Symptom network models in depression research

van Borkulo, Claudia Debora

CHAPTER 9

THE CONTACT PROCESS AS A MODEL FOR PREDICTING NETWORK DYNAMICS OF PSYCHOPATHOLOGY

Adapted from:

Van Borkulo, C. D., Wichers, M.C., Boschloo, L, Schoevers, R.A., Kamphuis, J. H., Borsboom, D. & Waldorp, L. J. (2015). The contact process as a model for predicting network dynamics of psychopathology. Manuscript submitted for publication.

It is well-established that the symptomatology of depressed patients is dynamic; symptoms are not continuously present or absent, but instead show patterns of change over time. Here, we present a dynamic network account of these changes in symptomatology. This dynamic network account is based on models to describe the spread of a virus across a population of individuals. Translated to a mental disorder, its dynamics entail the spread of activity across an individual's symptoms — which can activate each other due to causal relationships (e.g., feeling guilty can lead to sleep problems, which in turn may lead to concentration problems, etc.). In this so-called *contact process*, the topology of the network plays a crucial role, because an activated symptom can only directly activate neighboring symptoms in the network structure. We propose a maximum likelihood approach to estimate the infection and recovery rates for such networks on the basis of time-series data that capture symptom dynamics. On the basis of this model, the ratio between infection and recovery results in a Percolation Indicator (PI), which is indicative of behavior of the system in the long run, and may be used to anticipate future behavior of the symptom network. The quality of the proposed model estimates is investigated with simulations and application to real data of patients with depressive disorder. In addition, we derive a t-test to determine differences between individuals and to test the PI against a critical value. Results indicate that the quality of the estimates is good for a range of differently sized time series of the symptoms. Application of the t-test also reveals that a substantial proportion of patients has a PI above the critical value. We conclude that the contact process model is a promising method for the analysis of symptom dynamics. Future research is needed in which extensions of the model and predictive quality of PI are assessed.

9.1 Introduction

For a patient with depressive disorder, let's call her Alice, the loss of her job induced self-reproach, which in turn led to insomnia, fatigue and other symptoms of depression, which caused a vicious circle that spiraled her into a state of mental disorder. Bob, Alice's colleague, also got fired, but did not go on to develop such a full-blown depressive episode. While such differences between individuals are familiar to psychiatrists and clinical psychologists involved in the care for patients with depressive disorder, it is not yet clear why some people develop depressive

episodes in response to adverse life events, while others do not. Recent work has proposed that the key to understanding such differences lies in the way these symptoms interact with each other in a network structure (Borsboom & Cramer, 2013; Borsboom et al., 2011; Cramer et al., 2010). The current chapter exploits this perspective to develop a statistical model that can be used to analyze symptom dynamics in a way that can help to assess whether a given person is likely to develop a disorder, or is more likely to recover. This may provide clues to further understand the etiology of the disorder and may also be relevant to the challenge of identifying optimal targets for therapeutic interventions (Baglioni & Riemann, 2012; Rosmalen et al., 2012).

Classical approaches to study depression based on putative central psychological mechanisms or their neural substrates has lead to limited insights in the etiology and structure of the disorder. In fact, the inability to develop an adequate model for the etiology of disorders might be one of the crucial issues underlying the question of “what kinds of things" psychiatric disorders are (Kendler et al., 2011). Although current classifications of disorders suggest coherent categories, many aspects of the structure and dynamics of mental disorders are not well understood. For example, patients with the same disorder may show different symptomatology (Van Loo et al., 2012); in addition, these different symptoms are predisposed by different risk factors (Fried, Nesse, et al., 2014; Lux & Kendler, 2010). Furthermore, different symptoms may have a different impact on psychosocial functioning (Fried & Nesse, 2014; Tweed, 1993) and similar stressful life events have different impact on individual impairment (Bonanno, 2004), as illustrated in the case of Alice and Bob. It has been argued that the limited insights in these phenomena calls for *symptomics* (Fried, Boschloo, et al., 2015); modeling the individual building blocks of mental disorders (symptoms). This will lead to understanding the mechanisms of psychopathology.

