2.16 (1.05)	1.76 (0.98)
Interference due to obsessive thoughts	O-Int	2.39 (1.06)	1.99 (0.93)	1.77 (0.95)	1.41 (0.89)
Distress associated with obsessive thoughts	O-Distr	2.55 (1.02)	2.12 (0.89)	1.97 (0.91)	1.64 (0.87)
Resistance against obsessions	O-Res	1.72 (0.99)	1.42 (0.85)	1.27 (0.85)	1.01 (0.88)
Degree of control over obsessive thoughts	O-Cont	2.32 (0.96)	1.91 (0.83)	1.74 (0.81)	1.42 (0.79)
Time spent performing compulsive behaviours	C-Time	2.17 (1.16)	1.81 (1.03)	1.67 (0.99)	1.41 (0.95)
Interference due to compulsive behaviours	C-Int	2.11 (1.16)	1.74 (1.04)	1.54 (0.97)	1.28 (0.95)
Distress associated with compulsive behaviour	C-Distr	2.36 (1.06)	2.05 (0.93)	1.91 (0.93)	1.56 (0.90)
Resistance against compulsions	C-Res	1.89 (1.06)	1.40 (0.84)	1.23 (0.85)	0.95 (0.82)
Degree of control over compulsive behaviour	C-Cont	2.16 (1.00)	1.74 (0.86)	1.56 (0.83)	1.28 (0.79)

Table 6.1: The Yale-Brown Obsessive Compulsive Scale Self-Report (Y-BOCS-SR) with their assigned item labels and mean (SD) scores per time point.

direct causes. This process is then repeated for all nodes in the graph. For a more detailed description of how the ICP-algorithm performs in a simulation study, see Kossakowski et al. (2020) and Meinshausen et al. (2016).

6.3.2 Hidden Invariant Causal Prediction

When we run the ICP-algorithm, we assume that we have measured all possible causes of every single node. However, it is very difficult to satisfy this assumption, especially in psychopathology. Many aspects or events can affect features of OCD that are not captured by the Y-BOCS-SR. For example, research has shown that the catastrophic misinterpretation of obsessive thoughts (Rachman, 1997) can influence OCD symptoms. Also, OCD tends to co-occur with other mental disorders such as anxiety or mood disorders (see Ruzzano et al., 2015; McNally et al., 2017; Jones et al., 2018). Therefore, we ran the HICP-algorithm to account for the influence of *hidden variables*: variables that are not measured, but that may affect those we did measure (Meinshausen et al., 2016; Peters et al., 2017).

A hidden variable that affects both the cause and the target node inflates the relation between the two variables if the hidden variable is not accounted for. When hidden variables affect variables of interest, the ICP-algorithm can no longer estimate the causal relation between two variables accurately. Accordingly, to estimate the true causal relation between two variables, the HICP-algorithm employs an *instrumental variable*. An instrumental variable influences only the cause, but not the target node. It is the same variable that we used in the ICP-algorithm to split the data according to the different time points in the dataset. An instrumental variable splits the data into four parts that correspond to the four assessments that we have for our data. By comparing every pair of time points, the HICP-algorithm partials out the effect of the hidden variable. For more information on the HICP-algorithm and how it performs in a simulation study, see Kossakowski et al. (2020).

6.3.3 Literature Search

We aimed to complete the causal graph that results from the ICP and HICP-algorithm by means of a literature study. Using Google Scholar and Web of Science, we searched for articles on OCD. Within these studies, we looked for sentences that signify a causal relation between specific aspects of OCD. For example, Marker et al. (2006, p. 390) state that "Obsessive-compulsive disorder (OCD) is a common anxiety disorder characterised by intrusive thoughts that are difficult to dismiss and that increase anxiety." A sentence such as this expresses a causal relation from *O-Time* (time occupied by obsessive thoughts) to the subsequent *O-Distr* (distress associated with obsessive thoughts). An independent researcher unaware of our results double-checked our findings as a safeguard against any possible subjectivity and confirmation bias.

6.4 Results

6.4.1 Literature Study

The two researchers – the first two authors of this paper – found a total of 74 causal relations in 15 articles and two books. Researcher R1, who also searched for the literature, found 53 causal relations, and researcher R2 56. Of these causal relations, initially 21 were found by both researchers in the same materials. The initial lack of overlap between the two researchers was due to a different interpretation of the resistance items (*O-Res* and *C-Res*). After deliberation between the two researchers, the overlap increased to 35 causal relations, of which nine represent unique causal relations. We decided to include only those causal relations found by both researchers. Table 6.2 gives an overview of all the edges that are in the causal graph with the corresponding source.

6.4.2 Causal Graph

Figure 6.2 presents the causal graph, in which both show the split results from the algorithms and the literature review (Figures 6.2a and 6.2b), and a causal graph (Figure

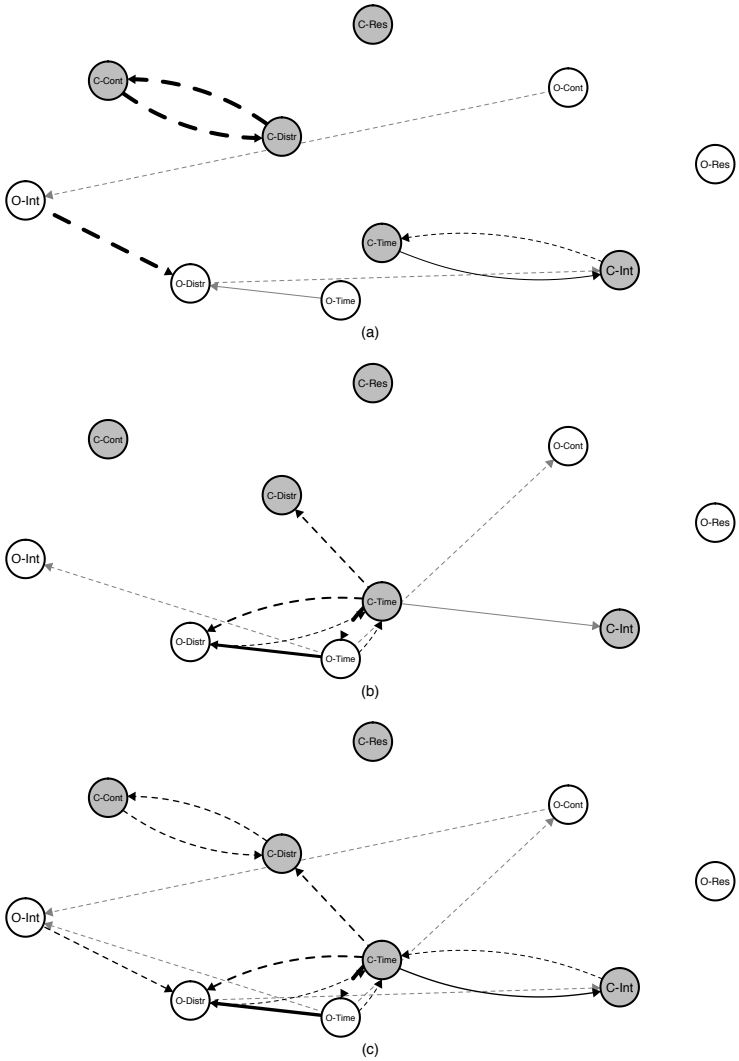


Figure 6.2: Causal graph of the Y-BOCS-SR items, detected by the ICP and HICP algorithms (a), found in the literature (b), and combined (c). Item labels are described in Table 6.1. White nodes signify aspects of obsession, whereas grey nodes signify aspects of compulsions. Solid edges are those detected by either of two algorithms *and* in the literature. Dashed edges are those found with either of two algorithms, or in the literature. The edge thickness and colour saturation shows how often an edge is detected, the thicker and more saturated an edge, the more often it was detected.

6.2c), where we combined the results from the ICP, the HICP-algorithm, and the literature review. Table 6.2 gives an complete overview of the sources justifying inclusion of edges in Figure 6.2. The first characteristic that stands out in this causal graph is the lack of overlap between the different sources. Out of the 14 edges that were detected, three edges were detected by both the ICP-algorithm and the HICP-algorithm, and only two edges were detected by either an algorithm and the literature review. At the same time, many edges appeared at least two studies or sources; only five edges had only one source (algorithm or literature).

Edge	Algorithm	Literature
O-Time → O-Int		Goodman et al. (2000, chapter 1, p. 33)
O-Time → O-Distr	ICP	Goodman et al. (2000, chapter 1, p. 33)
		Goodman et al. (2000, chapter 7, p. 239)
		Goodman et al. (2000, chapter 11, p. 395)
		Okasha (2002, chapter 1, p. 2)
		Jacoby et al. (2016, p. 179)
		Janeck et al. (2003, p. 182)
		Blakey et al. (2017, p. 113)
		Marker et al. (2006, p. 390)
O-Time → O-Cont		Janeck et al. (2003, p. 182)
O-Time → C-Time		Goodman et al. (2000, chapter 6, p. 223)
		Goodman et al. (2000, chapter 11, p. 404)
O-Int → O-Distr	ICP + HICP	
O-Distr → C-Time		Najmi et al. (2009, p. 494)
		Manos et al. (2010, p. 700)
O-Distr → C-Int	ICP	
O-Cont → O-Int	ICP	
C-Time → O-Time		Goodman et al. (2000, chapter 6, p. 223)

Continued on next page

Table 6.2 – *Continued from previous page*

Edge	Algorithm	Literature
		Goodman et al. (2000, chapter 6, p. 229)
		Goodman et al. (2000, chapter 7, p. 241)
		Goodman et al. (2000, chapter 11, p. 393)
		Goodman et al. (2000, chapter 11, p. 399)
		Najmi et al. (2009, p. 495)
		Hezel & McNally (2016, p. 222)
		Hezel & McNally (2016, p. 223)
		Hezel & McNally (2016, p. 223)
		McNally & Ricciardi (1996, p. 17)
		Tolin et al. (2002, p. 1256)
C-Time → O-Distr		Goodman et al. (2000, chapter 7, p. 241)
		Goodman et al. (2000, chapter 11, p. 393)
		Goodman et al. (2000, chapter 11, p. 395)
		Goodman et al. (2000, chapter 11, p. 399)
		Najmi et al. (2009, p. 495)
C-Time → C-Int	HICP	Goodman et al. (2000, chapter 1, p. 33)
C-Time → C-Distr		Goodman et al. (2000, chapter 1, p. 33)
		Goodman et al. (2000, chapter 7, p. 239)
		Luchian et al. (2007, p. 1657)
		Okasha (2002, p. 3)
C-Int → C-Time	HICP	
C-Distr → C-Cont	ICP + HICP	
C-Cont → C-Distr	ICP + HICP	

Table 6.2: Overview of the source of each edge that is in Figure 6.2.

Another interesting aspect of this causal graph is that it depicts four cycles. The first

cycle (detected by both the ICP and the HICP algorithms) exists between *C-Distr* and *C-Cont*, which indicates that, when for example the score on *C-Distr* (distress associated with compulsive behaviour) changes, this causes the score on *C-Cont* (degree of control over compulsive behaviour) to change as well, and vice versa. This means that, on average, when patients score higher on *C-Cont*, and thus have a lower level of control over their compulsions, their score on *C-Distr* increases as well, and they on average experience more distress associated with their compulsions.

The second cycle in Figure 6.2a and 6.2c is between *C-Time* and *C-Int*. This cycle, which is only detected by the HICP-algorithm, may indicate that, when patients' score on *C-Time* (time spent on compulsive behaviours) changes, this causes the score on *C-Int* (interference due to compulsive behaviours) to change as well, and vice versa. This means that, on average, when patients score higher on *C-Time*, and thus spend more time performing compulsive behaviours, their score on *C-Int* changes as well, and they on average experience more interference in their daily life due to their compulsive behaviours.

The third cycle was between O-Distr and C-Time (Figure 2b and c). This cycle appeared only in the literature and suggests that when patients' score on O-Distr (distress associated with obsessive thoughts) changes, this causes the score on C-Time (time spent on compulsive behaviours) to change as well, and vice versa. This means that, on average, when patients score higher on O-Distr, and thus on average experience more distress due to their obsessive thoughts, their score on C-Time increases as well, and they spend more time performing compulsive behaviours.

The fourth and last cycle was between *C-Time* and *O-Time*, and also only appeared in the literature. This cycle indicate that, as patients' score on *O-Time* (time spent on obsessive thoughts) changes, their score on *C-Time* (time spent on compulsive behaviours) changes as well, and vice versa. Thus, spending more time on obsessive thoughts results in spending more time on compulsive behaviours, which in turn leads to spending more time on obsessive thoughts.

Interestingly, Figure 6.2 shows that the item *O-Time* is the cause of several other

nodes and only received causal influence from the item *C-Time*. This could imply that the amount of time spent obsessing is the key driver of the causal system depicted in the graph, at least for those individuals already diagnosed with OCD. This seems to confirm the (dominant) view that OCD starts with obsessive thoughts, although not everyone agrees with this (Gillan & Sahakian, 2015). Lastly, two clinical features are disconnected from the rest of the causal graph, *O-Res* and *C-Res*. It is possible that this symptom is disconnected because it is part of the treatment (exposure and response prevention) that was given to the patients that were part of this study.

Notably, only three causal relations connect the obsessive items with the compulsive ones, *O-Time* $\longleftrightarrow$ *C-Time*, *O-Distr* $\longleftrightarrow$ *C-Time* (both found in the literature) and the causal relation *O-Distr* $\longrightarrow$ *C-Int* (found with the ICP-algorithm), which in turn is also causally related to *C-Time*. This implies that there are only a few pathways whereby obsessive thoughts can lead to compulsions in patients, either directly or via some alternative path.

6.5 Discussion

In this exploratory study we introduced a causal graph approach and applied it to data of patients with OCD who completed the Y-BOCS-SR four times before, (twice) during, and after treatment. Using two algorithms, we estimated a causal graph for OCD by combining observational and interventional data. The addition of interventional data enables the estimation of causal relations that are hard to detect with only observational data. The resultant causal graph generated by the algorithms was complemented with a literature study. Results exposed several cyclic patterns between features of OCD. These may signify subsystems that maintain (parts of) the disorder throughout therapy. The ICP and HICP-algorithms did not estimate many causal relations between the items of the Y-BOCS-SR. The ICP-algorithm estimated six causal relations, and the HICP-algorithm estimated five causal relations, three of which were estimated by both algorithms.

In a graph where 45 different edges are possible, it is unclear whether the few causal

relations detected by these algorithms exhaust all of them or whether limited statistical power may have rendered others difficult to detect. Indeed, Kossakowski et al. (2020) showed that the ICP-algorithm can be quite conservative and will only use the intersection of all the subsets that are investigated. This means that edges will only be estimated when a specific edge appears in all the subsets that have significant causal relations with the target variable (i.e., the item that is the recipient of the causal influence). The conservative character of the ICP-algorithm is also shown by its high accuracy when the intervention is strong with a small variance, which is an aspect of interventional data that is difficult to ascertain. At the same time, Kossakowski and colleagues showed that the HICP-algorithm has a high sensitivity, but only a moderate to high specificity depending on the number of edges in the graph and the strength of the intervention. The lack of variance within the items could also have caused the small number of causal relations to be estimated. However, as shown in Table 6.1, the standard deviations of the items per time point are high enough to suggest otherwise.

The causal graph that shows the combined result of the ICP and HICP-algorithm visualised a discrepancy between these two algorithms and the results of the literature study. Only two causal relations that were detected by either of the two algorithms were confirmed with the literature study. It is quite possible that the causal relations estimated with the algorithms exist in the literature, but that we have not found the right studies that show evidence for these relations. Even though many articles were found, not all articles explicitly stated a causal relation between any two features of OCD, other than its two global, defining symptoms: obsessions and compulsions. A search on Web of Knowledge uncovered more than 5500 articles exist with the term "obsessive-compulsive disorder" in their title, so 17 different sources may not be representative of the existing literature on OCD. Another disadvantage of the current literature study is that the used sources did not experimentally test if one feature of OCD (e.g., time spent on obsessive thoughts) had a causal effect on another (e.g., distress associated with compulsions).

The ICP and HICP-algorithm state that a causal relation between two variables can

only be invariant when the residual distributions between time points are equal. This means that the causal relations found by the ICP and the HICP-algorithm reflect those causal relations that exist throughout therapy. Causal relations that change due to therapy are not included, as these relations cannot be picked up by either the ICP-algorithm or the HICP-algorithm. Although it is interesting to explore the causal relations that are not affected by therapy, it can also be interesting to study those causal relations that vanish as a result of therapy. Furthermore, in this study we examined patients who underwent treatment. We did not have access to a control or waitlist condition for which we could run the same analyses and compare the results. In future studies, adding a control condition and comparing results could illuminate differences attributable to therapy. Another option would be to use therapy as a dummy variable (when the data consists of an experimental and control condition) and estimate the causal graph with that dummy variable, similar to the approach that Blanken et al. (2019) used.

To our knowledge, this is the first study that estimates a causal graph for OCD using both observational and interventional data. By combining results from different algorithms and sources, we may uncover causal relations between features that otherwise are left unknown. The result of this causal graph approach may potentially be used to set up intervention strategies in a therapy setting. Even though we have to investigate to what extent this causal graph approach is useful and meaningful for clinical practice, we do believe that the causal graph approach has a great potential for illuminating the causal structure of psychopathology.

Part III

Discussion

7

Discussion

7.1 Overview

In this dissertation, I focused on two questions, (1) can we, and if so how, assess to what extent a complex dynamical system is in such a state that it can transition between two stable states, and (2) how well are we able to estimate a causal graph when we combine observational and experimental data. Chapter 2 provided an overview on various techniques that can be used to estimate network structures. Two models, the Cramer model and the Empirical Mean Field Approximation, were described and illustrated using empirical data. In chapter 3, we theoretically showed that it is possible to reduce a multidimensional dynamical system to a single equation, which in turn may be used to estimate a system's dynamical properties. In other words, the mean field model that is introduced here can be used to infer whether or not a system is in a space where two stable states exist, or in a space where only one stable state exists. chapter 4 expanded on this work and combined the mean field model with maximum likelihood estimation to estimate the parameter of interest in the mean field model. With this parameter, we could then assess the expectancy of an individual to transition between two stable states.

The second part of this dissertation focused on causality. Causality is a difficult topic. Many researchers strive to find causal relations between constructs, but few actually manage to detect these causal relations. Chapter 5 studied different algorithms to estimate

causal graphs. Here, we argued that using observational data alone will not give the entire causal picture. By combining observational data with experimental data, it is possible to detect meaningful causal relations that would otherwise stay undetected. Not only did we show the advantage of the combination of observational and experimental data in a simulation study, chapter 6 showed that this approach also results in interesting and meaningful causal relations when empirical data are used.

In this discussion I will reflect on both the mean field model and the (hidden) invariant causal prediction algorithms. Every model suffers from limitations, and it is important to acknowledge these. Specifically, I will cover some of the questions that I have had myself whilst writing this dissertation. What have we learned from these models, and what steps need to be taken to make them accessible for clinical researchers? Furthermore, how can these methods aid therapists in their goal to help patients who suffer from mental disorders? In answering these questions, I will also look forward and elaborate on possible extensions for the mean field model and the causal graph approach.