Many studies have investigated psychopathology from a network perspective. (see Fried, Van Borkulo, et al., 2016, for a review). However, many of these studies are based on cross-sectional data (Beard et al., 2016; Bekhuis, Schoevers, Claudia Borkulo, Rosmalen, & Boschloo, 2016; Boschloo, Van Borkulo, Borsboom, & Schoevers, 2016; Boschloo et al., 2015; Fried, Bockting, et al., 2015; Fried, Epskamp, Nesse, Tuerlinckx, & Borsboom, 2016; Isvoranu, Borsboom, Os, & Guloksuz, 2016; McNally et al., 2015; Rhemtulla et al., 2016; Robinaugh et al., 2014; Ruzzano et al., 2014; Van Borkulo et al., 2015). It is, therefore, not clear how these group-level

results relate to individuals. One promising way of studying these mechanisms in individuals is by conceiving the multivariate dynamics of symptomatology as a network, that is driven by a system of causal relations between symptoms (Borsboom & Cramer, 2013; Borsboom et al., 2011; Bringmann et al., 2013; Cramer, van der Sluis, et al., 2012; Cramer et al., 2010; Kendler et al., 2011; Pe et al., 2015; Wichers, 2014; Wigman et al., 2015). Since these networks can differ across individuals (Bringmann et al., 2013), this may offer a scientific inroad to address the question of why some people develop mental disorders while others do not. In Alice′s case, this could entail that feeling worthless leads to a depressed mood, which in turn leads to suicidal thoughts: worthless → depressed mood → suicidal thoughts. In Bob′s case, however, feeling worthless may not lead to other symptoms, such that he does not get caught up in a vicious circle, and after a while he may start to feel better again. Knowing an individual′s pattern of symptom dynamics, visualized as a network, could thus be highly relevant for uncovering prognostic information. An individual′s pattern of symptom dynamics might reveal why Alice is prone to develop a full-blown disorder while Bob is not, even before the disorder emerges; perhaps, the pattern of symptom dynamics in Alice′s case is such that symptoms more easily influence each other, thereby culminating more easily into a full-blown disorder than in Bob′s case. Following the same line of reasoning, one could hypothesize that when Alice and Bob both suffer from a depressive episode, their patterns of symptom dynamics might reveal why Bob will recover soon and Alice′s depression is persistent.

Symptoms that interact with each other can be viewed as an *interacting particle system* (Liggett, 2004). A commonly used model to study such systems is the contact process model (Grimmett, 2010) — a probabilistic equivalent of the epidemiological *SIS* model (susceptible-infected-susceptible; Keeling & Rohani, 2011; Newman, 2010) — which will form the basis for modeling symptom dynamics in this chapter. Such models describe, for example, the spread of a disease (e.g., virus) across a population of individuals. After an individual has been infected, the individual recovers and becomes susceptible to infection again (Newman, 2010). Behavior of such models is influenced by the specific topology of the interactions between individuals (Keeling & Eames, 2005). Considering mental disorders by analogy, a symptom within a person can influence other symptoms within that person by turning them on.

Although symptoms can be argued to be continua, viewing psychopathology as a SIS or contact process model, in which symptoms can only be *on* or *off*, can be seen as a reasonable approximation for several reasons. First, the Diagnostic statistical Manual (DSM-5; American Psychiatric Association, 2013) contains symptoms that are scored as being present or absent. Second, Keeling and Eames (2005) showed that network topology has an impact on the evolution of SIS models. Third, similar results are recently found in symptom networks; the connectivity of a symptom network was associated with the future course of depression (Van Borkulo et al., 2015). The group-based baseline network of depressed patients who have persistent depression at two-year follow-up were found to be more densely connected than the baseline network of depressed patients who remitted at two-year follow-up. Apparently, a more densely connected network at baseline is related to a worse prognosis, which is also confirmed in comparing healthy controls to individuals with a diagnosis (Pe et al., 2015; Wigman et al., 2015).

Building on the above-mentioned group-level findings, we can use the SIS or contact process model to make actual predictions about the behavior of an individual's psychopathology network models by using percolation theory (Grimmett, 2010). The key concept of percolation theory is the *basic reproduction number* (Kolaczyk, 2009), defined as the ratio ρ between the infection and recovery rate λ and μ ($\rho = \lambda/\mu$), respectively (Fiocco & van Zwet, 2004; Grimmett, 2010; T. E. Harris, 1974; Newman, 2010). Here, we apply the concept of basic reproduction number ρ. If ρ is larger than a critical value $\rho > \rho_c$, the process will continue forever, whereas if $\rho \leq \rho_c$ the process will die out (Fiocco & van Zwet, 2004). In our application to psychopathology, this basic reproduction number can be interpreted as the balance between infection and recovery of symptoms. We do not know ρ_c, but we entertain the idea that $\rho_c = 1$ marks the transition between a state in which the disorder will die out or not. We call ρ the *percolation indicator* (PI). PI represents a quantification of the concept that certain characteristics of a symptom network (e.g., stronger and more connections, or a specific topology) increase the risk of becoming or remaining disordered. Note that when applying these models to psychopathology, $\rho > 1$ would imply lifelong pathology for the individual under consideration. Obviously, this is not realistic since individuals are not static models. The model can be influenced by various external factors that we do not consider here and it is plausible that an individual's symptom

network can change over time. We, therefore, assume that the ratio between λ and μ is indicative of behavior of the network in the *nearby* future. In which time span ρ is predictive for psychopathology remains an open question and requires further investigation.

Modeling the dynamics of a disorder requires multiple repeated measurements of symptomatology, which is challenging for two reasons. First, models that are currently often used are autoregressive models, which assume stationarity (Startz, 2008). It is doubtful, however, whether this assumption is tenable in daily clinical practice; it is conceivable that the probability that a symptom can turn on one moment is different from that at another moment. For example, one is more likely to have concentration problems after a period of insomnia than after a good night′s sleep. Second, in many dynamic models the measurements are required to be equidistant, which implies, for instance, that the patient completes the questionnaire at exactly the same time every day. Patients may plan their activities around the fixed time to fill out the questionnaire, thereby missing out on emotions during these activities. Therefore, it would be desirable to randomly prompt the patient to fill out the questionnaire and, consequently, capture as much of the emotions and feelings. The method of data collection by randomly prompting is called Experience Sampling Method (ESM; aan het Rot et al. 2012; Bolger, Davis, and Rafaeli 2003; Larson and Csikszentmihalyi 1983; Oorschot et al. 2012; Rosmalen et al. 2012). As we show in this chapter, the contact process model can both help to bypass the stationarity assumption and the assumption of equal time intervals between measurement occasions, and as such may lead to improvements in the analysis of ESM data. Moreover, PI can be used to find evidence for the idea that the topology of the network has an impact on the risk of a mental disorder (Cramer et al., 2016).

This chapter describes the contact process as a model for the dynamic relationships between symptoms within individuals, and develops PI as a predictive tool. Moreover, the methodology is applied to an ESM dataset involving time series of depression symptomatology. Since variables in empirical data can be influenced by external factors that are not measured (i.e., an event in daily life can affect one ore more symptoms), the assumption that the system under consideration is closed is violated. We adjusted some model parameters accordingly to take external influences into account. Note that, since our data does not allow this, we do not assess predictive quality here.

The chapter is organized as follows. Section 9.2 is about the model specification, followed by a section on how PI can be estimated (9.3). Section 9.4 contains a validation study to assess the quality of the estimator. In Section 9.5, the utility of the presented method is illustrated with data of patients with depressive disorder. Here, we explain which parameters of the original model should be adjusted when applying the model to real data. Finally, Section 9.6 covers a discussion on future directions. The method for estimating PI is implemented in R in the PercolationIndicator package with a built-in example and accompanying documentation (Van Borkulo, Epskamp, & Milner, 2016).

9.2 Model specification

The basis for modeling symptom dynamics is the contact process model (Grimmett, 2010), defined originally in T. E. Harris (1974). Classic SIS models (equivalent to the contact process model) are defined on a population assumed to be *fully mixed*, meaning that every symptom (or other entity) is equally likely to interact with all other symptoms (Newman, 2002, 2010). According to the network approach to psychopathology, however, it is this particular configuration of interactions between symptoms that constitutes what a disorder is; the limited set of interactions defined by the topology explain the behavior of a system. Therefore, following Keeling and Eames (2005), we use the contact process and take into account the specific graph topology (i.e., instead of a fully mixed graph), in which symptoms can only interact with direct neighboring symptoms (Newman, 2002, 2010).

The contact process model can be viewed as an undirected network (see Figure 9.1 for a graphical representation) and is characterized by two independent Poisson processes: one for infection and one for recovery. Infection is a process that involves neighboring nodes. This means that a node can be infected by an already infected node that is connected. In Figure 9.1, infection of one node by another is indicated by an arrow. Recovery, according to the contact process, is considered an autonomous process: it does not involve interactions with other nodes. When infected, recovery takes place at a certain rate, independent of the other nodes (Fiocco & van Zwet, 2004; Grimmett, 2010; T. E. Harris, 1974). The ratio of infection and recovery rates (reproduction number; Kolaczyk, 2009) are the main interest of this chapter, since the ratio between the two is indicative of

future behavior of the system (Fiocco & van Zwet, 2004); intuitively, if the recovery rate is higher than the infection rate, activity in the system will die out, whereas if the infection rate is higher than the recovery rate, the system will stay activated. In this section, we describe the model we use to estimate the infection and recovery rate.