7.2 The research that is

The five chapters that form the body of this dissertation are the result of four years of doing research. In these four years, I have developed a mean field model with which we can assess if individuals are likely to transition between two stable mood states (chapter 2, 3 and 4), and I have taken a step into the domain of causality (chapter 5 and 6) where I looked into techniques for estimating a causal graph. These models offer a new view on (clinical) psychology. For the mean field model, this means that the first steps have been taken towards a goal where one can predict if an individual may experience a transition between two stable mood states. Predicting this transition may be of vital importance for the clinical field; predicting may ultimately result in preventing a transition from a healthy to a depressed mood state. The causal graph approach proves that a mental disorder like obsessive-compulsive disorder (OCD) cannot be captured with an acyclic graph. Most researchers agree that mental disorders contain feedback loops that represent

the vicious circles that maintain the disorder. For OCD, such a feedback loop could be *spending time on compulsions* $\longrightarrow$ *distress associated with obsessions* $\longrightarrow$ *spending time on compulsions*. Locating these feedback loops and other causal relations may enhance our understanding of a mental disorder, how it develops and how it is maintained. Gaining a complete overview of a mental disorder may result in the improvement of treatments of mental disorders by localising the most important chains in the causal graph.

Despite the promising results for both the mean field model and the causal graph approach, neither of them are, in their current state, suitable to use in clinical practice. Like most statistical models, the mean field model and the causal graph approach suffer from limitations, some more severe than others. Here, I will elaborate on three of these. I will discuss in a later section how the future may look for the mean field model and the causal graph approach.

7.2.1 An abundance of assumptions

Every statistical model comes with its own set of assumptions that need to be satisfied in order to interpret the results. The mean field model and the causal graph approach are no exception to this. However, when a model has so many assumptions that it places too many restrictions on the data, we may wonder to what extent this bundle of assumptions also restricts the results and consequential interpretations. For example, the mean field model assumes that all nodes in a network display the same behaviour and change from active to inactive (or vice versa) in a similar manner. Moreover, the model assumes that the system can only transition between two stable states, if it is in a space where transitions can occur. The set of assumptions for the mean field model is complemented by assumptions with respect to the data, some of which will be discussed later in more detail.

We are required to make these assumptions in order to present a working model, but it can be debated if these assumptions can be met when it comes to clinical psychology. This debate is open-ended at the time of writing; we do not know yet how the interaction between symptoms differs for each pair, if at all. Also, the assumption that every node in

a (psychological) network changes from active to inactive (or vice versa) the same way may not be satisfied in clinical practice. It is likely that a node like "sleep problems" develops differently, and thus activates in a different manner, than say, "suicidal thoughts". Since it is possible that nodes change differently, they may also display other behaviours, which is also an assumption. Currently, we have no way of knowing to what extent these assumptions are violated when collecting psychopathological data. Measuring individual symptoms over time, or perturbing individual symptoms may give us this knowledge on how symptoms develop and affect one another, although perturbing individual symptoms is an issue on its own, which will be discussed later.

Some agree that mental disorders like depression have two stable states: a healthy and a depressed mood state, while others think that some individuals develop depression is a more gradual manner. Also, the two-stable system may not fit every mental disorder that appears in the diagnostic and statistical manual of mental disorders (American Psychiatric Association, 2013). More specifically, mental disorders like developmental or personality disorders may not adhere to this view. The results that are discussed in this dissertation should be interpreted in the light of the assumptions that are made for that specific model, and users should be aware of these assumptions as they may limit inferred conclusions or generalisations to a broader population.

7.2.2 Completeness of data

Both the mean field model and the causal graph approach require complete, non-missing data. This resulted in the exclusion of participants and the replacement of missing data. Both ways of dealing with missing data are less than ideal. Excluding participants may lead to a (great) loss of data. To illustrate, in chapter 6, we started with 3474 patients, and included only 903 patients, which is close to 26% of the original sample size. In chapter 4, we included approximately 60% and 81% of the original sample sizes. Excluding data is never the preferred solution when it comes to missing data. It is costly, both in time and financial sense, to collect data. Filling out questionnaires can be burdensome for

participants, especially when one has to complete questionnaires multiple times per day. We do not do justice to participants when we ask them to complete several questionnaires only to throw them away during the analysis phase of our research, even though arguments for excluding data and/or participants is justified. I believe that we as researchers have the obligation to develop statistical models that require less data, thereby lowering the burden on the participants, and more optimal solutions in dealing with missing data.

In chapter 4, we replaced any missing measurements with either the previous measurement, or by removing the measurement entirely. While both methods returned similar results, one may call into question whether this is an optimal way of dealing with missing measurements. Replacing missing measurements (while keeping all other thing equal) often results in a decrease in both the mean and the variance of an item. This in turn may lead to lower estimation of the probability parameter p and a bigger chance of concluding that an individual is in a space where a transition between two stable states is possible. Regardless of the current mood state of the participant, under- or overestimating the probability parameter p may lead to the wrong conclusion and possibly to a faulty treatment plan if the model is used in clinical practice. In studies where non-time-series data are used, data imputation is an available technique that conserves the properties that exist in the data.

Before imputing missing data, it important to first decide what type of missing data one is dealing with. Missing data can be *completely at random* (MCAR; there is no underlying process for the missing data, all missing data occur at random. An example of MCAR data includes a researcher forgetting to hand out a questionnaire), *at random* (MAR; the location of the missing data depends on participant characteristics, the previous measurement or on some other process. An example of MAR includes only observing missing data in a subgroup), and *not at random* (MNAR; the missing data is influenced by unobserved processes, such as spontaneous recovery; Little & Rubin, 2002). While MCAR is the preferred type of missing data, MAR is the most occurring type. Data imputation for time-series data is not trivial. More often than not, missing data in time-series

data is MNAR. For example, participants may not complete any measurement when they are feeling bad. Sun et al. (2019) studied missingness in time-series data and found that missing data was hard to predict. This research suggests that missing data may be less of an issue than we thought, although more research has to be conducted to get a definitive answer. At the time of writing this discussion, data imputation is difficult and suboptimal for time-series data.

7.2.3 Perturbations

In chapter 5 I argued that using both observational and experimental data may return the most reliable causal graph. Experimental data here are data on which some perturbation has taken place. The issue here lies in what is meant by "some perturbation". According to Webster's International Dictionary (2008), a perturbation is "the action of perturbing or condition of being perturbed", and to perturb means to "to throw into disorder or disturb". A perturbation is thus something that disturbs something else. Disturbing something does not bring us any closer to a concrete definition of what a perturbation is. A perturbation can take place on two levels in a dynamical system (see Figure 2.8): we can perturb the behaviour, by actively pushing specific variables. Examples of this include giving a patient sleep medication to help with his or her insomnia. With these kinds of perturbations, an individual may experience a transition, but this is only a short-term effect. By perturbing the "stress" variable of the system, we aim to change the dynamical landscape that an individual is in, resulting in (hopefully) long-term effects in terms of psychological well-being. Even though the causal graph algorithms developed by Meinshausen et al. (2016) and Peters et al. (2017) do not set any requirements to the nature of a perturbation in a clinical sense, when using these algorithms one needs to consider what kind of perturbation is placed upon participants and their corresponding data, and what the resulting causal graph means in light of this perturbation.

In psychology, to perturb often means that something is manipulated. Examples include giving patients either a pill with working elements or a placebo, or giving treatment

versus being put on a waiting list. We can use this type of experimental data to estimate a causal graph, but then we may run into interpretation problems. The edges in this causal graph represent those causal relations that are shared by both the experimental and the control condition, whereas most experiments are performed to find if there is a difference between the two. A classic experimental design with two (independent) groups may not be the best design to use when estimating a causal graph.

Another issue with the types of perturbations that are often used is that they perturb the entire system. I wonder if we can apply node-specific perturbations to psychology, similar to the gene knock-out experiments for which the DR-FFL and the TRANSWESD algorithms were developed (Pinna et al., 2010). A perturbation that is often used in psychology include therapy and/or medication. While some specific medications like sleeping medication may only perturb one symptom, it is more likely that they "attack" the entire system that is the mental disorder. Here, we also have to consider how we define symptoms. It is beyond the scope of this dissertation discuss this problem to its full extent. Research by Fried (2017) showed that different questionnaires often result in different operationalisations of depression symptoms, which may lead to validity issues: to what extent do we measure the symptoms of a mental disorder like depression?

7.3 The research to come

The results in this dissertation are in no way the final destination of these methods in psychology, they are merely the beginning. When you are at the start of a road, many options seem possible. This also holds for the mean field model and the causal graph approach. In this dissertation I presented starting points for both methods that work when many assumptions are satisfied. Some of the possible extensions have been discussed in chapters 4 and 6, but I will elaborate on them and others in more detail here.

7.3.1 Complex System

The mean field model is designed to take the essence of a complex dynamical system and place this in a space where either one or two stable states are possible. By only looking at the number of active nodes at each time point, we can estimate a probability parameter p and compare this to a bifurcation diagram, which then tells us if the system is able to make a transition between two stable states or not. The data published by Kossakowski et al. (2017) were used to validate this approach. Taking only the data up until the transition, we showed that the mean field model is able to detect whether an individual can transition between two stable (mood) states.

While the mean field model as described in chapters 2, 3 and 4 already works well, there are several extensions possible that may enhance the performance of the mean field model. Take for example the mean field model as it is formulated in equation (3.2). Here, we take the entire complex dynamical system as a whole and reduce it to just the number of active nodes per time point that is then used to determine the probability parameter p. Often, these complex dynamical systems consist of items that can be categorised in positive and negative items. The questionnaire used by van der Krieke et al. (2015) contained items such as "I feel nervous" and "I feel calm". It is possible to apply not just one complex dynamical system, but two interacting systems: one for the positive items, and one for the negative items. By separating the items that measure positive affect and negative affect, we can adjust the mean field model to fit those sets of items, thereby improving the estimation of the probability parameter p and the inference based on that parameter. The extension to two complex dynamical systems also makes sense from a clinical point of view. In order to get a major depressive disorder diagnosis, a patient must experience either a depressed mood, or a loss of interest (American Psychiatric Association, 2013). In other words, a patient must either experience an increase in negative mood, or a decrease in positive mood. We can even extend the number of dynamical systems to match the number of items that are part of the original complex dynamical system. As many questionnaire items or symptoms present themselves differently in different individuals,

attaching a dynamical system to each item further personalises the mean field model and may improve the estimation of not only the probability parameter, but also the conclusions taken from that parameter.

The mean field model as it is formulated in chapters 3 and 4 is currently only suitable for dichotomous data: data with only two response categories. In psychology, very few (if any) constructs or symptoms are dichotomous in nature (Borsboom et al., 2016). For example, there is a degree to which one may suffer from sleep problems. Also, one major depressive episode may be stronger than another. While using dichotomous data is a good starting point, they may not accurately represent how an individuals truly feels. Fortunately, it is possible to rewrite the mean field model in such a way that, instead of a binomial distribution, a Gaussian distribution can be used to describe the probability for a node to be active (equation (3.2)). With a Gaussian distribution, the mean field model becomes accessible for continuous data, which is a vast improvement from dichotomous data. A next step would be to combine the mean field model with ordinal data, or even a combination of dichotomous, continuous and ordinal data. Many questionnaires exist that measure major depressive disorder, and each questionnaire has its own set of items and response scales (Fried, 2017). Developing a mean field model that accommodates different questionnaires makes the model more accessible for clinical researchers, who cannot always adjust their study design to the model that is to be used later on in the analysis stage of their research.

The probability parameter p operationalises whether or not an individual can experience a transition between two stable (mood) states. In the mean field model as I have discussed it in this dissertation, we estimate this parameter only once for the entire time series. It thus gives an indication of an individual's position for the entire data collection period. If the parameter is smaller than the critical value, two stable states exist and an individual can experience a transition between the two states. If the parameter is greater than the critical value, only one stable state exists and a transition is not possible. We do not interpret the value of the probability parameter, other than to make a comparison

to the critical value. An interesting next step would be to employ a "moving window" approach and estimate a time-varying probability parameter p. One can imagine that the probability for a node to be active may change as time progresses, and may also change as a result of significant life events. A time-varying parameter may aid therapists as it may indicate when a patient, who is in a depressed mood state, can experience a transition to a healthier mood state.

7.3.2 Causal System

A causal graph visualises how variables influence each other. An arrow from variable X to variable Y indicates that, when a change occurs in variable X, a change will also occur in variable Y. A causal graph combines all these causal relations and gives an overview of a (psychological) construct and how all its individual bits and pieces are connected. All the causal relations that exist within this graph denote relations that survived some perturbation. This also means that, if a causal relation changes after some perturbation has taken place, it will not show up in the final causal graph.

In chapter 5, I showed how to estimate a causal graph and what the advantages and disadvantages of various algorithms are. Two algorithms showed the most promise for (clinical) psychology: the Invariant Causal Prediction algorithm (ICP; Meinshausen et al., 2016) and the Hidden Invariant Causal Prediction algorithm (HICP; Peters et al., 2017). Where the ICP-algorithm can be quite conservative and return a small number of connections, the HICP-algorithm can return many false positives. A combination of the two, like I showed in chapter 6, may therefore be a good alternative.

The conservativeness of the ICP-algorithm is possibly due to the design of the algorithm: it investigates every possible subset, and will only accept the intersection of the subsets as the set of nodes that are the cause of the target (causally receiving) node. With more nodes comes a larger group of subsets to investigate, and thus a smaller chance of causal relations to be accepted as "true". When the perturbation effect is strong and precise, the number of subsets becomes just a computational issue. However, when analysing

empirical data, the strength and precision of the perturbation is unknown, and the issue is no longer just a computational one. In order to be able to interpret an empirical causal graph, one needs to know if all causal relations are picked up. We saw in chapter 5 that the Peter and Clark algorithm (PC; Kalisch & Buhlmann, 2007) can accurately find the skeleton graph that forms the base of a causal graph. Since the PC-algorithm only uses observational data, the strength and precision of a perturbation do not have an influence on the performance of the PC-algorithm. Perhaps we can combine the skeleton finding skills of the PC-algorithm, and combine it with the ICP-algorithm. In this scenario, the skeleton graph that results from the PC-algorithm serves as input for the ICP-algorithm, and thus specifies the subsets and target nodes that are to be investigated. Here, we would decrease the number of subsets to be investigated, thereby increasing the chance for causal relations to be accepted.

Where the ICP-algorithm can investigate a great number of subsets, the HICP-algorithm only uses the entire group of remaining nodes as input. Since there is no intersection to select, every causal relation that is significant is accepted and thus exists in the causal graph. As a result, spurious causal relations may appear in the causal graph. The higher the sample size, the more spurious causal relations we observe in the causal graphs (see chapter 5). In an extension, the subset design that is implemented in the ICP-algorithm could also be employed by the HICP-algorithm. The computational limits that exist for the ICP-algorithm would also exist for the HICP-algorithm, possibly to a greater extent since more steps are involved in the HICP-algorithm to account for possible hidden variables. Naturally, the combination with the PC-algorithm to decrease the number of subsets can be applied here as well.

7.3.3 A Complex Causal System

Throughout this dissertation I discussed the two main topics (complex systems and causal systems) separately. An interesting next step would be to combine the two systems. Can we combine the two into one complex causal system? One could first estimate a causal

graph using time-series data, after which this graph serves as input for the mean field model. The causal graph indicates which items are causally related, and therefore also indicates which items are neighbours of each other. The latter is needed for the mean field model so that it can apply the majority rule (presented in equation (3.1)) properly. Although this may seem like a simple extension, there are some aspects in both the mean field model and the causal graph that need to be thoroughly considered before attempting this. In the mean field model, the first issue that we are presented with is the choice of network structure. In its current version, we described two network structures that return good results in the estimation of the probability parameter p: the random graph (Bollobás, 2001) and the small-world graph (Watts & Strogatz, 1998). These network structures are suitable for undirected networks, and thus can not be used when we want to incorporate a causal graph with the mean field model. A directed alternative is the Price model (Newman, 2010), but this specific network structure comes with its own advantages and disadvantages.

For the causal graph, an issue that we need to consider is the division of the data in order to create (at least) two environments, which are needed to apply the (hidden) invariant causal prediction algorithms. One can use time-series data to estimate a causal graph, but this type of data are often observational, meaning that no perturbations are implemented in the design of the study. Theoretically, this is easily solved: one has to add some perturbation design whilst collecting data. In practice, this turns out not to be a trivial matter, as we saw earlier in this discussion. A design like the one used in the data published by Kossakowski et al. (2017) is an option for participants who are on medication and want to reduce their medication intake. For participants that do not fall into this category, a possibility is to collect time-series data in two periods, with a "resting" period in between in which no questionnaires are completed.

7.4 Conclusion

In this dissertation I have introduced a new model for complex systems and an approach for estimating a causal system. For the former, I showed that we can average the system in such a way that we assess whether or not the system is under such an amount of pressure that a transition can take place between two stable states. I showed for the latter that we can estimate a causal graph when we not only observe a system, but also put a system under some pressure by perturbing it. With this combination causal relations can be accurately estimated.

Both models represent a different view on the problem that they are aiming to solve. For the mean field model (chapters 2, 3 and 4), this means that, next to trying to explain mental disorders such as depression, we wanted to know if we could accurately assess whether an individual can experience a transition or not. Here, the pressure that is put on the system is passive in nature: a complex dynamical system is not actively forced towards the tipping point between a healthy and a depressed mood state, but instead shifts towards this tipping point because of circumstances inside and outside the individual.

For the causal graph approach (chapter 5 and 6), I offered a new point of view in the estimation of causal graphs in (clinical) psychology. Causality can be a difficult and sensitive topic. Many researchers attempt to find causal relations, but few manage to accurately estimate them. I illustrated the potential of the causal graph approach in the explanation of mental disorders like OCD. With the causal graph approach, pressure is actively put on the causal system by means of perturbations. With a perturbation in place, causal flow can be observed since a change in the cause will lead to a change in the effect.

All in all, this dissertation may be used as the beginning of new lines of research, where the focus shifts from assessing transitions to predicting them, and from observing to perturbing in order to estimate causal graphs. These two goals in psychological research, explanation and prediction, cannot be just be seen separately, they go hand in hand. One cannot predict without understanding (parts of) a mental disorder, and prediction in turn may lead to a greater understanding of a mental disorder. This dissertation

showed how important both aspects are, and it may bring us new insights into systems that are actively or passively put under pressure.