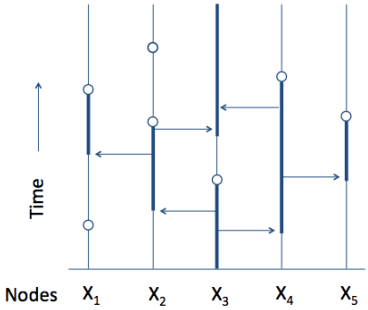

FIGURE 9.1. The contact process (after Grimmett 2010). On the horizontal axis are nodes X_1 through X_5 that are connected in a linear fashion. The vertical axis represents time. The bold lines indicate the time that a node is infected and arrows represent the time point on which a node infects a neighboring node. The mark ∘ represents recovery.

Consider Figure 9.1, in which random variables X_1 through X_5 can be viewed as processes on a graph that can be described as follows. Nodes (variables) can interact if they are connected (i.e., if they are depicted next to each other in Figure 9.1). An already infected node can infect a neighboring node at a certain time point (indicated by an arrow in Figure 9.1). In the contact process, recovery is considered an autonomous process: it does not involve interactions with other nodes.

To formally define the model, let $G = (V, E)$ be an undirected graph with $|V|$ nodes and $|E|$ edges. A node $x \in V$ represents an entity — regarding psychopathology, this can be a symptom or an emotion — and an edge $(x, y) \in E$ represents a relationship between the entities (e.g., a statistical association). Nodes can be in one of two states, healthy or infected, with the state of node $x \in V$ at time s given

by

$$\xi_s(x) = \begin{cases} 0 & \text{if node } x \text{ is healthy at time } s \\ 1 & \text{if node } x \text{ is infected at time } s \end{cases} \tag{9.1}$$

At a certain time point s, three types of events can be observed: infection, recovery or no observed switch

$$0 \to 1, \quad 1 \to 0, \quad \text{and} \quad 0 \to 0 \quad \text{or} \quad 1 \to 1.$$

The contact process $\xi_s(x)$ ($s \geq 0$) is a type of nearest neighbor process where the rates of changes in state for each node in $x \in V$ are as follows. For each $x \in V$, the number of neighbors at time s capable of infecting node x, can be defined by the random variable $\xi_s(x)$ as

$$k_s(x) = (1 - \xi_s(x)) \sum_{y \in V : (x,y) \in E} \xi_s(y) \tag{9.2}$$

We assume that the probability of infection and recovery are given by (Grimmett, 2010)

$$\begin{aligned} \mathbb{P}(\xi_{s+h}(x) = 1 \mid \xi_s(x) = 0) &= \lambda k_s(x) h + o(h) \\ \mathbb{P}(\xi_{s+h}(x) = 0 \mid \xi_s(x) = 1) &= \mu h + o(h) \end{aligned} \tag{9.3}$$

as $h \downarrow 0$. This defines a $|V|$-dimensional continuous-time Markov process with Poisson probabilities (Norris, 1997). A process is called Markov if it satisfies the Markov property that the stochastic process is memoryless; the future state of the system is independent from the past, given the present. A node x can be infected ($0 \to 1$) with rate $\lambda k_s(x)$. It becomes clear, then, that the structure of the network plays an important role in the process of infection, since the number of infected neighbors is determined by the topology. Indeed, infected nodes are only counted as capable of infecting node x when x is not infected already ($\xi_s(x) = 0$). The total number of neighbors in the network G capable of infecting some other connected node at time s is then $k_s(V) = \sum_{y \in V} k_s(y)$. Note that the same infected node can appear multiple times in this summation, since an infected node can be capable of infecting multiple neighboring nodes. Recovery of a node is defined for each node separately and independently, and changes the state of the node $1 \to 0$ at rate μ.

We can equivalently define the process by its transition probabilities (Norris, 1997, Theorem 2.4.3). The transition probability of $0 \to 1$ and $1 \to 0$ are, respectively,

$$p_s(x) = \frac{\lambda k_s(x)}{\lambda k_s(x) + \mu}, \quad q_s(x) = \frac{\mu}{\lambda k_s(x) + \mu}. \tag{9.4}$$

The 2×2 transition probability matrix $P_s(x)$ of node x at time s of the two state Markov process is as follows (Brzezniak & Zastawniak, 2000; Grimmett, 2010, see Appendix C, Section C.1 for a derivation)

$$P_s(x) = \begin{pmatrix} 1 - p_s(x) & p_s(x) \\ q_s(x) & 1 - q_s(x) \end{pmatrix} \tag{9.5}$$

The conditional probabilities $p_s(x)$ and $q_s(x)$ indicate, respectively, the transition probability of node x being infected given that it was previously recovered and the probability that node x recovers given that it was previously infected.

An important property of the contact process is the basic reproduction number, defined as the ratio between infection and recovery

$$\rho = \frac{\lambda}{\mu}. \tag{9.6}$$

As stated in the introduction, we call ρ the Percolation Indicator (PI).

In practice, we observe events at several (random) time points, and, consequently, do not have access to the fully continuous process. We therefore assume that $\xi_s(x)$ for $s \geq 0$ is a right-continuous Markov process (Norris, 1997), meaning that when a node is in one of two states (e.g., healthy or infected), it stays in that state until the time of the next event; then it switches to the other state (e.g., infected or healthy). The number of events during the time interval $[0, s]$ is modeled as a Poisson process (Brzezniak & Zastawniak, 2000; Grimmett, 2010; Norris, 1997). Consequently, the time between events (i.e., holding time) is assumed to be exponentially distributed (Norris, 1997). We then have the discrete time Markov chain $\xi_i(x)$ for $i = 1, 2, \ldots$. It is this discrete-time Markov process that we will use for estimation.

9.3 Estimation procedures

In this section, our main focus is on ρ, the estimator of PI. However, in order to obtain ρ, we require a network. In general we do not have the underlying symptom

network of psychopathology for a specific person. We therefore need to estimate the network structure in order to estimate ρ. The basic tool to obtain a network that provides statistical dependencies between symptoms is the graphical model (Lauritzen, 1996; Wainwright & Jordan, 2008). While there is much to say about how to best infer the network structure from data, we will use the graphical VAR model (Wild et al., 2010). We describe the network estimation method based on this model in this section and provide a simulation study in Section C.2, where it can be seen that the method we use works well for the length of time series we have in our type of data.

9.3.1 Percolation Indicator estimation

The ratio $\rho = \lambda/\mu$ is of particular interest to us, as we can use it to determine whether someone is at risk ($\rho > 1$) or not ($\rho \leq 1$). Here we derive the maximum likelihood estimates for λ and μ similar to Fiocco and van Zwet (2004) such that the ratio ρ can be determined.

Let the time points of events (randomly chosen measurement points) be indicated by T_i for $i = 1, 2, \ldots, n$. The assumption that in the time interval $[T_{i-1}, T_i)$ there is a single change, leads to the rates of the Poisson process

$$r_i(x) = \begin{cases} \lambda k_{i-1}(x) & \text{if } \xi_{i-1}(x) = 0 \\ \mu & \text{if } \xi_{i-1}(x) = 1 \end{cases} \tag{9.7}$$

This rate of change can also be written as a sum of mutually exclusive events

$$r_i(x) = \lambda k_{i-1}(x)(1 - \xi_{i-1}(x)) + \mu \xi_{i-1}(x). \tag{9.8}$$

In the Poisson process, holding times $T_i - T_{i-1}$ for each $x \in V$ are independently, exponentially distributed with density $\lambda k_{i-1}(x) \exp[-\lambda k_{i-1}(x)(T_{i-1} - T_i)]$ for infection and independently $\mu \xi_{i-1}(x) \exp[-\mu \xi_{i-1}(x)(T_{i-1} - T_i)]$ for recovery. Therefore, the likelihood of the observed process for the interval $[0, t]$ (with $T_0 = 0$ and $T_N = t$) is defined as (Fiocco & van Zwet, 2004)

$$L(\lambda, \mu) = \prod_{i=1}^{n} \prod_{x \in V} r_i(x) \exp[-r_i(x)(T_i - T_{i-1})] \tag{9.9}$$

Consequently, the log-likelihood is

$$\log L(\lambda, \mu) = \sum_{i=1}^{n} \sum_{x \in V} \log r_i(x) - \sum_{i=1}^{n} \sum_{x \in V} r_i(x)(T_i - T_{i-1}) \tag{9.10}$$