Part IV

Appendices

A

Supplementary Material to Chapter 3

Proof (Lemma 2) Let the Kullback-Leibler divergence between $p+\varepsilon$ and p be defined as a function of $0 \leq \varepsilon \leq p$

$$h_+(\varepsilon) = (p+\varepsilon)\log\frac{p+\varepsilon}{p} + (1-p-\varepsilon)\log\frac{1-p-\varepsilon}{1-p}$$

and similarly, define $h_-(\varepsilon) = h_+(-\varepsilon)$. Then Chernov's bound (Lesigne, 2005; Venkatesh, 2012) for the density ρ_t of a grid with n nodes and its mean at time t, $p_{\text{grid}}(\rho_t$ defined in (3.2), for $0 < \varepsilon < \min\{p_{\text{grid}}, 1-p_{\text{grid}}\}$ immediately gives

$$\mathbb{P}(|\rho_t - p_{\text{grid}}(\rho_t)| > \varepsilon) \leq \exp(-nh_+(\varepsilon)) + \exp(-nh_-(\varepsilon))$$

The Kullback-Leibler divergence can be approximated quadratically by

$$h_+(\varepsilon) = \frac{\varepsilon^2}{2p(1-p)} + O(\varepsilon^3)$$

Similarly for $h_-(\varepsilon)$ gives

$$\mathbb{P}(|\rho_t - p_{\text{grid}}(\rho_t)| > \varepsilon) \leq 2\exp(-\varepsilon^2/2\sigma^2_{\text{grid}}(\rho_t)) \tag{A.1}$$

where $\sigma^2_{\text{grid}} = p_{\text{grid}}(1-p_{\text{grid}})/n$. Let $\delta = 2\exp(-\varepsilon^2/2\sigma^2_{\text{grid}}(\rho_t))$ such that $\varepsilon = \sqrt{2\sigma^2_{\text{grid}}\log(2/\delta)}$. Then we obtain the result with probability at least $1-\delta$. □

Proof (Equation 3.8) To obtain (3.8) assume a fixed value ν for all neighbourhood sizes k, maj_p uses the value $k=\nu$. Then maj_p only depends on the number of active neighbours r. First note that

$$\binom{k}{r}\binom{n-1}{k}=\binom{n-1}{r}\binom{n-r-1}{k-r}$$

Second, by changing the order of summation and reordering the sums, we get

$$\sum_{r=0}^{n-1}\mathrm{maj}_p(r,n-1)\binom{n-1}{r}\rho_t^r p_e^r\sum_{k=r}^{n-r-1}\binom{n-r-1}{k-r}(p_e(1-\rho_t))^{k-r}(1-p_e)^{n-k-1}$$

In the sum on the right we can use the binomial theorem with $m=k-r$ and $N=n-r-1$, which gives

$$\sum_{m=0}^{N}\binom{N}{m}(p_e(1-\rho_t))^m(1-p_e)^{N-m}=(p_e(1-\rho_t)+1-p_e)^N$$

which leads to (3.8).

For the approximation error, write $\mathrm{maj}_p(r,k)=p\mathbb{1}\{r\le k/2\}+(1-p)\mathbb{1}\{r>k/2\}$ and recall that ν is fixed. Then

$$\begin{aligned}&p_{\mathrm{rg}}(\rho_t)-p^{\nu}_{\mathrm{grid}}(\rho_t)=\\&\quad\sum_{k=0}^{n-1}\sum_{r=0}^{k}\binom{k}{r}\rho_t^r(1-\rho_t)^{k-r}\binom{n-1}{k}p_e^k(1-p_e)^{n-k-1}[\mathrm{maj}_p(r,k)-\mathrm{maj}_p(r,\nu)]\end{aligned}$$

Using Hölder's inequality with the ℓ_∞ and ℓ_1 norms, gives

$$\begin{aligned}&\left|\sum_{k=0}^{n-1}\sum_{r=0}^{k}\binom{k}{r}\rho_t^r(1-\rho_t)^{k-r}\binom{n-1}{k}p_e^k(1-p_e)^{n-k-1}[\mathrm{maj}_p(r,k)-\mathrm{maj}_p(r,\nu)]\right|\\&\le\sum_{k=0}^{n-1}\sum_{r=0}^{k}\binom{k}{r}\rho_t^r(1-\rho_t)^{k-r}\binom{n-1}{k}p_e^k(1-p_e)^{n-k-1}\max_{r,k}|\mathrm{maj}_p(r,k)-\mathrm{maj}_p(r,\nu)|\end{aligned}$$

The binomial theorem for the first term of the right hand side gives

$$\begin{aligned}&\sum_{k=0}^{n-1}\sum_{r=0}^{k}\binom{k}{r}\rho_t^r(1-\rho_t)^{k-r}\binom{n-1}{k}p_e^k(1-p_e)^{n-k-1}\\&=\sum_{r=0}^{n-1}\binom{n-1}{r}(\rho_t p_e)^r(1-\rho_t p_e)^{n-r-1}=1.\end{aligned}$$

For each r,k let $\mathrm{maj}_p(r,k)$, where k denotes the neighbourhood size, such that $r\le k$ we have that

$$\text{maj}_p(r,k) - \text{maj}_p(r,\nu) =$$
$$p(\mathbb{1}\{r \leq k/2\} - \mathbb{1}\{r \leq \nu/2\}) + (1-p)(\mathbb{1}\{r > k/2\} - \mathbb{1}\{r > \nu/2\})$$

The term $|\text{maj}_p(r,k) - \text{maj}_p(r,\nu)|$ is at most $2p-1$ if $\nu < k$ or $1-2p$ if $\nu \geq k$ for any r,k, which gives the size of the error bound. □

Proof (Proposition 3) If we fix $\nu = \lfloor p_e(n-1) \rfloor$, the expectation of the random variable for each node of the possible number of neighbours $B(n-1,p_e)$, such that each $k = \nu$ in the part for the density we obtain

$$\sum_{k=0}^{n-1} \binom{n-1}{k} p_e^k (1-p_e)^{n-k-1} \left(\sum_{r=0}^{\nu} \binom{\nu}{r} \text{maj}_p(r) \rho_t^r (1-\rho_t)^{\nu-r} \right),$$

from which we obtain $p^{\nu}_{\text{grid}}(\rho_t)$. The approximation error for the probabilities is then

$$|p_{\text{rg}}(\rho_t) - p^{\nu}_{\text{rand}}(\rho_t)| = \left| \sum_{k=0}^{n-1} (p^k_{\text{grid}}(\rho_t) - p^{\nu}_{\text{grid}}(\rho_t)) \binom{n-1}{k} p_e^k (1-p_e)^{n-k-1} \right|.$$

The probability of obtaining a neighbourhood size k close to the expected number of neighbours $\nu = p_e(n-1)$ can be obtained from the Chernov bound in Lemma 2, giving $\mathbb{P}(|k-\nu| \leq t) \geq 1 - 2\exp(-(n-1)\varepsilon^2/p_e(1-p_e))$, for $\varepsilon \searrow 0$. This leads to the difference being bound by

$$|p_{\text{rg}}(\rho_t) - p^{\nu}_{\text{rand}}(\rho_t)| \leq \left| \sum_{k=0}^{n-1} (p^k_{\text{grid}}(\rho_t) - p^{\nu}_{\text{grid}}(\rho_t)) 2\exp(-(n-1)\varepsilon^2/p_e(1-p_e)) \right|.$$

Using Hölder's inequality with the sup and ℓ_1 norms, we find that the above is

$$\leq \max_k |p^k_{\text{grid}}(\rho_t) - p^{\nu}_{\text{grid}}(\rho_t)| \sum_{k=0}^{n-1} 2\exp(-(n-1)\varepsilon^2/p_e(1-p_e)).$$

The difference $p^k_{\text{grid}}(\rho_t) - p^{\nu}_{\text{grid}}(\rho_t)$ is determined by the mismatch between k and ν and is at most $2p-1$ if $\nu < k$ and $1-2p$ if $\nu \geq k$ for any r,k. And so we obtain

$$|p_{\text{rg}}(\rho_t) - p^{\nu}_{\text{grid}}(\rho_t)| \leq |p - 1/2| 2\exp(-(n-1)\varepsilon^2/p_e(1-p_e) + \log(n)),$$

completing the proof. □

Proof (Lemma 6) We showed already in Section 3.4.1 that the transition kernel is a contraction and so with t increasing we will decrease the distance in total variation to 0 between $\pi_0 K_{0,t}$ and $\pi_0' K_{0,t}$. For this to work we need that for each t, $K_{t,t+1}$, the single step transition probability, is irreducible. This implies that the transition probability $\mathbb{P}(Y_{t+1} = m \mid Y_t = k)$ from state k to state m

$$K_{t,t+1}(k,m) = \binom{n}{m} p_{\text{grid}}(\tfrac{k}{n})^m (1 - p_{\text{grid}}(\tfrac{k}{n}))^{n-m}$$

with

$$p_{\text{grid}}(\rho_t) = \sum_{r=0}^{|\Gamma|} \text{maj}_p(r) \binom{|\Gamma|}{r} \rho_t^r (1-\rho_t)^{|\Gamma|-r}$$

in (3.16) cannot be 0 or 1. From Section 3.3.2 it is clear that p_{grid} converges to 0 as $t \to \infty$ when ρ_t is in the stable set $S(0)$ of p_{grid}. Hence we require that ρ_t is outside of $S(0)$. Likewise, we require that ρ_t is not in the stable set $S(1)$. So, only if p_{grid} converges to 0 will $K_{t,t+1}$ converge to 0. For any p in the majority function maj_p such that it is not in $S(0) \cup S(1)$ we define $\eta_t = \min_{t,k,m} K_{t,t+1}(k,m)$, which implies that for each time point t and any states k and m there is an η_t such that $K_{t,t+1}(k,m) \geq \eta_t$, which is uniform (over states) irreducibility. We can therefore conclude that whenever ρ_t is not in $S(0) \cup S(1)$, then we obtain merging. Since

$$\pi_{t+1}(k) = \sum_{r=0}^{n} \mathbb{P}(Y_{t+1} = k \mid Y_t = r) \mathbb{P}(Y_t = r)$$

and

$$\min_{r,k} \mathbb{P}(Y_{t+1} = k \mid Y_t = r) \sum_{r=0}^{n} \mathbb{P}(Y_t = r) \leq \sum_{r=0}^{n} \mathbb{P}(Y_{t+1} = k \mid Y_t = r) \mathbb{P}(Y_t = r)$$

we obtain the inequality $\min_{k,r} \mathbb{P}(Y_{t+1} = k \mid Y_t = r) \leq \pi_{t+1}(k)$. Similarly for the other inequality, giving the range for all probabilities $\pi_{t+1}(k)$. □

B

Supplementary Material to Chapter 4: Data from 'Critical Slowing Down as a Personalized Early Warning Signal for Depression'

Abstract

We present a dataset of a single (N=1) participant diagnosed with major depressive disorder, who completed 1478 measurements over the course of 239 consecutive days in 2012 and 2013. The experiment included a double-blind phase in which the dosage of anti-depressant medication was gradually reduced. The entire study looked at momentary affective states in daily life before, during, and after the double-blind phase. The items, which were asked ten times a day, cover topics like mood, physical condition and social contacts. Also, depressive symptoms were measured on a weekly basis using the Symptom Checklist Revised (SCL-90-R). The data are suitable for various time-series analyses and studies in complex dynamical systems.

B.1 Overview

B.1.1 Collection Date(s)

The data were collected between August 13, 2012, and April 11, 2013.

B.1.2 Background

We present a dataset of a single participant, with a history of major depressive disorder (MDD), whose daily life experience was monitored over the course of 239 days (Wichers et al., 2016). The partici-

pant, who initiated the experiment, wanted to obtain more personal insight during a period in which the anti-depressants were gradually reduced. The aim of the participant was to know whether or not he would become more vulnerable to develop a new depressive episode when the anti-depressants were reduced, and whether this vulnerability could be detected in the data. The study design was set up at the initiative of, and in collaboration with, the participant, who agreed upon the set of items that was selected and added some items that were relevant to the participant. The participant had also found a pharmacist who provided a dose-reduction scheme that was randomly chosen out of several dose-reduction schemes that were designed in collaboration with the participant. The chosen dose-reduction scheme was unknown to both the participant and the researchers involved in the experiment. The participant was monitored on a momentary basis during a baseline period, a period of dose-reduction and a post-reduction period. The participant also initiated a follow-up measurement period in which his daily life experiences kept being monitored.

B.2 Methods

B.2.1 Sample

The participant is a 57-year old male with a history of MDD. The participant has been using antidepressants for 8.5 years (Groot, 2010). The participant completed on average 6.2 assessments per day ($SD = 1.9$).

B.2.2 Materials

The dataset consists of items that were measured momentary, daily, and weekly.

Momentary

Momentary items (no. items is 50) were collected using the experience sampling method (ESM; Csikszentmihalyi & Larson, 1987). Items were selected based on previous experience with these types of items regarding within-person variation, loading on a negative or positive affect component, the relevance for the current type of psychopathology and specific personal characteristics of the participant. The momentary assessment questionnaire consisted of items pertaining to mood states (e.g. feeling relaxed, feeling irritated etc.), self-esteem, the company the participant

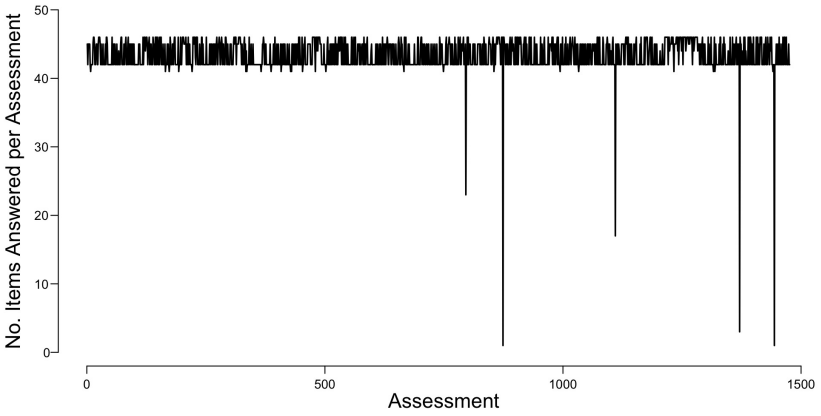


Figure B.1: Number of items answered per assessment.

was in at the moment of assessment, the pleasantness or unpleasantness of being in company, physical condition, the activity that the participant was doing at the moment of assessment, and to an important event that had occurred since the last assessment.

33 items were measured on a 7-point Likert scale, ranging from 1 (not) to 7 (very). The items concerned with feeling down, lonely, anxious and guilty were measured on a 7-point Likert scale, ranging from -3 (not) to $+3$ (very). This scale differs from the majority of the items as a pilot trial showed that the participant reported more variation with a 7-point Likert scale ranging from -3 to $+3$ than a 7-point Likert scale ranging from 1 to 7. Items with respect to the company the participant was in, the activity the participant was enrolled in or any event that had taken place since the last assessment were categorised accordingly. Two items concerning the pleasantness and importance of the event were measured on a 7-point Bipolar scale, ranging from -3 (unpleasant/unimportant) to $+3$ (pleasant/important). Figure B.1 shows the number of items answered per assessment. The average number of items answered per assessment is 43.4 ($SD = 2.7$).

Daily

Two separate sets of items were completed daily: a six-item set after waking up (item labels start with 'mor'), with which information is collected on the quality of sleep. The other set consists of six items and were asked to be filled out right before the participant went to sleep (item labels start with 'evn'). This set of items was concerned with the quality of the day the participant had. At both occasions, the participant was asked whether or not he took his medication either yesterday (morning item), or today (evening item). Three of the morning items were categorised accordingly. Two of the morning items and four of the evening items were measured on a 7-point Likert scale, ranging from 1 (not) to 7 (very). The medication items together with one evening item was measured dichotomously (yes/no).

Weekly

Once a week the participant completed 13 items of the depression subscale of the Symptom Checklist Revised (item labels start with SCL-90-R; Derogatis et al., 1976). Each item is scored on a 5-point Likert scale, with 0 meaning that the participant wasn't bothered by that specific thought or feeling at all, and 4 meaning that he was extremely bothered by it.

B.2.3 Missing Data

Missing data for momentary items

Each day, 43.4 items were filled out on average ($SD = 1.3$) per day at each assessment. In total, out of 1478 assessments, only five were aborted before completion. At 1478 assessments, items were completed on average 1280.4 times ($SD = 378$). Figure B.2 depicts the number of assessments completed per day. The average number of completed assessments per day is 6.2 ($SD = 1.9$).

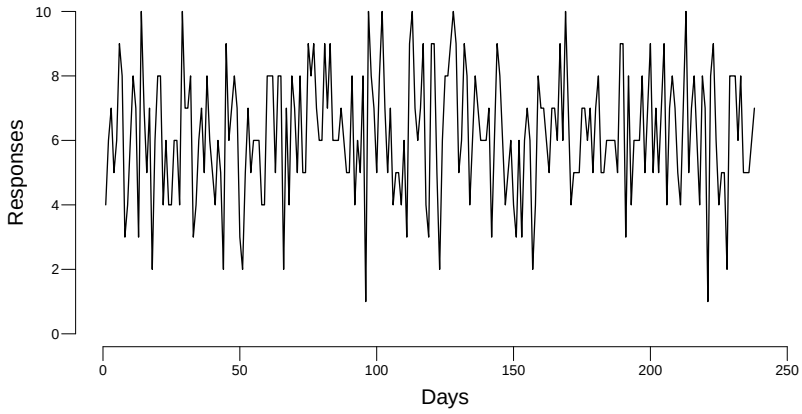


Figure B.2: Number of assessment completed per day.

Missing data for daily items

During the entire study, 59.7% of the set of morning items and 81.5% of the set of evening items were completed. Each day, either the entire set of morning/evening items was answered, or none. On average, 3.6 ($SD = 2.9$) of the morning items and 4.9 ($SD = 2.3$) of the evening items were completed.

Missing data for weekly items

The study lasted for 34 weeks. During this period, depressive symptoms were measured on a weekly basis. This weekly questionnaire was completed 28 times (= 82.4%). On each of these occasions, all items were answered.

B.2.4 Procedures

The entire study comprised 5 phases: (1) a baseline measurement period that lasted four weeks, (2), a double-blind period in which the anti-depressant dosage was not yet reduced, which lasted between zero and six weeks, (3) a double-blind period in which the anti-depressant dosage was gradually reduced from 150 mg (venlafaxine) to 0 mg, which lasted eight weeks, (4), a post-assessment period in which the anti-depressant dosage was not changed, which lasted again eight weeks, and (5), a follow-up period that lasted twelve weeks.

The dose-reduction scheme issued in phase 3 was set up by the pharmacist who provided the anti-depressants during the

study. Several reduction schemes were developed, which varied with respect to the length in weeks before the dose reduction started (phase 2). During the experiment, the participant and the researchers involved were unaware of the dose-reduction scheme, although they did know that the anti-depressant dosage was going to be reduced. The participant reported after the experiment that he had not been able to figure out which eventual dose-reduction scheme had been used.

At the start of the experiment, the participant received a PsyMate (a digital device with touch screen; Wichers et al., 2011), which was set up to send out a beep-signal at random moments within each of ten 90-minute intervals between 07.30 AM and 10.30 PM every day. At each beep-signal, the participant completed a 50-item questionnaire. Each beep-signal was accompanied by a ten-minute window in which the questionnaire was available to the participant. Assessments were started on average within 2.16 minutes (SD = 21 seconds). At the beginning and ending of each day, the participant was asked to complete an extra set of six items. On Mondays, the participant's depressive symptoms were measured using the depression subscale of the SCL-90-R (Derogatis et al., 1976).

B.2.5 Quality Control

All questionnaires were administered by means of a digital device (PsyMate). In a few cases, the SCL-90-R was completed on paper and e-mailed to the researchers, who added the scores to the dataset.

B.2.6 Ethical Issues

The participant (the 2nd author of this paper) initiated the study and expressed that he wanted the data to be published. Approval from the Maastricht University ethical committee was therefore unnecessary and not obtained. The participant gave his consent for collecting and (re)using the data.

B.3 Dataset description

B.3.1 Object name

The datafile is named "ESMdata.zip". This zip file contains the data "ESMdata.xls", "ESMdata.csv", "ESMdata.txt", a code-

book “Codebook.pdf” and a supplement “MissingnessMomentaryItems.pdf”.