The second term on the right hand side can be decomposed into an infection and a recovery part. For the infection, by definition of $k_i(x)$ in (9.8) infected neighbors are counted only if x is not infected, that is if $\xi_{i-1}(x) = 0$. Moreover, since we assumed that in the time interval $[T_{i-1}, T_i)$ there is no change in $\xi_s(x)$ (and hence no change in $k_s(x)$) for any $x \in V$, we find that for each i

$$\sum_{i=1}^{n} \sum_{x \in V} \lambda k_{i-1}(x)(1 - \xi_{i-1}(x))(T_i - T_{i-1}) = \lambda \sum_{i=1}^{n} k_{i-1}(V)(T_i - T_{i-1}) = \lambda A_t,$$

in which A_t is the sum of infected neighbors that are capable of infecting, over all nodes and all n time points. Similarly, for recovery we obtain

$$\sum_{i=1}^{n} \sum_{x \in V} \mu \xi_{i-1}(x)(T_i - T_{i-1}) = \mu \sum_{i=1}^{n} m_{i-1}(V)(T_i - T_{i-1}) = \mu B_t,$$

in which $m_{i-1}(V)$ is the number of infected nodes at time point T_{i-1}, and, consequently, B_t is the sum of infected nodes over all n time points. Putting them together results in

$$\sum_{i=1}^{n} \sum_{x \in V} r_i(x)(T_i - T_{i-1}) = \lambda A_t + \mu B_t. \tag{9.11}$$

For the first term in (9.10) we consider the product over the infections and recoveries in $r_i(x)$ for all n time points and $|V|$ nodes. For each time interval $[T_{i-1}, T_i)$, the rate $r_i(x)$ is either $\lambda k_{i-1}(x)$ or μ. Consequently, we obtain for the log of the product over time intervals and over nodes

$$\log \prod_{i=1}^{n} \prod_{x \in V} r_i(x) = U_t \log \lambda + D_t \log \mu + \sum_{i=1}^{n} \sum_{x \in V} \log k_{i-1}(x), \tag{9.12}$$

in which

$$U_t = \sum_{i=1}^{n} \sum_{x \in V} \mathbb{1}\{\xi_{i-1}(x) = 0\} \qquad D_t = \sum_{i=1}^{n} \sum_{x \in V} \mathbb{1}\{\xi_{i-1}(x) = 1\} \tag{9.13}$$

are the number of upward and downward jumps across all nodes in V and time intervals $[T_{i-1}, T_i)$, respectively, and $\mathbb{1}\{A\}$ is the indicator function for the set A. The log-likelihood can now be written as

$$\log L(\lambda, \mu) = U_t \log \lambda + D_t \log \mu - \lambda A_t - \mu B_t + c(k), \tag{9.14}$$

in which $c(k)$ is the last term in (9.12) and does not depend on λ or μ. Differentiating with respect to λ and μ yields the maximum likelihood estimates

$$\hat{\lambda} = \frac{U_t}{A_t}, \qquad \hat{\mu} = \frac{D_t}{B_t}. \tag{9.15}$$

This means that $\hat{\lambda}$ is defined as the ratio between the number of upward jumps (U_t) and the number of infected neighboring nodes that are capable of infecting (A_t) in time interval $[0, t]$, across all variables. Estimator $\hat{\mu}$ is defined as the ratio between the number of downward jumps (D_t) and the number of infected nodes (B_t) in time interval $[0, t]$, across all variables. According to Fiocco and van Zwet (2004), PI is defined as

$$\hat{\rho} = \frac{\hat{\lambda}}{\hat{\mu}} = \frac{U_t B_t}{A_t D_t}. \tag{9.16}$$

From equation (C.9), it follows that the intervals $[T_i - T_{i-1}]$ themselves are irrelevant; only the total time interval $[0, t]$ is important. Fiocco and van Zwet (2006) have shown that the estimates for λ, μ, and ρ are consistent and asymptotically normal. However, considering psychological and psychiatric measurements, the assumption that the process is fully observed is untenable: that is, events can be missed. For simplicity, we set the rate of missing events to 1, thereby assuming that in ordinary ESM settings (i.e., in which symptoms or emotions are repeatedly measured during several days) we do not miss more than 1 event between measurements. We are interested in the behavior of $\hat{\rho}$ in the context of finite samples as well as assuming to have missed one event.

9.3.2 Network estimation

The basis for the method to estimate ρ, is the work of Fiocco and van Zwet (2004). In their work, the network that generated the data has a known structure. In many fields, such as psychology and psychopathology, network structures are unknown and have to be inferred from data. To infer the network structure, one can use any method that is consistent to the true underlying network. In this chapter, we use the graphical VAR method (implemented in R package graphicalVAR, version 0.1.2). This method is based on the graphical VAR model (Wild et al., 2010) in which parameters are estimated through ℓ_1-regularization (Friedman et al., 2008), and is combined with the extended Bayesian Information Criterion

(EBIC) to choose the optimal tuning parameters (J. Chen & Chen, 2008). For a detailed outline of the estimation procedure, see Abegaz and Wit (2013) and Rothman et al. (2010). This method estimates two networks: one directed and one undirected network. The directed network describes temporal effects, whereas the undirected *contemporaneous* network describes relationships within time points after controlling for the temporal effects (Epskamp, Waldorp, Mõttus, & Borsboom, 2016). Since the contact process requires an undirected network, we use the temporal network in which parameters > 0 are set to 1, and otherwise is set to 0 (i.e., negative relationships are ignored, since the contact process does not accommodate this).

Although the graphical VAR method has proven to yield accurate network estimations (Abegaz & Wit, 2013), we assessed the quality under circumstances similar to our data. For detailed information and the results of our validation study with graphicalVAR, see Section C.2 in the Appendix.

9.4 Validation study

To assess the quality of our method to estimate ρ, we investigated the Mean Squared Error (MSE), which consists of the squared bias and the variance of the estimate. To investigate the bias and the variance, we generated simulated data. First, we describe the design of validation study (the conditions, how the data was simulated, and how the simulated data was analyzed). Second, we present the results of the validation study.

9.4.1 Design

To simulate data, we generated three types of network structures. We generated a lattice structure, for which estimator ρ originally was developed (Figure 9.2 left panel). In such a structure each node has 4 neighbors and is created with R package igraph. To see how well ρ is estimated with a different network structure, we also generated a network with 50% of the edges of a lattice randomly replaced by other edges, connecting different nodes (Figure 9.2 middel panel), and a structure with 100% of the edges randomly replaced (Figure 9.2 right panel). Network size was chosen to be close to the number of variables in our real data. For each network type, we simulated four main data sets with four different parameter

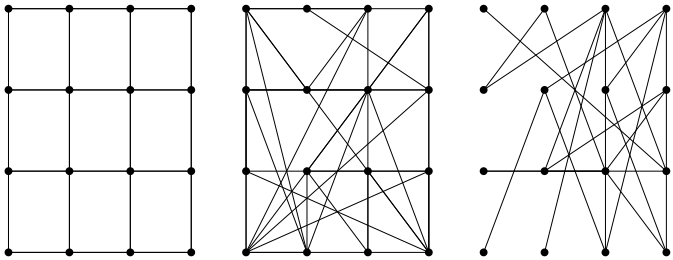

FIGURE 9.2. Simulated networks. Three types of networks in the validation study: a pure 2D lattice structure (left panel), a network in which 50% of the edges of a lattice are randomly replaced (middle panel), and a network in which 100% of the edges of a lattice are replaced (right panel). All networks contain 32 edges as in the lattice structure, in which all nodes have 4 edges (note that some edges are not visible, e.g., the edge from the upper left node to the upper right node in the left panel is indistinguishable).

values (true parameters): ρ = .5, 1, 1.5, and 2. These four main data sets contain eight groups of data sets with each 100 simulations. Sample size n ranges from 50, 70, 90, up to 190. This resulted in a $3 \times 4 \times 8$ factorial design, with the factors network type (lattice, 50% replaced, 100% replaced), ρ value (.5, 1, 1.5, 2), and sample size (90, 70, ..., 190). Each of the 96 conditions was replicated 100 times.

9.4.1.1 Simulation process

With n the sample size, we drew n numbers from a Poisson distribution with rate $\gamma = \lambda + \mu$, where λ and μ are the infection and recovery rates. In our case, the rate parameter γ represents both the infection and recovery parameters λ and μ. The rate parameter γ is actually $\lambda + \mu$. Since we use ρ = .5, 1, 1.5, and 2, while keeping $\mu = 1$ across all conditions, we obtain λ = .5, 1, 1.5, and 2. This means that when, for instance, $\lambda = 2$ and $\mu = 1$, $\gamma = 3$. As a result, if we want to simulate 50 observations, we draw 50 numbers from a Poisson distribution with $\gamma = 3$. This returns 50 numbers that indicate the number of events per observation. These events are simulated with transition probabilities as in equation C.4 (Section C.1 in the Appendix). Now, we have simulated the fully observed process. From this, we draw our observations from the Poisson distribution (with the same rate $\gamma = 3$).

This second draw does not perfectly coincide with the simulated events, resulting in realistic missing of events. See Section C.3 in the Appendix for the R code.

9.4.1.2 Analysis of Mean Squared Error: Bias and variance

The first component of the MSE is the bias, i.e., the expected absolute difference between the estimate and the true value. The bias is assessed by inspecting the distributions of $\hat{\rho}$ from our Monte Carlo simulations. Violin plots are used for this purpose, in which a box plot and a kernel density plot are combined (Hintze & Nelson, 1998). The second component of the MSE is the variance. Since the variance is unknown, it has to be estimated. The usual way to do this is to calculate the variance from the Fisher information. Fiocco and van Zwet (2004) stated, however, that the sample variance yields a better estimate. We investigated both Fisher information and sample variance by comparing them to the Monte Carlo variance in Section C.4 in the Appendix. The Monte Carlo variance is the variance of $\hat{\lambda}$ and $\hat{\mu}$ of 100 simulations.