B.3.2 Data type

All datafiles are primary data, with the exception of one variable called “dep”. This is a mean score of the SCL-90-R items as is mentioned as such in the codebook.

B.3.3 Format names and versions

The data are provided in three different formats: .xls, .csv and .txt format. The accompanying codebook and the supplement are in .pdf format.

B.3.4 Data Collectors

Peter Groot and Marieke Wichers designed the entire study and the experiment. Frenk Peeters was involved as a psychiatrist in the design phase of the experiment, Claudia Simons was responsible for the ESM briefing and technical assistance regarding the use of the PsyMate to collect the data.

B.3.5 Language

English.

B.3.6 License

The data have been deposited under a CC-By Attribution 4.0 International (CC-By) License.

B.3.7 Embargo

Not applicable.

B.3.8 Repository location

http://osf.io/j4fg8

B.3.9 Publication date

The data have been published online since November 30, 2016.

B.4 Reuse potential

The dataset contains around 1500 measurements and over 50 items. Furthermore, items have been completed at different time scales: momentary, daily and weekly. It is a very extensive time-series dataset that can be used for several purposes. First, Wichers et al. (2016) showed that the participant experienced a critical transition and that symptoms behaved conform principles of complex dynamical systems. Therefore, these data are extremely

suitable for researchers to validate new methods for predicting the onset of a critical transition. Second, there have been recent developments into estimating time-varying networks. These data can be used as an empirical example to show how time-varying networks can be estimated and how the network develops over time. Lastly, since items were measured at different time scales, this dataset can aid research that aims to combine (time-series) data from different time scales.

C

Supplementary Material to Chapter 5: Psychological Perturbation Data on Attitudes Towards the Consumption of Meat

Abstract

We present a dataset on participants' attitudes towards the consumption of meat (N = 30). Participants were presented with a baseline questionnaire entailing 11 statements. After a baseline measurement, we perturbed the participant's opinion on one of the 11 items, after which the participant completed the same questionnaire. By repeating this procedure for each of the 11 items, we measured to what extent the perturbation changed the participant's baseline score. In addition, we asked participants to draw the influence of a specific item onto the other items in a network format. The data are suitable for various purposes, like causal inference and the malleability of attitudes.

C.1 Overview

C.1.1 Collection Date(s)

The data were collected between June 8, 2017, and July 7, 2017.

C.1.2 Background

We present a dataset of 30 participants on their attitudes towards the consumption of meat. The complexity of the issues surrounding meat consumption in this era is extensive. The consumption of meat is

linked to several health problems such as heart and vascular disease (Walker et al., 2005), has a negative influence on the well-being of animals (Grandin, 2014), and is highly damaging for the environment (Steinfeld et al., 2006). Consumers should arguably consider the influence the consumption of meat has on their well-being, animal welfare and their environmental footprint.

The aim of the original study was to investigate to what extent causal influences between different components of attitudes toward the consumption of meat can be estimated. For this study we adapted a questionnaire developed by Dorresteijn (2017) *"Attitude Towards the Consumption of Meat Questionnaire"*. Following the tripartite model of attitudes (Bagozzi et al., 1979), the questionnaire captures three important aspects of attitudes: cognition, affect and behaviour.

C.2 Methods

C.2.1 Sample

The sample consists of 30 participants recruited through the psychology lab at the University of Amsterdam from June 8, until July 7, 2017. Participants ranged in age from 19 to 57, with a median age of 20 ($SD = 9$ years). The majority of the participants were female (21). Of the 30 participants, 15 reported to perceive their dietary life style to be omnivores, 14 flexitarians (dietary life style in which individuals choose not to eat meat at least 3 days a week; Voedingscentrum, 2018) one participant reported to be vegetarian and no participants reported to perceive their dietary life style as vegan. Participation was either compensated with 10 euros per hour or research credit.

C.2.2 Materials

The dataset consists of data collected with an altered version of *"Attitude Towards the Consumption of Meat Questionnaire"*, developed by Dorresteijn (2017). The original questionnaire contained 22 items regarding affect (6), behaviour (10), cognition (6), and six demographic questions. We selected 11 items (6 cognition items and 5 affect items) that we felt were the easiest to perturb. Furthermore, we used 15 behaviour items and selected three demographic items that were asked at the end of

the experiment.

Note that we have more behaviour items than the original questionnaire, since we created separate items to investigate how many days during the week a participant eats cheese, dairy products or eggs. These three items were combined into one item in the original questionnaire.

Baseline questionnaire

The baseline questionnaire consists of 11 items regarding the participants' attitude towards the consumption of meat, which was measured on a 6-point Likert-scale ranging from 1 (completely disagree) to 6 (completely agree), four behaviour questions measured on a 7-point scale (days of the week), and three demographic questions, namely, age, gender, and perceived lifestyle (omnivore, flexitarian, vegetarian, vegan). The baseline questionnaire demonstrated internal consistency ($\alpha = 0.70$), assessed with Cronbach's alpha, using the omega() function in the R-package *psych* (version 1.8.4; Revelle, 2018). The omega hierarchical estimate, determined with the same function was 0.51. The questionnaire can be found in the supplementary file "questionnaire.pdf" at https://osf.io/8tm5f/.

Hypothetical scenarios

Each of the 11 questionnaire items corresponds to a hypothetical scenario. These scenarios were written with the intention to perturb the participant's initial response, i.e., to alter their answer on the baseline questionnaire, either positively or negatively. The goal of the hypothetical scenario was to perturb the participant's attitude in the opposite direction of their initial attitude regarding the particular statement. Participants were asked to identify themselves with the hypothetical scenario and to contemplate how this would change other aspects of their attitude towards the consumption of meat. After the participant was presented with a hypothetical scenario, the participant completed the questionnaire again (including the item that was perturbed), keeping the hypothetical scenario in mind. This procedure was repeated for each of the 11 items.

For example, after a participant answered the item *"The production of meat is harmful for the environment"* in the baseline questionnaire with *"completely dis-*

agree" (value = 1), they would receive a hypothetical scenario that would alter their initial statement positively (values 4–6). In this case the participant would receive the following hypothetical scenario: *"The meat and dairy industry has a huge CO-2 emission and is therefore harmful for the environment. How does this influence your attitude towards the consumption of meat?"* All the 22 hypothetical scenarios can be found in the supplementary file "scenarios.pdf" at https://osf.io/8tm5f/.

Adjacency matrix

When participants completed the questionnaire after an individual item was perturbed, participants drew their own network to visualise the effect of the statement on remaining items of the questionnaire. Participants received instructions and a blue print of a network containing 11 nodes (corresponding to the 11 items) and no edges. They were asked to draw an arrowhead line from node A to B if they thought node A had a causal influence on node B. Figure C.1 (upper panel) shows an example of an empty network as it was presented to the participants. The middle panel of Figure C.1 depicts a network of a participant who drew the effect of item 7 ("I like/do not like the taste of meat") on the remaining items of the questionnaire. In order to create one adjacency matrix for each participant, we added all 11 adjacency matrices (one for each of the statements). The lower panel of Figure C.1 shows a network that depicts all causal influences that a participant drew.

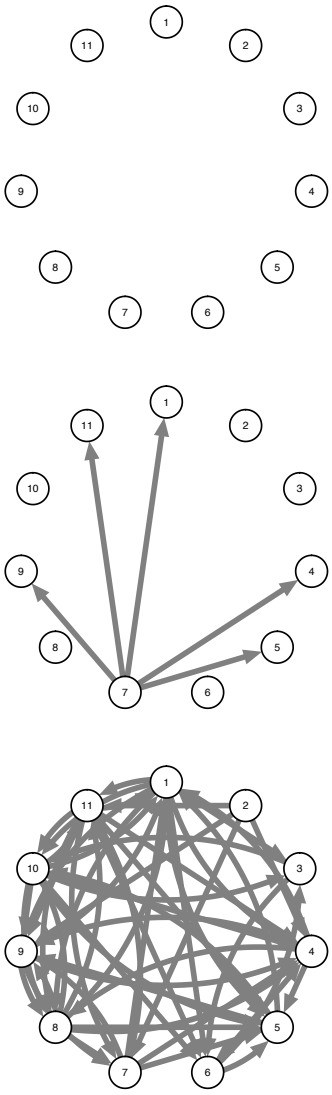


Figure C.1: The upper panel depicts an empty network that was presented to the participants. The middle panel depicts a network of a participant who drew the effect for item 7 (“I do/don’t like the taste of meat”). Note that in the experiment, participants drew the causal effects on paper. For printing purposes, this drawing was digitalised. The lower panel depicts all causal influences a participant drew.

C.2.3 Procedures

Participants first received a baseline questionnaire regarding their attitude towards the consumption of meat. Based upon these scores, participants received 11 hypothetical scenarios, which corresponded to the 11 items. After each scenario the participant completed the questionnaire again. In addition to the questionnaire that had to be filled out after each hypothetical scenario, participants were asked which items were influenced by the item that was perturbed and had to draw these causal influences in an empty network. The questionnaire items were randomised for each hypothetical scenario. We also randomised the order of the hypothetical scenario per participant.

C.2.4 Quality Control

The questionnaire was administered on paper. Hypothetical scenarios (positive/negative) were selected and administered based on the baseline measurement for each individual participant. The data were digitalised by Ria Hoekstra and checked by Jolanda Kossakowski.

C.2.5 Ethical Issues

This study was approved by the ethical review board of the University of Amsterdam. All participants signed an informed consent form before participating in the study. The data were anonymised before publication.

C.3 Dataset description

C.3.1 Object name

The datafile is named "data.zip". This zip file at https://osf.io/8tm5f/ contains the data "PerturbationData.xls", "PerturbationData.csv", "PerturbationData.txt", the individual adjacency matrices "adj1.txt – adj30.txt", a codebook "Codebook.pdf", a supplement holding the used hypothetical scenarios "scenarios.pdf", and a supplement containing the questionnaire "questionnaire.pdf".

C.3.2 Data type

All datafiles are primary data.

C.3.3 Format names and versions

The primary data are provided in three different formats: .xls, .csv, and .txt format.

The individual adjacency matrices are provided in one format: .txt. The accompanying codebook, the scenario supplement and the questionnaire are provided in .pdf format.

C.3.4 Data Collectors

Ria Hoekstra, Jolanda Kossakowski and Han van der Maas designed the entire study and the experiment. Ria Hoekstra was responsible for the actual data collection.

C.3.5 Language

English.

C.3.6 License

The data have been deposited under a CC-By Attribution 4.0 International (CC-By) License.

C.3.7 Embargo

Not applicable.

C.3.8 Repository location

http://osf.io/8tm5f.

C.3.9 Publication date

The data have been published online since February 21, 2018.

C.4 Reuse potential

This dataset contains data from 30 participants who completed the same questionnaire 12 times. We perturbed the participant's opinion on each of the 11 items and measured to what extent this changed the participant's scores on the questionnaire. The dataset also contains adjacency matrices for each individual, holding information about their perceived causal relations between questionnaire items. It is a unique dataset that can be used for several purposes. First, the questionnaire data can aid research that aims to infer causal relations between variables. Since the data contain both observational and experimental data, algorithms like the downward ranking of feed-forward loops (DR-FFL; Pinna et al., 2013), the invariant causal prediction (ICP; Meinshausen et al., 2016), or newly created algorithms can be used to investigate to what extent questionnaire items causally influence each other. Second, these data

can be used to study the malleability of a person's attitude towards meat consumption, a "hot topic" (Dalege et al., 2017). Third, these data are suitable for studies that look into attitude differences between participants that maintain different dietary lifestyles, like vegetarians and omnivores. Lastly, the adjacency matrices can aid research that aims to combine information from different participants on perceived causal relations between items into one set of perceived causal relations.

D

Supplementary Material to Chapter 5

D.1 D-separation in a DAG

Earlier we described the four different causal structures that can exist in a causal graph (see Figure 5.1). By determining the conditional (in)dependencies between sets of variables, the PC-algorithm estimates a DAG. The notion of conditional independencies can be extended to the idea of *directed (d) separation*. D-separation generalises a conditional independence relation between two variables. Say we have a path from variable X to variable Y. Variables X and Y are d-separated (denoted by $\mathrm{X} \perp\!\!\!\perp_{\mathrm{G}} \mathrm{Y} \mid \mathrm{Z}$) given a variable Z if Z blocks the path from any node in X to any node in Y. D-separation is relatively easy in the case of a chain or a common cause structure. With these specific structures, X and Y are d-separated given Z when Z is observed. When Z is observed, it "blocks" the path from X to Y. The reverse is true for the collider structure: X and Y are d-separated given Z as long as Z, or any of its descendants are not conditioned on. For a disjoint set of random variables X, Y and Z with joint probability distribution $\mathbb{P}$, we note that X is conditionally independent of Y given Z by $\mathrm{X} \perp\!\!\!\perp_{\mathbb{P}} \mathrm{Y} \mid \mathrm{Z}$. From this notion of d-separation, two assumptions follow:

Assumption 1: *We assume that for disjoint sets of variables X, Y and Z the* causal Markov condition *is satisfied, which specifies that*

$$X \perp\!\!\!\perp_G Y \mid Z \implies X \perp\!\!\!\perp_{\mathbb{P}} Y \mid Z \tag{D.1}$$

This assumption guarantees that, when we find that two variables are d-separated, these two variables are conditionally independent given a third variable.

Assumption 2: *We assume that for disjoint sets of variables X, Y and Z the* causal faithfulness condition *is satisfied, which specifies that*

$$X \perp\!\!\!\perp_{\mathbb{P}} Y \mid Z \implies X \perp\!\!\!\perp_G Y \mid Z \tag{D.2}$$

This assumption ensures that, when two variables are conditionally independent given a third, they are also d-separated given that third variable.

D.2 Theory of transitive reduction

Both the DR-FFL and the TRANSWESD algorithm use transitive reduction to estimate a causal graph. Both algorithms first draw up a perturbation graph in which causal relations between variables exist that exceed a pre-specified threshold. Often (but not always), the correlation between two variables is used. The idea here is that, when a correlation between two variables is nonzero, then there must be either a direct or an indirect relation between these two variables. Transitive reduction aims to remove direct effects where there should not be one, by considering alternative paths between two variables.

To illustrate transitive reduction, we have set up two examples, visualised in Figure D.1. Here, we consider the causal relation between variables W and Y. Wright (1921) showed that the correlation between W and Y is sum of the product of the path coefficients, denoted by β_{ij}. In the first example, shown in the left panel of Figure D.1, two paths exist from W to Y: $W \longrightarrow X \longrightarrow Y$ and $W \longrightarrow Z \longrightarrow Y$. The correlation ρ_{WY} then becomes $(0.20)(0.20) + (0.20)(0.20) = 0.08$. The following criterion is used to remove a direct effect from the perturbation graph:

$$\min\{|\rho_{WX_1}^{\{\emptyset,s\}}|, \ldots, |\rho_{X_kY}^{\{\emptyset,s\}}|\} > |\rho_{WY}^{\{\emptyset,s\}}|, \tag{D.3}$$

where $\{\emptyset, s\}$ denote the observational environment in which no perturbations have taken place ($\emptyset$), and the experimental environment in which perturbations have taken place on variable s. The variables $X_1 \ldots X_k$ denote the variables that lie on the path from W to Y. In other words, (D.3) states that if the smallest absolute path coefficient is larger than the direct effect between two variables, then the direct effect is to be removed from the perturbation graph. In our illustration, the correlation between W and Y ($\rho_{WY} = 0.08$) is smaller than the smallest path coefficient on either path (all path coefficients are 0.20), and thus there should not be a direct effect from W to Y. Therefore the left panel of Figure D.1 shows that transitive reduction is able to come to the right conclusion. However, the criterion in (D.3) is *necessary*, but it turns that that it is not *sufficient* to find the true

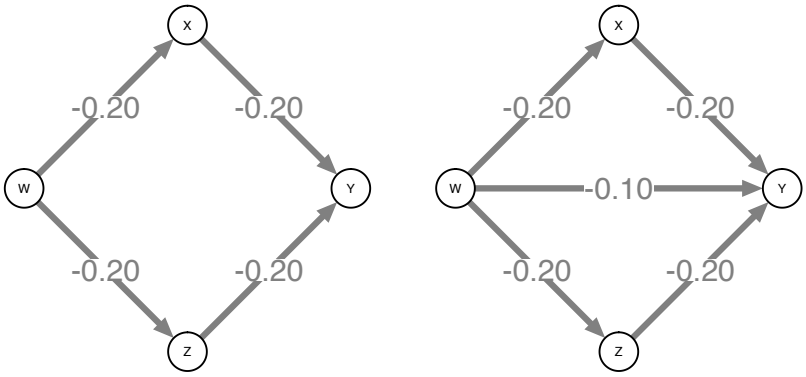


Figure D.1: Two examples of perturbation graphs one for which transitive reduction is appropriate (left panel) and one for which it is not (right panel).

causal graph. This is shown with the example in the right panel of Figure D.1. Here, there is a direct effect from W to Y. Now ρ_{WY} becomes $(0.20)(0.20)+(0.20)(0.20)-0.10=-0.02$, which is still smaller than the smallest path coefficient. Here, transitive reduction would erroneously remove the direct effect from W to Y. This shows that transitive reduction may not reach the correct causal graph, especially when the path coefficients are small. Specifically, the criterion in (D.3) will not work when the sum of the direct effect and ρ_{ij} is smaller than the smallest absolute path coefficient on any path between i and j.

D.3 Formal description of the HICP-algorithm

The HICP-algorithm controls for hidden variables by employing an instrumental variable Z. This instrumental variable cannot directly influence the target node Y, as shown in Figure 5.8. By using the instrumental variable Z, the regression of the target node onto the remaining variables will be split for the different time points, and the difference between these time points is used to estimate the causal effect. The causal effect from X to Y, denoted by $\hat{\alpha}$, is calculated as follows:

$$\hat{\alpha} = \frac{\text{cov}[X,Y]}{\text{var}[X]} = \alpha + \frac{\delta\gamma\,\text{var}[H]}{\text{var}[X]} \tag{D.4}$$

$$\begin{aligned}
\hat{\alpha} &= \frac{\text{cov}[X,Y]}{\text{var}[X]} \\
&= \frac{\text{cov}[X, \alpha\beta Z + (\alpha\gamma+\delta)H]}{\text{var}[X]} \\
&= \frac{\alpha\beta\ \text{cov}[X,Z] + (\alpha\gamma+\delta)\ \text{cov}[X,H]}{\text{var}[X]} \\
&= \frac{\alpha\beta\ \text{cov}[\beta Z + \gamma H, Z] + (\alpha\gamma+\delta)\ \text{cov}[\beta Z + \gamma H, H]}{\text{var}[X]} \\
&= \frac{\alpha\beta^2\ \text{var}[Z] + \gamma(\alpha\gamma+\delta)\ \text{var}[H]}{\hat{\beta}\ \text{var}[Z] + \gamma^2\ \text{var}[H] + \text{var}[N_x]} \\
&= \frac{\alpha\beta^2 + \gamma^2\alpha + \delta\gamma}{\beta^2 + \gamma^2 + \sigma_x^2} \\
&= \frac{\alpha(\beta^2+\gamma^2) + \delta\gamma}{(\beta^2+\gamma^2)} \\
&= \alpha + \frac{\delta\gamma}{\text{var}[X]},
\end{aligned} \tag{D.5}$$

where α, β, δ and γ represent relations between X, Y, H and Z. See Figure 5.8 for a visual representation. We can rewrite equation (D.5) as follows:

$$\begin{aligned}\hat{\alpha} &= \frac{\text{cov}[X,Y]}{\text{var}[X]} \\ &= \frac{\text{cov}[X,\alpha X]+\text{cov}[X,\delta H]}{\text{var}[X]} \\ &= \frac{\alpha\ \text{var}[X]+\delta\gamma\ \text{var}[H]}{\text{var}[X]} \\ &= \alpha+\frac{\delta\gamma\ \text{var}[H]}{\text{var}[X]}\end{aligned} \tag{D.6}$$

where the term $\frac{\delta\gamma\ \text{var}[H]}{\text{var}[X]}$ will be 0 when there are no hidden variables. In order to estimate $\hat{\alpha}$, the HICP-algorithm employs a two-step procedure. It first estimates $\hat{\beta}$ (the effect from the instrumental variable Z to variable X), after which $\hat{\beta}$ is used to estimate $\hat{\alpha}$:

$$\begin{aligned}\hat{\alpha} &= \frac{\text{cov}[\hat{\beta}Z,Y]}{\beta^2\ \text{var}[Z]} \\ &= \frac{\hat{\beta}\ \text{cov}[Z,\hat{\beta}Z+e]}{\beta^2\ \text{var}[Z]} \\ &= \frac{\hat{\beta}^2\ \text{var}[Z]}{\beta^2\ \text{var}[Z]}\end{aligned} \tag{D.7}$$

Note that the equations here are employed after the target node is selected. The computational steps that are taken to estimate all the causal effects are described below for a target node. We programmed a wrapper function that repeats these steps for every variable in the data. The HICP-algorithm uses the instrumental variable to divide the data into two subsets. The first subset contains data from the first environment (often that part of the data in which no perturbation has taken place). The second subset consists of all the remaining data. We can rewrite equation D.5 to make it computationally appropriate:

$$\begin{aligned}\hat{\alpha} &= (X'X)^{-1}X'y \\ &= \left[\frac{X_1'X_1}{n_1}-\frac{X_2'X_2}{n_2}\right]^{-1}\cdot\left[\frac{X_1'Y_1}{n_1}-\frac{X_2'Y_2}{n_2}\right] \\ &= \frac{\text{cov}[X_1,Y]-\text{cov}[X_2,Y]}{\text{var}[X_1]-\text{var}[X_2]},\end{aligned} \tag{D.8}$$

where X_1 and X_2 represent the predictor variables for the two environments, and Y_1 and Y_2 denote the scores on the target node for the two environments. The parameters n_1 and n_2 denote the number of participants that exist in the two environments. The result of equation (D.8) is a $p \times 1$ matrix that holds all the regression coefficients from every remaining node to the target node. After calculating $\hat{\alpha}$, we proceed with the calculation of Z-values for all participants per environment:

$$Z_{i,\varepsilon} = -\tau \cdot \sum_{p=1}^{P} \tau \left[\frac{X_1' Y_1}{n_1} - \frac{X_2' Y_2}{n_2} \right] + Y_{i,\varepsilon} \tau, \tag{D.9}$$

where $\tau = \left[\frac{X_1' X_1}{n_1} - \frac{X_2' X_2}{n_2} \right]^{-1} \cdot X_{i,\varepsilon}$. The parameter τ is created for each participant i and environment ε individually. The matrix $X_{i,\varepsilon}$ is a $1 \times p$ vector that holds the observational data for participant i and p variables. Two separate $n \times p$ matrices emerge from this equation: one for the first environment, and one for the second environment. The next step includes the calculation of σ:

$$\sigma = \sqrt{\text{diag}\left(\frac{s^2(Z_1)}{n_1} + \frac{s^2(Z_2)}{n_2} \right)} \tag{D.10}$$

where $s^2(Z_1)$ and $s^2(Z_2)$ denote the covariance matrix of the Z-values that we calculated previously in (D.9) for environments 1 and 2, and n_1 and n_2 the number of participants in the first and second environment. The term *diag* here indicates that we only take the diagonal of the result of $\frac{s^2(Z_1)}{n_1} + \frac{s^2(Z_2)}{n_2}$. In the last step we calculate the p-values associated with $\hat{\alpha}$. These are calculated in the following manner:

$$p = \max \begin{cases} 2K \cdot 1 - t(|\hat{\beta}| / \max(10^{-10}, \sigma)) \\ 1 \end{cases} \tag{D.11}$$

where K is the number of environments (2 in this study). The parameter $t()$ denotes the critical value in a t-distribution for a value of $\frac{|\hat{\beta}|}{\max(10^{-10}, \sigma)}$, with degrees of freedom $n-1$ (the total sample size). In order to estimate the maximal effect for each variable, we first determine the Z-value:

$$Z = \text{qnorm}\left[\max(0.5, 1-\alpha/(2K))\right]\sigma \tag{D.12}$$

which is then used in combination with $\hat{\alpha}$ to calculate the maximimal effect:

$$\theta = \text{sign}(\hat{\beta}) \cdot \max(0, |\hat{\beta}| - Z) \tag{D.13}$$

The maximal effects for insignificant variables is set to be 0 due to the max term that exists in θ.

D.4 The Creation of a Directed Acyclic Graph

When constructing a *Directed Acyclic Graph* (DAG), one needs to specify the number of nodes in the graph p, and the proportion of edges that should be present in the graph d, the density of the graph. In a graph $p^2 - p$ edges are possible (self-loops are not possible), and therefore the number of edges that will be drawn in the graph is $d \cdot \left(p^2 - p\right) = e$. We added a constraint to the number of edges that states that e cannot be lower than $p - 1$. We added this constraint to ensure that all nodes in the graph have at least one connection.

We start out with a $p \times p$ adjacency matrix that consists of solely 0s. Then, we randomly select e cells and set them to 1s: a 0 indicates the absence of an edge, and a 1 the presence of an edge. The diagonal of the matrix is always 0, as self-loops are not permitted at this point in time. To illustrate, say one wants to create a DAG consisting of $p = 5$ nodes, with $d = 0.25$, which translates to $e = 5$. This process may result in an adjacency matrix like this:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

After this initiation process, the DAG, which we call G, is checked to see if it satisfies the following conditions: (1) the number of edges in G must equal e, (2) each node in G must have at least one connection, (3) all edge weights in G must equal 1, (4) all nodes in G must be connected in one component and (5), G cannot contain any cycles. This process may result in the following adjacency matrix:

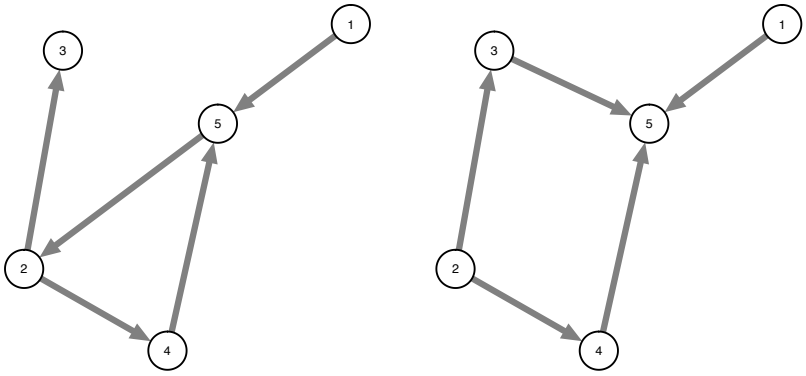


Figure D.2: Visualisation of the DAG simulation. The left panel depicts a graph after the initiation process, and the right panel depicts a graph after it satisfies all conditions.

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Figure D.2 shows the visualisation of the graph G after the initiation process (left panel) and after all criteria are satisfied (right panel). After the DAG is finalised, its adjacency matrix is ordered so that all nonzero elements are in the lower-diagonal part of the matrix. This process does not alter the DAG itself, solely its representation. This ordering process is needed to properly simulate data, which is discussed in section 5.3.

D.5 Numerical evaluation of causal inference algorithms with hidden variables

Next to the simulation study reported in section 5.4, we ran a second simulation study using data that contained hidden variables. Figure D.3 shows the MCC for the five algorithms that we investigated. The results are very similar to the results using data without hidden variables. For the PC-algorithm, we again see that the graph density d influences the MCC, where it reaches the highest numbers when $d = 0.25$. This effect only appears when the graph size $p = 5$. With $p = 10$, the MCC is generally low (average MCC = 0.09) at $d = 0.5$ and will only increase to mediocre (average MCC = 0.56) values when $d = 0.1$.

The picture we painted in section 5.4 for the DR-FFL and the TRANSWESD algorithms does not improve when hidden variables are included. When the graph size $p = 5$, the average MCC lies around 0.15, whereas when $p = 10$, the average MCC is around 0.03. The sample size n does not seem to influence the performance of both algorithms. On the other hand, the threshold parameter β has a big impact. The lower β, the higher the MCC is. To illustrate, when $\beta = 0.5$, the average MCC is 0.50, whereas when $\beta = 2.58$, the average MCC is close to zero. The threshold parameter β influences how many edges are retained after the first step in both the DR-FFL and the TRANSWESD algorithm. The higher the threshold, the lower the number of edges that are present in the perturbation graph, and thus the lower the MCC.

The ICP-algorithm has the best performance when there is a medium number of edges in the graph ($d = 0.25$). As we saw before, we observe high MCC values with the smaller graphs ($p = 5$). When p is increased to 10, the ICP-algorithm becomes more conservative, resulting in a lower MCC. Only when the mean of the perturbation distribution ($\overline{m}$) is high and the standard deviation is small can the ICP-algorithm accurately estimate causal graphs. This indicates that the ICP-algorithm needs a strong and effective perturbation in order to correctly identify causal relations.

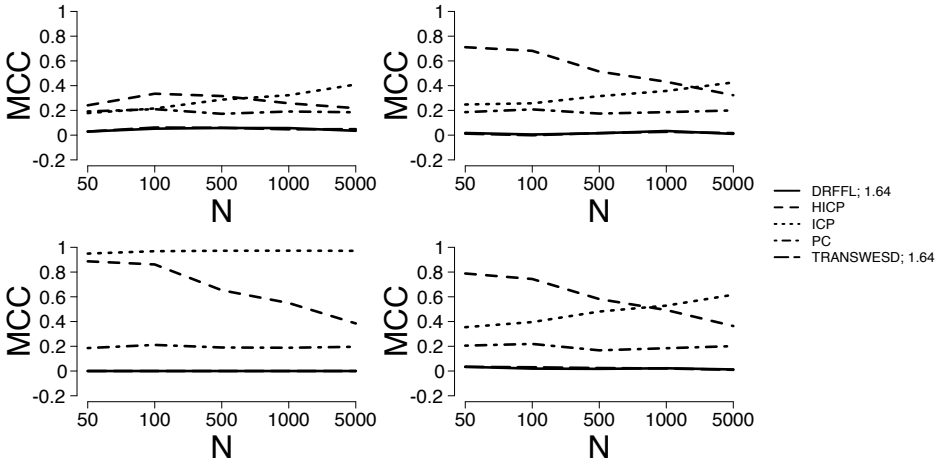


Figure D.3: Matthew's correlation coefficient (MCC) for $p = 10$ nodes with a network density of $d = 0.25$ with the addition of hidden variables. Top left = $\overline{m} = 1$, $sd = 0.5$, top right = $\overline{m} = 1$, $sd = 5$, bottom left = $\overline{m} = 5$, $sd = 0.5$, bottom right = $\overline{m} = 5$, $sd = 5$.

The mixed performance that we saw earlier with respect to the HICP-algorithm is also present when we add hidden variables to the data. This means that the HICP-algorithm can accurately estimate causal graphs with a small sample size. When the sample size increases, the accuracy decreases. This effect is present in almost every simulation condition. The only exception is when the graph density is low ($d = 0.1$). In that case, the MCC increases when the sample size increases.

Figure D.4 paints a similar picture that we saw in section 5.4. The lack of accuracy of the DR-FFL and the TRANSWESD algorithm is clearly visible, as are the spurious edges that are estimated by the HICP with a large sample size. Even though hidden variables are added to these data, the ICP-algorithm shows the highest number of true positives, combined with the lowest number of false positives for this simulation condition. Lastly, the PC-algorithm can have issues with determining the direction of an edge. This problem emerges independent of the presence of hidden variables, as we have seen this in section 5.4 as well. All in all, these results may suggest that the ICP-algorithm is the saver option when one wants to estimate a causal graph.

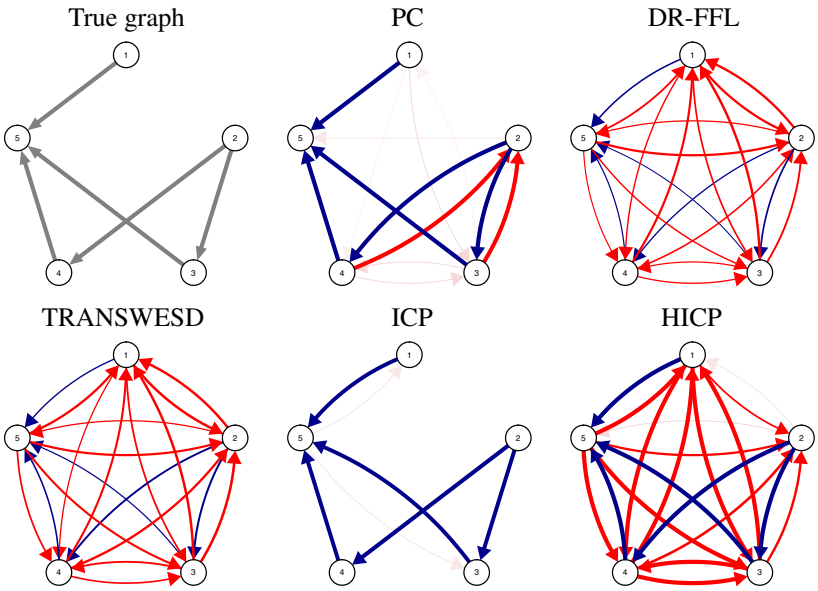


Figure D.4: Visualisation of the number of true positives and false positives for $p = 5$, $d = 0.25$, $n = 5000$, $\overline{m} = 5$, $sd = 0.5$ and $\beta = 0.5$ with the addition of hidden variables. Blue edges indicate true positives, and red edges indicate false negatives. The saturation and thickness of the edge represents how often that edge was (in)correctly estimated. Upper left = true graph, upper middle = PC, upper right = DR-FFL, lower left = TRANSWESD, lower middle = ICP, lower right = HICP.

Bibliography

Alligood, K. T., Sauer, T. D., & Yorke, J. A. (1996). *Chaos: An introduction to dynamical systems*. New York, USA: Springer.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596.

Bagozzi, R. P., Tybout, A. M., Craig, C. S., & Sternthal, B. (1979). The construct validity of the tripartite classification of attitudes. *Journal of Marketing Research*, *16*, 88–95.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*, 754–94.

Balister, P., Bollobás, B., & Kozma, R. (2006). Large deviations for mean field models of probabilistic cellular automata. *Random Structures & Algorithms*, *29*, 399–415.

Barrat, A., Barthelemy, M., & Vespignani, A. (2008). *Dynamical processes on complex networks*. Cambridge, UK: Cambridge University Press.

Blakey, S. M., Abramowitz, J., Reuman, L., Leonard, R. C., & Riemann, B. C. (2017). Anxiety sensitivity as a predictor of outcome in the treatment of obsessive-compulsive disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *57*, 113–117.

Blanken, T., van der Zweerde, T., van Straten, A., van Someren, E., & Borsboom, D. (2019). Introducing network intervention analysis to investigate sequential, symptom-specific treatment effects: A demonstration in co-occurring insomnia and depression. *Psychotherapy and Psychosomatics*, *88*, 52–54.

Bollobás, B. (2001). *Random graphs*. Cambridge, UK: Cambridge University Press.

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, *64*, 1089–1108.

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*, 5–13.

Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*, 91–121.

Borsboom, D., Epskamp, S., Kievit, R. A., Cramer, A. O. J., & Schmittmann, V. D. (2011). Transdiagnostic networks: Commentary on Nolen-Hoeksema and Watkins (2011). *Perspectives on Psychological Science*, *6*, 610–614.

Borsboom, D., Rhemtulla, M., Cramer, A. O. J., van der Maas, H. L. J., Scheffer, M., & Dolan, C. V. (2016). Kinds versus continua: a review of psychometric approaches to uncover the structure of psychiatric constructs. *Psychological Medicine*, *46*, 1567–1579.

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., ... Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS ONE*, *8*, e60188.

Callaway, D. S., Newman, M. E. J., Strogatz, S. H., & Watts, D. J. (2000). Network robustness and fragility: percolation on random graphs. *Physical Review Letters*, *85*, 5468–5471.

Carpenter, S. R., & Brock, W. A. (2006). Rising variance: a leading indicator of ecological transition. *Ecology Letters*, *9*, 311–318.

Chen, J., & Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, *95*, 759–771.

Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mottus, 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.

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 patterns of depressive symptom inter-correlations. *Psychological Medicine*, *42*, 957–965.

Cramer, A. O. J., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L. J., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamical system. *PloS ONE*, *11*, e0167490.

Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., ... Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality*, *26*, 414–431.

Cramer, A. O. J., Waldorp, L. J., van der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: a network perspective. *Behavioral and Brain Sciences*, *33*, 137–193.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *The Journal of Nervous and Mental Disease*, *175*, 526–536.

Dakos, V., Scheffer, M., van Nes, E. H., Brovkin, V., Petoukhov, V., & Held, H. (2008). Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences USA*, *105*, 14308–14312.

Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas, H. L. J. (2016). Toward a Formalized Account of Attitudes: The Causal Attitude Network (CAN) Model. *Psychological Review*, *123*, 2–22.

Dalege, J., Borsboom, D., van Harreveld, F., Waldorp, L. J., & van der Maas, H. L. J. (2017). Network structure explains the impact of attitudes on voting decisions. *Scientific reports*, *7*, 4909. doi: https://doi.org/10.1038/s41598-017-05048-y

Derogatis, L., Rickels, K., & Rock, A. (1976). The SCL-90 and the MMPI: a step in the validation of a new self-report scale. *British Journal of Psychiatry*, *128*, 280–289.

Dolan, C. V., Oort, F. J., Stoel, R. D., & Wicherts, J. M. (2009). Testing measurement invariance in the target rotated multigroup exploratory factor model. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 295–314. doi: 10.1080/10705510902751416

Dorresteijn, L. (2017). *Attitudes towards meat consumption; A network perspective.* Unpublished Master's Thesis. Retrieved from `https://scripties.uba.uva.nl/search?id=639037`

Drton, M., & Richardson, T. (2004). Iterative conditional fitting for Gaussian ancestral graph models. In *Proceedings of the 20th conference on uncertainty in artificial intelligence* (pp. 130–137).

Durett, R. (2007). *Random graph dynamics*. Cambridge, UK: Cambridge University Press.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*, 155–174.

Emde, R. M., & Harmon, R. J. (1984). *Continuities and discontinuities in development*. New York, USA: Plenum Press.

Epskamp, S. (2015). *IsingSampler: Sampling methods and distribution functions for the Ising model.* Retrieved from `https://cran.r-project.org/package=IsingSampler`

Epskamp, S. (2016). *Regularized Gaussian psychological networks: Brief report on the performance of extended BIC model selection.* Submitted for publication. Retrieved from `https://pure.uva.nl/ws/files/26756857/1606.05771v1.pdf`

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: a tutorial paper. *Behavior Research Methods*, *50*, 195–212.

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.

Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, *23*, 617–634.

Epskamp, S., Waldorp, L. J., Mottus, R., & Borsboom, D. (2018). The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research*, *53*, 453–480.

Fleming, T. R., & Harrington, D. P. (1978). Estimation for discrete time nonhomogeneous Markov chains. *Stochastic Processes and their Applications*, *7*, 131–139.

Fontenelle, L. F., Mendlowicz, M. V., & Versiani, M. (2006). The descriptive epidemiology of obsessive–compulsive disorder. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *30*, 327–337.

Fort, H., Mazzeo, N., Scheffer, M., & van Nes, E. H. (2010). Catastrophic shifts in ecosystems: spatial early warnings and management procedures. *Journal of Physics Conference Series*, *246*, 1–15.

Foygel, R., & Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. In *Nips* (pp. 604–612).

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.

Fried, E. I., & Cramer, A. O. J. (2017). Moving forward: challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science*, *12*, 999–1020.

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*, 1–10.

Gardner, M. (1970). Mathematical games: The fantastic combinations of John Conway's new solitaire game "life". *Scientific American*, *223*, 120–123.

Gillan, C. M., & Sahakian, B. J. (2015). Which is the driver, the obsessions or the compulsions, in OCD? *Neuropsychopharmacology*, *40*, 247–248.

Glymour, C., & Scheines, R. (1986). Causal modeling with the TETRAD program. *Synthese*, *68*, 37–63.

Golubitsky, M., & Stewart, I. (2003). *The symmetry perspective: from equilibrium to chaos in phase space and physical space*. Basel, Switzerland: Birkhäuser.

Goodman, W. K., Price, L. K., Rasmussen, S. A., Mazure, C., Delgado, P., Heninger, G. R., & Charney, D. S. (1989). The Yale-Brown Obsessive Compulsive Scale. II. Validity. *Archives of General Psychiatry*, *46*, 1012–1016.

Goodman, W. K., Price, L. K., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., ... Charney, D. S. (1989). The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. *Archives of General Psychiatry*, *46*, 1006–1011.

Goodman, W. K., Rudorfer, M. V., & Maser, J. D. (2000). *Obsessive-compulsive disorder: Contemporary issues in treatment* (1st ed.). London, UK: Routledge.

Gordijn, M., Beersma, D., Bouhuys, A., Reinink, E., & van den Hoofdakker, R. (1994). A longitudinal study of diurnal mood variation in depression: characteristics and significance. *Journal of Affective Disorders*, *31*, 261–273.

Gordijn, M., Beersma, D., Bouhuys, A., & van den Hoofdakker, R. (1998). Mood variability and sleep deprivation effect as predictors of therapeutic response in depression. *Sleep-wake research in the Netherlands*, *9*, 41–44.

Grandin, T. (2014). Animal welfare and society concerns finding the missing link. *Meat Science*, *98*, 461–469.

Granger, C. W. J. (1980). Testing for causality. *Journal of Economic Dynamics and Control*, *2*, 329–352.

Groot, P. C. (2010). Patients can diagnose too: how continuous self-assessment aids diagnosis of, and recovery from, depression. *Journal of Mental Health*, *19*, 352–362.

Guloksuz, S., Pries, L. K., & Van Os, J. (2017). Application of network methods for understanding mental disorders: pitfalls and promise. *Psychological medicine*, *47*, 2743–2752.

Gulyás, L., Kampis, G., & Legendi, R. O. (2013). Elementary models of dynamic networks. *The European Physical Journal Special Topics*, *222*, 1311–1333.

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.

Hartman, C. A., Hox, J., Mellenbergh, G. J., Boyle, M. H., Offord, D. R., Racine, Y., ... Sergeant, J. A. (2001). DSM-IV internal construct validity: When a taxonomy meets data. *Journal of Child Psychology and Psychiatry*, *42*, 817–836.

Haslbeck, J. M. B., & Waldorp, L. J. (2016a). mgm: Structure Estimation for time-varying mixed graphical models in high-dimensional Data. *Submitted for publication*. Retrieved from `https://arxiv.org/pdf/1510.06871.pdf`

Haslbeck, J. M. B., & Waldorp, L. J. (2016b). Structure estimation for mixed graphical models in high-dimensional data. *Submitted for publication*. Retrieved from `https://arxiv.org/pdf/1510.05677.pdf`

Hasselblatt, B., & Katok, A. (2003). *A first course in dynamics*. Cambridge, UK: Cambridge University Press.

Hezel, D. M., & McNally, R. J. (2016). A theoretical review of cognitive biases and deficits in obsessive–compulsive disorder. *Biological Psychiatry*, *121*, 221 – 232.

Hirsch, M. W., Smale, S., & Devaney, R. L. (2004). *Differential equations, dynamical systems, and an introduction to chaos*. Oxford, UK: Academic press.

Hoekstra, R. H. A., Kossakowski, J. J., & Van der Maas, H. L. J. (2018). Psychological perturbation data on attitudes towards the consumption of meat. *Journal of Open Psychology Data*, *6*, 1–3.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945–960.

Holmgren, R. (1996). *A first course in discrete dynamical systems*. New York, USA: Springer Science & Business Media.

Hosenfeld, B., Bos, E. H., Wardenaar, K. J., Conradi, H. J., van der Maas, H. L. J., Visser, I., & de Jonge, P. (2015). Major depressive disorder as a nonlinear dynamic system: bimodality in the frequency distribution of depressive symptoms over time. *BMC Psychiatry*, *15*, 1–9.

Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei*, *31*, 253–258.

Jacoby, R. J., Leonard, R. C., Riemann, B. C., & Abramowitz, J. (2016). Self-punishment as a maladaptive thought control strategy mediates the relationship between beliefs about thoughts and repugnant obsessions. *Cognitive Therapy Research*, *40*, 179–187.

Janeck, A. S., Calamari, J. E., Riemann, B. C., & Heffelfinger, S. K. (2003). Too much thinking about thinking? Metacognitive differences in obsessive-compulsive disorder. *Anxiety Disorders*, *17*, 181–195.

Janson, S., Kozma, R., Ruszinko, M., & Sokolov, Y. (2019). A modified bootstrap percolation on a random graph coupled with a lattice. *Discrete Applied Mathematics*, *258*, 152–165.

Janson, S., Luczak, T., Turova, T., & Vallier, T. (2012). Bootstrap percolation on the random graph $G_{n,p}$. *The Annals of Applied Probability*, *22*, 1989–2047.

Janssens, K. A. M., Bos, E. H., Rosmalen, J. G. M., Wichers, M., & Riese, H. (2018). A qualitative approach to guide choices for designing a diary study. *BMC Medical Research Methodology*, *18*, 1 – 12.

Jones, P. J., Mair, P., Riemann, B. C., Mugno, B. L., & McNally, R. J. (2018). A network perspective on comorbid depression in adolescents with obsessive-compulsive disorder. *Journal of Anxiety Disorders*, *53*, 1–8.

Kalisch, M., & Buhlmann, P. (2007). Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *Journal of Machine Learning Research*, *8*, 613–636.

Kalisch, M., Mächler, M., Colombo, D., Maathuis, M. H., & Bühlmann, P. (2012). Causal inference using graphical models with the R package pcalg. *Journal of Statistical Software*, *47*, 1–26. Retrieved from `http://www.jstatsoft.org/v47/i11/`

Kessler, R. C., Adler, L. A., Berglund, P., Green, J., McLaughlin, K. A., Fayyad, J., ... Zaslavsky, A. M. (2014). The effects of temporally secondary co-morbid mental disorders on the associations of DSM-IV ADHD with adverse outcomes in the US

National Comorbidity Survey Replication Adolescent Supplement (NCS-A). *Psychological Medicine*, *44*, 1779–1792.

Kessler, R. C., Berglund, P., Chiu, W. T., Demler, O., Heeringa, S., Hiripi, E., ... Zheng, H. (2004). The US national comorbidity survey replication (NCS-R): Design and field procedures. *International Journal of Methods in Psychiatric Research*, *13*, 69–92.

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*, 593–602.

Kessler, R. C., & Ustun, T. B. (2004). The World Mental Health (WMH) survey initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research*, *13*, 93–121.

Kindermann, R., & Snell, J. L. (1980). *Markov random fields and their applications*. Providence, USA: American Mathematical Society.

Klamt, S., Flassig, R. J., & Sundmacher, K. (2010). TRANSWESD: Inferring cellular networks with transitive reduction. *Bioinformatics*, *26*, 2160–2168.

Kleczkowski, A., & Grenfell, B. T. (1999). Mean-field-type equations for spread of epidemics: The 'small world' model. *Physica A: Statistical Mechanics and its Applications*, *274*(1), 355–360.

Kossakowski, J. J. (2019). *Results from "Applying a dynamical systems nodel and network theory to major depressive disorder".* Retrieved from https://osf.io/edyzp/.

Kossakowski, J. J., & Cramer, A. O. J. (2019). Complexity, chaos and catastrophe: Modeling psychopathology as a dynamic system. In M. S. Vitevitch (Ed.), *Network science in cognitive psychology* (p. 45-79). New York, USA: Routledge.

Kossakowski, J. J., Epskamp, S., Kieffer, J. M., van Borkulo, C. D., Rhemtulla, M., & Borsboom, D. (2016). The application of a network approach to health-related quality of life (HRQoL): introducing a new method for assessing HRQoL in healthy adults and cancer patients. *Quality of Life Research*, *25*, 781–792.

Kossakowski, J. J., Gordijn, M. C. M., Riese, H., & Waldorp, L. J. (2019). Applying a dynamical systems model and network theory to major depressive disorder. *Frontiers in Psychology: Quantitative Psychology and Measurement*, *10*, 1762. doi: 10.3389/fpsyg.2019.01762

Kossakowski, J. J., Groot, P. C., Haslbeck, J. M. B., Borsboom, D., & Wichers, M. (2017). Data from 'Critical slowing down as a personalized early warning signal for depression'. *Journal of Open Psychology Data*, *5*, 1–3.

Kossakowski, J. J., & Waldorp, L. J. (2020). *Results from 'Mean field dynamics of stochastic cellular automata for random and small-world graphs'*. Retrieved from: http://osf.io/ewf2g.

Kossakowski, J. J., Waldorp, L. J., & van der Maas, H. L. J. (2020). The search for causality: A comparison of different techniques for causal inference graphs. *Submitted to Psychological Methods*.

Kozma, R., Puljic, M., Balister, P., Bollobás, B., & Freeman, W. J. (2004). Neuropercolation: a random cellular automata approach to spatio-temporal neurodynamics. *Conference on Cellular Automata*, 435–443.

Kozma, R., Puljic, M., Balister, P., Bollobás, B., & Freeman, W. J. (2005). Phase transitions in the neuropercolation model of neural populations with mixed local and non-local interactions. *Biological Cybernetics*, *92*, 367–379.

Kuznetsov, Y. A. (2013). *Elements of applied bifurcation theory*. New York, USA: Springer-Verlag.

Lebowitz, J. L., Maes, C., & Speer, E. R. (1990). Statistical mechanics of probabilistic cellular automata. *Journal of Statistical Physics*, *59*, 117–170.

Lesigne, E. (2005). Heads or tails: An introduction to limit theorems in probability. *Providence, USA: American Mathematical Society*.

Levin, D. A., Peres, Y., & Wimer, E. L. (2017). Markov chains and mixing times. *Providence, USA: American Mathematical Society*.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, USA: John Wiley & Sons.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states - comparison of the depression anxiety stress scale (DASS) with the beck depression and anxiety inventories. *Behaviour Research and Therapy*, *33*, 335–343.

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the depression anxiety stress scales* (2nd edition ed.). Sydney, Australia: Psychology Foundation.

Luchian, S. A., McNally, R. J., & Hooley, J. M. (2007). Cognitive aspects of nonclinical obsessive–compulsive hoarding. *Behaviour Research and Therapy*, *45*, 1657–1662.

Manos, R. C., Cahill, S. P., Wetterneck, C. T., Conelea, C. A., Ross, A. R., & Riemann, B. C. (2010). The impact of experiential avoidance and obsessive beliefs on obsessive-compulsive symptoms in a severe clinical sample. *Journal of Anxiety Disorders*, *24*, 700–708.

Marker, C. D., Calamari, J. E., Woodard, J. L., & Riemann, B. C. (2006). Cognitive self-consciousness, implicit learning and obsessive–compulsive disorder. *Anxiety Disorders*, *20*, 389–407.

Mataix-Cols, D., Fernandez de la Cruz, L., Nordsletten, A. E., Lenhard, F., Isomura, K., & Blair Simpson, H. (2016). Towards an international expert consensus for defining

treatment response, remission, recovery and relapse in obsessive compulsive disorder. *World Psychiatry*, *15*, 80–81.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta - Protein Structure*, *405*, 442–451.

McNally, R. J., Mair, P., Mugno, B. L., & Riemann, B. C. (2017). Co-morbid obsessive–compulsive disorder and depression: A bayesian network approach. *Psychological Medicine*, 1204–1214.

McNally, R. J., & Ricciardi, J. N. (1996). Suppression of negative and neutral thoughts. *Behavioural and Cognitive Psychotherapy*, *24*, 17–25.

McNally, R. J., Robinaugh, D. J., Wu, G. W. Y., Wang, L., Deserno, M. K., & Borsboom, D. (2014). Mental disorders as causal systems: A network approach to posttraumatic stress disorder. *Clinical Psychological Science*, *3*, 836–849.

Meinshausen, N. (2018). Invariantcausalprediction: Invariant causal prediction [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=InvariantCausalPrediction` (R package version 0.7-2)

Meinshausen, N., Hauser, A., Mooij, J. M., Peters, J., Versteeg, P., & Buhlmann, P. (2016). Methods for causal inference from gene perturbation experiments and validation. *Proceedings of the National Academy of Sciences*, *113*, 7361–7368.

Molenaar, P. C. (2007). On the implications of classic ergodic theorems: Analysis of developmental processes has to focus on intra-individual variation. *Developmental Psychobiology*, *50*, 60–69.

Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J., & Scholkopf, B. (2016). Distinguishing cause from effect using observational data: Methods and benchmarks. *Journal of Machine Learning Research*, *17*, 1–102.

Mott, J., & Schneider, H. (1957). Matrix norms applied to weakly ergodic Markov chains. *Archiv der Mathematik*, *8*, 331–333.

Najmi, S., Riemann, B. C., & Wegner, D. M. (2009). Managing unwanted intrusive thoughts in obsessive–compulsive disorder: Relative effectiveness of suppression, focused distraction, and acceptance. *Behaviour Research and Therapy*, *47*, 494–503.

Newman, M. E. J. (2010). *Networks: an introduction*. Oxford, UK: Oxford University Press.

Newman, M. E. J., & Watts, D. J. (1999a). Renormalization group analysis of the small-world network model. *Physics Letters A*, *263*, 4–6.

Newman, M. E. J., & Watts, D. J. (1999b). Scaling and percolation in the small-world network model. *Physical Review E*, *60*, 7332.

Norris, J. R. (1997). *Markov chains*. Cambridge, UK: Cambridge University Press.

O'Donnell, R. (2014). *Analysis of boolean functions*. Cambridge, UK: Cambridge University Press.

Okasha, A. (2002). Diagnosis of obsessive-compulsive disorder: A review. In *Obsessive-compulsive disorder* (2nd ed., pp. 1–19). Chichester, UK: John Wiley & Sons Ltd.

Olde Rikkert, M. G. M., Dakos, V., Buchman, T. G., de Boer, R., Glass, L., Cramer, A. O. J., ... Scheffer, M. (2016). Slowing down of recovery as generic risk marker for acute severity transitions in chronic diseases. *Critical Care Medicine*, *44*, 601–606.

Owis, M. I., Abou-Zied, A. H., Youssef, A.-B. M., & Kadah, Y. M. (2002). Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification. *IEEE Transactions on Biomedical Engineering*, *49*, 733–726.

Paz, A. (1971). *Introduction to probabilistic automata*. New York, USA: Academic Press.

Pearl, J. (2009). *Causality: models, reasoning, and inference* (2nd ed.). Cambridge, UK: Cambridge University Press.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. New York, USA: Basic Books.

Pearl, J., & Verma, T. S. (1991). A theory of inferred causation. In J. Allen, R. Fikes, & E. Sandewall (Eds.), *Knowledge representation and reasoning: Proceedings of the second international conference* (pp. 441–452). New York, USA: Morgan Kaufmann Publishers.

Peeters, F., Ponds, R., & Vermeeren, M. (1996). Affectiviteit en zelfbeoordeling van depressie en angst. *Tijdschrift voor de Psychiatrie*, *38*, 240–250.

Peters, J., Buhlmann, P., & Meinshausen, N. (2016). *Causal inference using invariant prediction: identification and confidence intervals.* Retrieved from `https://arxiv.org/pdf/1501.01332.pdf`

Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference: Foundations and learning algorithms*. Cambridge, USA: MIT Press.

Philip Babcock Gove (Ed.). (2008). *Webster's third new international dictionary*. USA: Merriam-Webster.

Pinna, A., Heise, S., Flassig, R. J., de la Fuente, A., & Klamt, S. (2013). Reconstruction of large-scale regulatory networks based on perturbation graphs and transitive reduction: Improved methods and their evaluation. *BMC Systems Biology*, *7*, 1–19.

Pinna, A., Soranzo, N., & de la Fuente, A. (2010). From knockouts to networks: establising direct cause-effect relationships through graph analysis. *PloS ONE*, *5*, e12912.

Powers, D. M. W. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, *2*, 37–63.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from `https://www.r-project.org/`

Rachman, S. (1997). A cognitive theory of obsessions. *Behavior Research and Therapy*, *35*, 793–802.

Raes, F., Daems, K., Feldman, G., Johnson, S., & van Gucht, D. (2009). A psychometric evaluation of the Dutch version of the responses to positive affect questionnaire. *Psychologica Belgica*, *49*, 293–310.

Rajarshi, M. B. (2012). *Statistical inference for discrete time stochastic processes*. New Delhi, India: Springer.

Rendell, P. (2011). A universal Turing machine in Conway's game of life. In *International conference on high performance computing & simulation* (pp. 764–772).

Revelle, W. (2018). psych: Procedures for psychological, psychometric, and personality research [Computer software manual]. Evanston, Illinois. Retrieved from `https://CRAN.R-project.org/package=psych` (R package version 1.8.12)

Rice, J. J., Tu, Y., & Stolovitzky, G. (2005). Reconstructing biological networks using conditional correlation analysis. *Bioinformatics*, *21*, 765–773.

Ruscio, A. M., Stein, D. J., Chiu, W. T., & Kessler, R. C. (2010). The epidemiology of obsessive–compulsive disorder in the National Comorbidity Survey Replication. *Molecular Psychiatry*, 53–63.

Rush, A., Bernstein, L., Trivedi, M., Camody, T., Wisniewski, S., Mundt, J., ... Fava, M. (2006). An evaluation of the quick inventory of depressive symptomatology and the Hamilton rating scale for depression: a sequenced treatment alternatives to relieve depression trial report. *Biological Psychiatry*, *59*, 493–501.

Rush, A., Trivedi, M., Ibrahim, H., Camody, T., Arnow, B., Klein, D., ... Keller, M. (2003). The 16-item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*, *54*, 573–583.

Ruzzano, L., Borsboom, D., & Geurts, H. M. (2015). Repetitive behaviors in autism and obsessive–compulsive disorder: New perspectives from a network analysis. *Journal of Autism and Developmental Disorders*, *45*, 192–202.

Saloff-Coste, L., & Zuninga, J. (2009). Merging for time inhomogeneous finite Markov chains, Part I: Singular values and stability. *Electronic Journal of Probability*, *14*, 1456–1494.

Saloff-Coste, L., & Zuninga, J. (2010). Merging and stability for time inhomogeneous finite Markov chains. Retrieved from `https://arxiv.org/pdf/1004.2296`

Sarkar, P. (2000). A brief history of cellular automata. *ACM Computing Surveys (CSUR)*, *32*, 80–107.

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., ... Sugihara, G. (2014). Early-warning signals for critical transitions. *Nature*, *46*, 53–59.

Scheffer, M., Carpenter, S., Foley, J. A., Folke, C., & Walker, B. (2001). Catastrophic shifts in ecosystems. *Nature*, *413*, 591–596.

Serra, R., Villani, M., & Semeria, A. (2004). Genetic network models and statistical properties of gene expression data in knock-out experiments. *Journal of Theoretical Biology*, *227*, 149–157.

Sethna, J. P., Dahmen, K. A., & Perkovic, O. (2004). Random-field Ising models of hysteresis. Retrieved from `https://arxiv.org/pdf/cond-mat/0406320.pdf`

Simon, H. A. (1952). On the definition of the causal relation. *The Journal of Philosophy*, *49*, 517–528.

Sornette, D. (2002). Predictability of catastrophic events: Material rupture, earthquakes, turbulence, financial crashes, and human birth. *Proceedings of the National Academy of Sciences*, *99*, 2522–2529.

Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search*. Cambridge, USA: MIT Press.

Sporns, O., & Honey, C. J. (2006). Small worlds inside big brains. *Proceedings of the National Academy of Sciences*, *103*, 19219–19220.

Steinfeld, H., Gerber, P., Wassenaar, T. D., Castel, V., & de Haan, C. (2006). *Livestock's long shadow: Environmental issues and options*. Rome, Italy: United Nations - Food and Agriculture Organization.

Steketee, G., Frost, R., & Bogart, K. (1996). The Yale-Brown Obsessive Compulsive scale: Interview versus self-report. *Behavior Research and Therapy*, *34*, 675–684.

Strogatz, S. H. (1994). *Nonlinear dynamics and chaos with applications to physics, biology, chemistry and engineering*. Boulder, USA: Westview Press.

Sun, J., Rhemtulla, M., & Vazire, S. (2019). Eavesdropping on missing data: What are people doing when they miss experience sampling reports? Retrieved from `https://psyarxiv.com/5tcwd`

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*, 267–288.

Tio, P., Epskamp, S., Noordhof, A., & Borsboom, D. (2016). Mapping the manuals of madness: Comparing the ICD-10 and DSM-IV-TR using a network approach. *International Journal of Methods in Psychiatric Research*. doi: 10.1002/mpr.1503

Tolin, D. F., Abramowitz, J., Przeworski, A., & Foa, E. (2002). Thought suppression in obsessive-compulsive disorder. *Behaviour Research and Therapy*, *40*, 1255–1274.

Tomassini, M., Giacobini, M., & Darabos, C. (2005). Evolution and dynamics of small-world cellular automata. *Complex Systems*, *15*, 261–284.

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, *24*, 637–657.

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*, 1–10.

van Borkulo, C. D., 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*, 1219–1226.

van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., ... Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences*, *111*, 87–92.

van der Krieke, L., Jeronimus, B. F., Blaauw, F. J., Wanders, R. B. K., Emerencia, A. C., Schenk, H. M., ... de Jonge, P. (2015). HowNutsAreTheDutch (HoeGekIsNL): A crowdsourcing study of mental symptoms and strengths. *International Journal of Methods in Psychiatric Research*, *25*, 123–144.

van der Maas, H. L. J., Dolan, C. V., Grasman, R., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842–861.

Venkatesh, S. S. (2012). *The theory of probability: Explorations and applications*. Cambridge, UK: Cambridge University Press.

Voedingscentrum. (2018). *Meer dan de helft van de Nederlanders is 'flexitarier'*. Retrieved from `https://www.voedingnu.nl/voeding-en-gedrag/nieuws/2015/06/meer-dan-de-helft-van-de-nederlanders-is-flexitarier-10111022`

Voelkle, M. C., Oud, J. H. L., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*, 176–192.

Von Neumann, J. (1951). The general and logical theory of automata. *Cerebral mechanisms in behavior*, 1–41.

von Zerssen, D. (1986). Clinical self-rating scales (CSRS) of the Munich psychiatric information system (PSYCHIS München). In *Assessment of depression* (pp. 270–303). Berlin, Germany: Springer.

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). *Psychological Methods*, *17*, 228–243.

Waldorp, L. J., & Kossakowski, J. J. (2020). Mean field dynamics of stochastic cellular automata for random and small-world graphs. *Under review at Journal of Mathematical Psychology*.

Walker, P., Rhubart-Berg, P., McKenzie, S., Kelling, L., & Lawrence, R. S. (2005). Public health implications of meat production and consumption. *Public Health Nutrition*, *8*, 348–356.

Ware Jr, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care*, *30*, 473–483.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality Social Psychology*, *54*, 1063–1070.

Watts, D. J. (1999). *Small-worlds: From order to chaos*. Princeton, USA: Princeton University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440–442.

Wichers, M., Groot, P. C., Psychosystems, ESM Group, & ESW Group. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics*, *85*, 114–116.

Wichers, M., Simons, C., Kramer, I., Hartmann, J., Lothmann, C., Myin-Germeys, I., ... van Os, J. (2011). Momentary assessment as a tool to help patients with depression help themselves. *Acta Psychiatrica Scandinavica*, *124*, 262–272.

Wild, B., Eichler, M., Friederich, H.-C., Hartmann, M., Zipfel, S., & Herzog, W. (2010). A graphical vector autoregressive modelling approach to the analysis of electronic diary data. *BMC Medical Research Methodology*, *10*, 1–13.

Wit, E., van den Heuvel, E., & Romeijn, J.-W. (2012). 'All models are wrong...': an introduction to model uncertainty. *Statistica Neerlandica*, *66*, 217–236.

Wolfram, S. (1984a). Cellular automata as models of complexity. *Nature*, *311*, 419–424.

Wolfram, S. (1984b). Computation theory of cellular automata. *Communications in Mathematical Physics*, *96*, 15–57.

World Health Organization. (2012). *Depression, a hidden burden.* Retrieved from: http://www.who.int.

Wright, S. (1921). Correlation and causation. *Journal of agricultural research*, *20*, 557–585.

Zhao, P., & Yu, B. (2006). On model selection consistency of lasso. *The Journal of Machine Learning Research*, *7*, 2541–2563.

Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., ... Diener, E. (2019). From data to causes I: Building a general cross-lagged panel model (GCLM). *Organizational Research Methods*. doi: DOI: 10.1177/1094428119847278

Publications

Main publications

Chapter 2 is published as:

Kossakowski, J. J., & Cramer, A.O.J. (2019). Complexity, chaos and catastrophe: Modeling psychopathology as a dynamic System. In M.S. Vitevitch (Eds.) *Network science in cognitive psychology* (pp 45-79). New York, USA: Routledge.

JJK generated the idea for the book chapter. AOJC ran the analyses for section 2.4, and JJK for section 2.5. JJK and AOJC wrote the draft of the book chapter and AOJC provided critical revisions.

Chapter 3 is submitted as:

Waldorp, L. J., & **Kossakowski, J. J.** (2020). Mean field dynamics of stochastic cellular automata for random and small-world graphs. *Under review at Journal of Mathematical Psychology*.

LJW generated the idea of the study. JJK and LJW designed the simulation study. JJK programmed the simulation study and analyzed the data under supervision of LJW. LJW drafted the manuscript. JJK wrote section 3.5. Both authors provided critical revisions.

Chapter 4 is published as:

Kossakowski, J. J., Gordijn, M. C. M., Harriette, R., & Waldorp, L. J. (2019). Applying a dynamical systems model and network theory to major depressive disorder. *Frontiers in Psychology: Quantitative Psychology and Measurement, 10*, 1762.

JJK and LJW jointly generated the idea for the study. JJK programmed the study. MCMG collected the data of the clinical sample. JJK wrote the analysis code and analyzed the data, LJW verified the accuracy of those analyses. JJK wrote the first draft of the manuscript, and all authors critically edited it.

Chapter 5 is submitted as:

Kossakowski, J. J., Waldorp, L. W., & van der Maas, H. L. J. (2020). The search for causality: A comparison of different techniques for causal inference graphs. *Under review at Psychological Methods*.

JJK and LJW jointly generated the idea for the study. JJK. programmed the study. JJK analyzed the data under the supervision of LJW and HLJM. JJK drafted the manuscript and LJW and HLJM provided critical revisions.

Chapter 6 is submitted as:

Kossakowski, J.J., van Oudheusden, L. B, McNally, R. J., Waldorp, L. J., Riemann, B. C., & van der Maas, H. L. J. (2020). Introducing the causal graph approach to psychopathology: An illustration in patients with obsessive-compulsive disorder. *Under review at Clinical Psychological Science*.

JJK, RJM, LJW, and HLJM jointly generated the idea for the study. JJK. programmed the study. BCR was responsible for the data collection. JJK and LJBO performed the literature study. JJK analyzed the data under the supervision of RJM, LJW and HLJM.

JJK drafted the manuscript and LJBO, RJM, BCR, LJW, and HLJM provided critical revisions.

Appendix B is published as:

Kossakowski, J. J., Groot, P. C., Haslbeck, J. M. B., Borsboom, D., & Wichers, M. (2017). Data from 'Critical slowing down as a personalized early warning signal for depression'. *Journal of Open Psychology Data*, *5*: 1, DOI: https://doi.org/10.5334/jopd.29.

PCG and MW designed the entire study and the experiment. Frenk Peeters was involved as a psychiatrist in the design phase of the experiment, Claudia Simons was responsible for the ESM briefing and technical assistance regarding the use of the PsyMate to collect the data. JJK drafted the manuscript, and PCG, JMBH, DB and MW provided critical revisions.

Appendix C is published as:

Hoekstra, R. H. A., **Kossakowski, J. J.,** & van der Maas, H. L. J. (2018) Psychological perturbation data on attitudes towards the consumption of meat. *Journal of Open Psychology Data, 6*: 3, DOI: http://doi.org/10.5334/jopd.37.

RHAH, JJK and HLJM designed the entire study and the experiment. RHAH was responsible for the actual data collection. RHAH and JJK drafted the manuscript, and HLJM provided critical revisions.

Other publications

Robinaugh, D. J., Haslbeck, J. M., **Kossakowski, J. J.**, Millner, A. J., Fried, E. I., Waldorp, L. J., McNally, R. J., van Nes, E. H., Scheffer, M., Kendler, K. S., & Borsboom, D. (2020). Advancing the network theory of mental disorders: A computational model of panic disorder as a complex system. *In preparation*.

Van den Berg, J. W., **Kossakowski, J. J.**, Smid, W., Babchishin, K. M., Borsboom, D., Janssen, E., & Gijs, L. (2020). Interrelationship of dynamic risk factors in adult males who committed sexual offences: A replication study. *In preparation*.

Van Borkulo, C.D., Waldorp, L. J., Boschloo, L.,, **Kossakowski, J.J.**, Tio, P., Schoevers, R.A., & Borsboom, D. (2020). Comparing network structures on three aspects: A permutation test. *In preparation*.

Van Oudheusden, L. J. B., **Kossakowski, J. J.**, Van Oppen, P., Meynen, G., & Van Balkom, A. J. L.M. (2020). Symptom dynamics in obsessive-compulsive disorder: a network analysis. *In preparation*.

Vial, A., van der Put, C., Stams, G. J. J. M., **Kossakowski, J. J.**, & Assink, M. (2020). Exploring the interrelatedness of risk factors for child maltreatment: A network approach. *Submitted to Child Abuse and Neglect*.

Van den Berg, J. W., Smid W., **Kossakowski, J. J.**, van Beek, D., Borsboom, D., Janssen, E., & Gijs, L. (in press). Network analysis applied to dynamic risk factors of adult male sex offenders. *Clinical Psychological Science*.

Meier, M., **Kossakowski, J. J.,** Jones, P. J., Riemann, B. C., Kay, B., & McNally, R. J. (2019). Obsessive-compulsive symptoms in eating disorders: A network investigation.

International Journal of Eating Disorders, 1-10.

Barbalat, G., van den Bergh, D., & **Kossakowski, J. J.** (2019). Assessing outcomes in mental health services: from individual scores to symptom networks. *BMC Psychiatry, 19*, 1-9. DOI: https://doi.org/10.1186/s12888-019-2175-7.

Derks, K. P., Burger, J., van Doorn, J. D., **Kossakowski, J. J.**, Atticciati, L., Beitner, J., . . . Wagenmakers, E.- J. (2018). Network models to organize a dispersed literature: The case of misunderstanding analysis of covariance. *Journal of European Psychology Students, 9*, 48-57. DOI: http://doi.org/10.5334/jeps.458.

Rouquette, A., Pingault, J-P., Fried, E., Falissard, B., **Kossakowski, J.J.,** Orri, M., Cote, S., & Borsboom, D. (2018). Behavioral symptom network structure in girls and links to anxiety disorders in early adulthood. *JAMA Psychiatry, 75*, 1173-1181. DOI: 10.1001/jamapsychiatry.2018.2119.

Santos Jr, H., **Kossakowski, J. J.**, Beeber, L., Schwartz, T., & Fried, E. I. (2018). Longitudinal network structure of depression symptoms and self-efficacy in low-income mothers. *PLos ONE*, *13*, e0191675. DOI: 10.1371/journal.pone.0191675.

Cheshin, A., Heerdink, M. W., **Kossakowski, J. J.**, & van Kleef, G. A. (2016). Pitching emotions: The interpersonal effects of emotions in baseball. *Frontiers in Psychology, 7*, 1-12. DOI: 10.3389/fpsyg.2016.00178.

Kossakowski, J. J., Epskamp, S., Kieffer, J. M., van Borkulo, C. D., Rhemtulla, M., & Borsboom, D. (2016). The application of a network approach to health-related quality of life (HRQoL): introducing a new method for assessing HRQoL in healthy adults and cancer patients. *Quality of Life Research*, *25*, 781-792.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, *349* (6251). DOI: 10.1126/science.aac4716.

English summary

Until recently, it was believed that psychological phenomena like depression or general intelligence could be explained by a latent (unobserved) variable. In the past years, *network theory* has gained popularity. This theory claims that a latent variable is not needed to explain psychological phenomena. Instead, network theory argues that a construct like depression or general intelligence is a *network* that comprises reciprocal (potentially causal) relations between variables, which can be used to gain a better understanding of a psychological construct. In the example of depression, such a network can be seen as a *complex system*: a set of variables that influence each other in such a way that critical transitions from one stable state to the other are possible.

Complex systems are complex for more than one reason. The system as a whole can quickly grow to be too complex to study in its entirety, and relations between variables can produce *critical transitions*: a sudden jump from one stable state to another. Systems of only few variables and relations do not pose a problem, but the systems that are studied in psychology often consist of ten or more variables, with many potential edges that are estimated. It is thus vital to simplify these systems. In the first part of this dissertation, we reduced these complex systems by means of a *mean field approximation* (MFA), where we approach the complex system as a *stochastic cellular automaton* (SCA).

A cellular automaton (CA) is a discrete dynamical system that has a deterministic, local update rule to move through time. Every node in a CA lies in a finite grid and can either be 'active' (1) or 'inactive' (0). An update rule that is often used is the majority

rule. This rules states that, when the majority of a node's neighbours is active, then the node itself will become active at the next time point. While the deterministic majority rule can be efficient, it is not an accurate representation of psychological constructs, as there is always some uncertainty in human beings, their behaviour, and their emotions. Therefore we introduced a SCA, in which the majority still considers a node's neighbours, but instead of fixing the outcome based on that node's neighbours, it assigns a probability p for that node to become active at the next time point. By assuming that all nodes in a system have the same update rule and the same number of neighbours, we can reduce that system to a one-dimensional discrete Markov chain. Since we also assume that every node can be in either one of two states, the process of updating a node's state is Bernoulli distributed. This simplifies the system, enabling us to analyse the entire system more easily.

With the MFA, one only needs to determine the state of the nodes at $t = 0$ and the probability parameter p for the majority rule, and the system will continue to update. At a certain point in this process, the system will find either one or two equilibria, based on the probability parameter p. If a system has two equilibria, it can show critical transitions between the two; with only one equilibrium critical transitions are theoretically impossible. By varying this probability parameter, we can map the space where the system is in, and pin point where one theoretically would expect critical transitions.

In theory, when one knows the probability parameter p and the space of the system, one could determine whether a system can experience critical transitions or not. In psychopathology, we could indicate whether or not an individual is mathematically in a place where critical transitions can occur. In practice, we can estimate this probability parameter from the data using maximum likelihood estimation. In both a simulation study and an empirical analysis, we showed that this estimation process works well, and that we were able to infer whether patients diagnosed with depression or healthy participants could experience critical transitions.

Relations between variables are often reciprocal, indicating that two variables may

influence each other. Establishing a causal relation between two variables is an essential next step; learning the cause of a symptom like concentration problems can help to reduce symptoms like these. A causal relation can indicate many things. Here, we define a causal relation to be a relation $X \longrightarrow Y$ where, when we change the cause (X), we observe a change in the effect (Y). This also means that, when we do not observe a change in Y, no change in X has occurred.

The type of data that is often used to estimate causal relations are *observational data*. These are data in which no perturbations (manipulations or interventions) have taken place. However, we cannot unravel all causal relations with only observational data. As implied by our definition, we need to perturb variables and observe its effects in order to establish causality between variables. Therefore, we need *experimental data*. These are data where some perturbation has taken place. By combining observational and experimental data, we may have enough information to properly estimate causal relations. Observational data is used to establish a connection between two variables, and experimental data is needed to estimate whether any (or both) of these variables change as a result of perturbing the other.

In the second part of this dissertation, we put several algorithms for estimating causal relations to the test in a simulation study, and found that two algorithms - the invariant causal prediction (ICP) and the hidden ICP (HICP) algorithm - in particular hold the best cards for correctly estimating causal relations. These algorithms search for the common denominator: only those causal relations that exist in all subsets of the data are retained in the final causal graph. The difference between the algorithms lies in the accountability of *hidden variables*, where the ICP-algorithm does not control for them, and the HICP-algorithm does.

In an empirical study, we combined the results of these two algorithms with a literature study in a causal graph approach to create a causal graph of obsessive-compulsive disorder. Although we saw a discrepancy between the algorithms and the literature, we uncovered causal relations that otherwise may have been left unknown, and showed the

potential of such an approach for psychopathology.

In my discussion, I took a critical stance towards my research. What steps have been taken, what have we accomplished, what do we need to do to achieve the goals where we can predict (instead of assessing) if individuals may experience critical transitions, and where we gain a full understanding of a disorder in terms of its causal relations? Although the research presented in this dissertation demonstrate the first steps towards these goals, we are not there yet. I elaborate on specific limitations, that may pose to be limitations for other studies as well, such as the great number of assumptions that accompany models, or (currently unsolved) issues of dealing with missing data. Despite these limitations, both the mean field model and the causal graph approach can be extended and adapted in many ways. It may be even possible to combine the two into a complex causal system.

This dissertation showed that both prediction and explanation are important in psychological research. By putting a system passively or actively under pressure, we may assess where a system can go to, where we may gain a better understanding of the internal workings of a system.

Nederlandse samenvatting

Tot voor kort geloofde men dat psychologische fenomenen zoals depressie of algemene intelligentie verklaard konden worden door een latente (niet-geobserveerde) variabele. In de laatste jaren is de *netwerktheorie* in populariteit gestegen. Deze theorie beweert dat een latente variabele niet nodig is om psychologische fenomenen uit te leggen. In plaats daarvan beargumenteert de netwerktheorie dat een construct zoals depressie of algemene intelligentie een netwerk is dat wederkerige (mogelijk causale) relaties tussen variabelen bevat, welke gebruikt kunnen worden om een beter begrip te krijgen van een psychologisch construct. In het voorbeeld van depressie kan zo'n netwerk gezien worden als een *complex systeem*: een set variabelen die elkaar beïnvloeden op zo'n manier dat kritische transities van een stabiele staat naar een andere mogelijk zijn.

Complexe systemen zijn complex om meerdere redenen. Het systeem als geheel kan uitgroeien tot zo'n groot systeem dat het te complex is om in zijn geheel te bestuderen, en relaties tussen variabelen kunnen *kritische transities* produceren: een plotselinge sprong van één stabiele staat naar een andere. Systemen van slecht enkele variabelen en relaties vormen niet zo'n probleem. Echter, de systemen die vaak bestudeerd worden in de psychologie bevatten vaak tien of meer variabelen, met vele potentiële verbindingen die geschat worden. Het is dus van essentieel belang om deze systemen te simplificeren. In het eerste deel van dit proefschrift hebben we deze complexe systemen gereduceerd door middel van een *gemiddelde-veld-approximatie* (GVA), waarin we het complexe systeem benaderen als een *stochastisch cellulair automaat* (SCA).

Een cellulair automaat (CA) is een discreet dynamisch systeem met een lokale, deterministische actualisatieregel om zich voort te bewegen door de tijd. Elke knoop in een CA ligt op een eindig rooster en kan 'actief' (1) of 'inactief' (0) zijn. Een actualisatieregel die vaak gebruikt wordt is de meerderheidsregel. Deze regel stelt dat, wanneer de meerderheid van de buren van een knoop actief is, de knoop zelf ook actief wordt op het volgende tijdpunt. Hoewel de deterministische actualisatieregel efficiënt kan zijn, is het geen accurate representatie van psychologische constructen omdat er altijd een vorm van onzekerheid zit in mensen, hun gedrag, en hun emoties. Daarom hebben we een SCA geïntroduceerd, waarin we nog steeds kijken naar de buren van een knoop. In plaats van het vastzetten van de uitkomst op basis van van de buren van een knoop wordt er een kans p toegewezen voor die knoop om actief te worden op het volgende tijdpunt. Door aan te nemen dat alle knopen in een systeem dezelfde actualisatieregel hebben en dezelfde hoeveelheid buren kunnen we het systeem reduceren tot een eendimensionale, discrete Markov-keten. Omdat we ook aannemen dat elke knoop zich slechts in één van twee staten kan bevinden is het proces van het bijwerken van de staat van een knoop Bernouilli verdeeld. Dit simplificeert het systeem waardoor we het gehele systeem makkelijker kunnen analyseren.

Met de GVA hoeft men alleen de staat van alle knopen op $t = 0$ te bepalen en de kans parameter p voor de meerderheidsregel om het systeem continue te laten bijwerken. Op een bepaald punt in dit proces zal het systeem één of twee evenwichten vinden, gebaseerd op de kans parameter p. Als een systeem twee evenwichten heeft kan het kritische transities vertonen tussen de twee; met een enkel evenwicht zijn kritische transities theoretisch onmogelijk. Door de kans parameter te laten variëren kunnen we de ruimte waar het systeem zich in bevindt in kaart brengen en vaststellen waar men theoretisch gezien kritische transities kan verwachten.

Wanneer men de kans parameter p en de ruimte van het systeem kent, kan men in theorie bepalen of een systeem kritische transities kan ervaren. In de psychopathologie zouden we kunnen inschatten of een individu mathematisch gezien op een plek is waar

kritische transities kunnen voorkomen. In de praktijk schatten we deze kans parameter uit de data door middel van maximale waarschijnlijkheidsschatting. In zowel een simulatiestudie als een empirische analyse hebben we laten zien dat dit schattingsproces goed werkt en dat we in staat zijn om te bepalen of patiënten met een depressie diagnose of gezonde deelnemers kritische transities konden verwachten.

Relaties tussen variabelen zijn vaak wederkerig. Een dergelijke relatie geeft aan dat twee variabelen elkaar beïnvloeden. Het vaststellen van een causale relatie tussen twee variabelen is een essentiële volgende stap; het vinden van de oorzaak van een symptoom zoals concentratieproblemen kan ons helpen om dergelijke symptomen te verminderen. Een causale relatie kan wijzen op meerdere dingen. Hier definiëren we een causale relatie als een relatie $X \longrightarrow Y$ waarbij, wanneer we de oorzaak (X) veranderen, we een verandering in het gevolg (Y) waarnemen. Dit betekent ook dat, wanneer we geen verandering in Y waarnemen, er geen verandering in X heeft plaatsgevonden.

Het soort data dat vaak gebruikt wordt om causale relaties te schatten zijn *observatiedata*. Dit zijn data waarop geen perturbaties (manipulaties of interventies) hebben plaatsgevonden. We kunnen echter niet alle causale relaties vinden met louter observatiedata. Zoals geïmpliceerd in onze definitie van een causale relatie moeten we variabelen perturberen en de effecten hiervan waarnemen om causaliteit tussen variabelen te kunnen vaststellen. Hiervoor hebben we *experimentele data* nodig. Dit zijn data waarop een vorm van een perturbatie heeft plaatsgevonden. Door observatie en experimentele data te combineren hebben we wellicht genoeg informatie om causale relaties naar behoren te schatten. Observatie data worden gebruikt om een verbinding tussen twee variabelen vast te stellen en experimentele data om te schatten of één (of beide) variabelen veranderen als gevolg van het perturberen van de andere variabele.

In het tweede deel van dit proefschrift hebben we verschillende algoritmes waarmee causale relaties geschat kunnen worden getest in een simulatiestudie. In deze studie vonden we dat twee specifieke algoritmes - het invariante causale predictie (ICP) algoritme en het verborgen invariante causale predictie (VICP) algoritme - de beste papieren hebben

om causale relaties te schatten. Deze algoritmes zoeken naar de gemene deler: alleen die causale relaties die in alle deelverzamelingen van de data voorkomen blijven behouden in de uiteindelijke causale graaf. De verschillen tussen de algoritmes liggen in het omgaan met *verborgen variabelen*, waarbij het ICP-algoritme niet controleert voor zulk soort variabelen, en het VICP-algoritme wel.

In een empirische studie hebben we de resultaten van beide algoritmes en een literatuurstudie gecombineerd in een causale-graaf-benadering om een causale graaf van obsessief-compulsieve stoornis te creëren. Hoewel we een gebrek aan overlap moesten vaststellen tussen de algoritmes enerzijds en de literatuurstudie anderzijds, ontdekten we causale relaties die anders wellicht onbekend zouden blijven. We hebben daarmee het potentieel laten zien van een dergelijke benadering voor de psychopathologie.

In mijn discussie heb ik een kritische blik geworpen op mijn onderzoek. Welke stappen zijn er genomen, wat hebben we bereikt, wat moeten we doen om de doelen te bereiken waarbij we kunnen voorspellen (in plaats van inschatten) of individuen kritische transities kunnen ervaren en waarbij we volledig begrip krijgen van een stoornis door middel van zijn causale relaties? Hoewel de onderzoeken die gepresenteerd zijn in dit proefschrift de eerste stappen vormen richting deze doelen, zijn we er nog niet. Ik ga in op specifieke beperkingen van de studies, die mogelijk ook beperkingen opleveren voor andere studies, zoals de hoeveelheid aannames die samen met modellen komen, of de (op dit moment nog niet opgeloste) problemen die ontstaan wanneer we om moeten gaan met missende data. Ondanks deze beperkingen kunnen zowel het gemiddelde-veldmodel als de causale-graaf-benadering op verschillende manieren uitgebreid en aangepast worden. Het is wellicht zelfs mogelijk om de twee te combineren in een complex, causaal systeem.

Dit proefschrift laat zien dat zowel predictie als verklaring belangrijk zijn in psychologisch onderzoek. Door een systeem passief of actief onder druk te zetten kunnen we inschatten waar een systeem naartoe gaat en krijgen we een beter begrip van de interne werkingen ervan.

Acknowledgements - Dankwoord

In a dark place we find ourselves, and a little more knowledge lights our way
(Yoda, Star Wars Episode III: Revenge of the Sith)

Ik ben geen extreem grote fan van Star Wars, maar deze epische filmserie liep als een rode draad door mijn promotietijd heen. Lourens, jij startte deze trend in je praatje tijdens mijn buluitreiking, waar je de link legde tussen Luke Skywalker en Yoda enerzijds, en ons twee anderzijds. Ik ben blij dat we dit niet letterlijk hebben genomen en dat ik je niet op mijn rug de afdeling rond hoefde te dragen, maar de figuurlijke gelijkenis staat nog steeds. Ik weet nog goed dat je me vroeg om een NWO-talentbeurs aan te vragen, wat het startsein bleek voor onze samenwerking. Toen al had je alle vertrouwen in een goede afloop, en samen met een grote dosis geduld van jouw kant ligt er dan ook een prachtig proefschrift. Ik kan in alle eerlijkheid zeggen dat dit niet was gelukt zonder jou. Je was nooit te beroerd om zaken die ik moeilijk vond nog een keer uit te leggen, en je stond altijd open voor mijn ideeën. Han, ondanks dat je het vaak erg druk had, kon je altijd een gaatje vinden als ik iets met je wilde bespreken. Je gaf een frisse en eerlijke kijk op de stukken die ik had geschreven. Je pragmatisme heeft zeker geholpen bij het publiceren van een aantal artikelen. Denny, Angélique, Hilde, Maarten en Marieke, dank jullie wel dat jullie de tijd en moeite hebben genomen om mijn proefschrift door te nemen. Ik ben vereerd dat jullie in mijn promotiecommissie plaats nemen.

Denny, je bent geen dan geen officiële promotor, maar ik heb wel je als een soort pleegpromotor gezien. Ik kon altijd bij je binnenlopen en had dan vaak een dosis goed

advies voor me. Mede door jouw begeleiding met mijn bachelorscriptie en stage groeide mijn enthousiasme voor onderzoek en de wetenschap. Zonder jou was ik waarschijnlijk verdronken in onderwijsuren, en dankzij jou werd het mogelijk om een periode in Boston door te brengen. Ik wil je bedanken voor alles dat je voor me hebt gedaan en betekend. Angélique, je hebt me vaak niet alleen op academisch, maar ook op persoonlijk vlak geholpen. Je wist moeiteloos de vinger op de zere plek te leggen en passend advies te geven. Ik heb altijd het gevoel gehad dat ik bij je terecht kon, ook al werkte je niet meer op dezelfde afdeling (of universiteit). Ik heb genoten van onze gesprekken, al dan niet met een lekker glas wijn erbij. Ik heb veel geleerd van onze samenwerking, waar hoofdstuk 2 het resultaat van is.

Ik wil iedereen van de PsychoSystems groep en Yield bedanken voor de interessante labmeetings en discussies. Tessa, dank je wel voor de hulp in mijn laatste maanden als AiO. Ik heb veel aan je tips en opmerkingen gehad! Maarten, ik waardeer het dat je altijd de tijd voor me nam als ik een vraag had, je enthousiasme is aanstekelijk. Sacha, dank je wel voor al je hulp die me in de afgelopen 6 jaar hebt geboden. Sharon, dank je wel dat je mij in 2011 hebt aangenomen als werkgroepbegeleider voor het vak SPSS. Ik heb met veel plezier met je samengewerkt bij verschillende vakken, en veel van je opgestoken wat betreft het geven van onderwijs. Ik wil de rest van de afdeling PML bedanken voor de mooie tijd aan deze geweldige afdeling. Zonder jullie was PMLabs nooit zo'n groot succes geworden! Louise, de theekransjes op jouw kamer waren altijd erg gezellig en zal ik zeker missen.

I want to thank my colleagues from Harvard University. Nicole, Emily, Payton, Ben, Grant, Mike, Jack, John and Katie-Ann, thank you for your warm welcome, I have had a wonderful time at your lab and I learned a lot. Marieke, thank you for being an amazing roommate, lunchmate, and friend! I enjoyed how much we could relate when we compared European and American habits, and you cheered me up when I was homesick. Richard, thank you for allowing me to visit your lab and conduct some research there, I greatly appreciate it and I learned a lot. Don, thank you for helping me get around Boston,

I really enjoyed our talks.

Alle artikelen die ik heb geschreven waren nooit tot stand gekomen zonder de hulp van (vele) co-auteurs. Ik wil jullie hier bedanken voor alle feedback en hulp bij het schrijven. Ik heb elke samenwerking als zeer prettig beschouwd!

Lisa en Pia, wat ben ik blij dat jullie naast mij staan op deze bijzondere dag. Lisa, we hebben elkaar leren kennen tijdens het PML-specialisatie jaar, en trekken sindsdien veel met elkaar op. Je hielp me bij Controverses en gaf me inspiratie voor nieuwe boeken om te lezen. Ook kan ik altijd bij je terecht voor een goed advies en een leuke avond met een mooie fles wijn en een leuk spelletje. Pia, wij hebben elkaar een jaar later ontmoet. Het duurde even, maar gedurende onze AiO-periode liepen we als buurman en buurman over de afdeling. Het maakt niet uit hoe gek de vraag, ik kan hem je altijd voorleggen. Ik vond heerlijk om je als kamergenoot te hebben, en miste onze theepauzes als een van ons in het buitenland zat. Ik kijk altijd uit naar de spelletjesavonden met jullie beiden, en ik hoop ook dat we deze nog vaak kunnen houden!

Lieve vrienden buiten de universiteit, jullie zorgden altijd voor een moment van ontspanning wanneer ik dit nodig had. Robbert, naast goede vrienden zijn we inmiddels donorbuddies, moddermaten en wederzijdse getuigen geworden. We hebben vaak de grootste lol en missen we maar weinig Marvel films. Ik ben er trots op dat je mijn beste vriend bent, en hoop nog vele jaren met je te kunnen lachen! Esmeralda, we spelen dat niet meer in hetzelfde team, toch zien we elkaar nog op regelmatige basis. Ik waardeer je goede adviezen en het luisterend oor dat je biedt. Tanja, het blijft leuk om over Schöne te zwijmelen, alhoewel hij inmiddels niet meer bij Ajax speelt. Ik beleef veel plezier aan onze etentjes samen met Esmeralda. Lieve meiden (en heren) van D2: de trainingen en wedstrijden vormden vaak een moment van rust en ontspanning in mijn hectische weken. Jullie staan altijd voor me klaar met een schouder of een koud biertje, het ABC-tje zorgt nog steeds voor een lach op mijn gezicht #liefsvanD2. Rachel en Sarah, we kennen elkaar al bijna tien jaar inmiddels, wat gaat de tijd toch snel. Mijn studietijd is door jullie significant leuker en gezelliger geworden. Ik hoop nog vaak in jullie gezelschap te mo-

gen verkeren. Halmer, Simone, Danny, Senta, Marten, Carolien, Patrick, Leon, David, Marloes, Nico, Anna, Bea en Douwe, dank jullie wel voor de gezellige avonden. Ik zie jullie naast Eelke's vrienden ook als mijn vrienden. Halmer, Simone, Danny en Senta, een bruiloft plannen en een proefschrift schrijven is geen ideale combinatie, maar mede dankzij jullie hulp bij de bruiloft is deze periode een stuk soepeler verlopen! Simone, dank je wel dat je, ondanks je eigen drukke tijden, ruimte hebt gevonden om de Nederlandse onderdelen van dit proefschrift na te lezen, dat waardeer ik enorm! Sophie en Timo, het is altijd ontzettend gezellig met jullie. Ik vind het fijn om jullie als vrienden in mijn (en ons) leven te hebben! Ingrid, Raymond, Catharina en Roelof-Jan, ik geniet van de gezellige samenkomsten die we hebben met kerst of in het land van Bartje.

Brenda, dank je wel voor het ontwerpen van de prachtige omslag van dit proefschrift. Ik ben jaloers op je creativiteit en je hebt de essentie van mijn proefschrift omgezet naar een geweldig ontwerp! Els, het is soms eng hoeveel we op elkaar lijken qua innerlijk (en soms uiterlijk). Je bent een geweldige vriendin, die altijd in is voor een beetje nerderigheid of een lekker stukje kaas.

Lieve mam, dank je wel dat me altijd gesteund hebt in alles wat ik wilde doen. Pap, toen ik vertelde dat ik ging promoveren riep je van enthousiasme dat ik professor werd. Dat laatste gaat hem waarschijnlijk niet worden, maar ik heb je steun altijd gewaardeerd. Mariska, wat ben ik trots dat ik je kleinste zus ben. Dankzij jou begrijp ik wiskunde veel beter! Je staat altijd voor me klaar, en dat waardeer ik enorm. Lieve Nankie, Rob, Ineke, Annemarie, Rolf, Teun en Elske, dank jullie voor het warme welkom in jullie familie. Vanaf moment één was ik een volwaardig onderdeel en dat ben ik met veel plezier.

Mijn laatste woorden zijn voor de persoon aan wie ik dit proefschrift opdraag. Mijn allerliefste Eelke, zonder jou had ik hier niet gestaan. Terwijl beursaanvragen, afwijzingen en buitenlandtripjes de revue passeerden was (en ben) je er altijd om mij te steunen. Ik kan mijn ei bij je kwijt en je steekt een hart onder mijn riem als ik weer eens onzeker was over een stuk of een presentatie. Je weet precies de goede woorden te zeggen die me geruststellen als ik me weer eens te veel druk maak over iets. Je hebt mij leren kennen als

studentje die in de lift naar het Ciel Blue nog stond na te kijken of ze haar scriptie wel echt had ingeleverd (wat natuurlijk het geval was). Inmiddels kun je er niet meer onderuit: het downdaten is begonnen. Wat de toekomst zal brengen weten we niet, maar ik weet wel dat we samen alles aankunnen!

Conway's Game of Life

Here you can play Conway's Game of life Gardner (1970), discussed in the introduction. Here are the rules to the game:

1. A live cell (black) that has less than two live neighbours will die.
2. A live cell with two or three live neighbours will live.
3. A live cell that has more than three live neighbours will die.
4. A dead cell that has exactly three live neighbours will be brought to live.

On the next page you will find several empty boards. Use the first one to colour in several cells, and play the game from there.

Enjoy!