9.4.2 Results validation study

To assess the quality of the estimation method, we investigated the MSE of the estimates. Inspecting the first component of the MSE (bias) reveals that $\hat{\rho}$ shows little bias; the estimate is consistently close to the true value for all conditions (see Figure 9.3 for data simulated with 100% replaced network structures and Figures C.3 and C.4 in Section C.5 of the Appendix for data simulated with lattice and 50% replaced edges). Therefore, we conclude that $\hat{\rho}$ is little biased.

The second component of the MSE is the variance. Since the variance is unknown, we estimate it with the sample variance (for empirical evidence that the sample variance is a better estimate than the Fisher information variance, see Figure C.2 in Section C.4 of the Appendix). The sample variances of the simulated data indicate that the variance of $\hat{\rho}$ is overall quite low (see Figure 9.4 with 100% replaced networks). The peaks arise from single data sets with one node having a deviant (high) estimate of ρ_x. For results of data simulated with other network structures (lattice and 50% replaced edges), see Figures C.5 and C.6 in Section C.6 of the Appendix).

To conclude, $\hat{\rho}$ shows little bias and the variance is overall low, which makes it a useful estimator in statistical analyses of the kind investigated here.

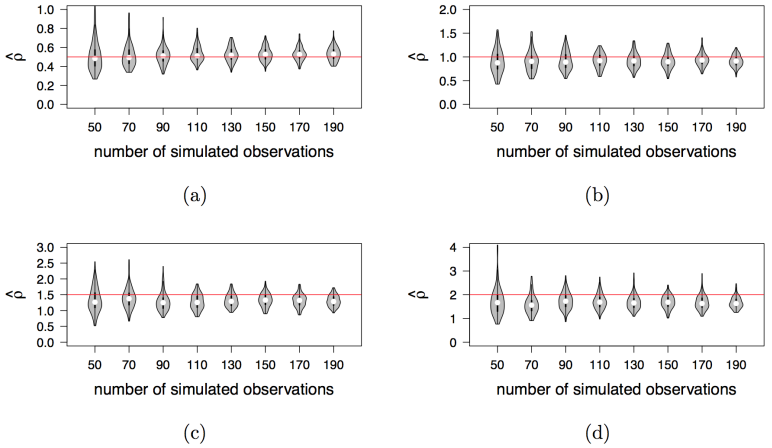

FIGURE 9.3. Violin plots of the estimates of ρ. Distributions of estimates of 100 simulated data sets with 100% replacement networks, increasing amount of observations (50, 70,..., 190) and with $\rho = .5$ (a), $\rho = 1$ (b), $\rho = 1.5$ (c), and $\rho = 2$ (d). The red lines indicate the true value of ρ, with which the data was simulated.

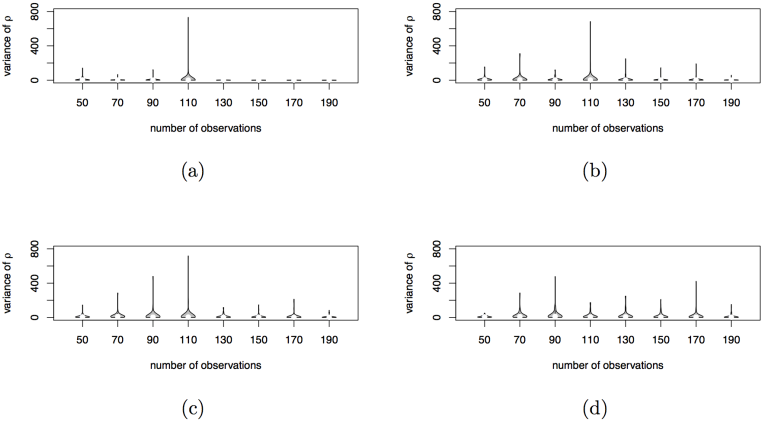

FIGURE 9.4. Sample variances of simulated data based on a network with 100% replacement structure and (a) $\rho = .5$, (b) $\rho = 1$, (c) $\rho = 1.5$, and (d) $\rho = 2$.

9.5 Application of method to real data

To illustrate the utility of the presented method, we analyzed real data of patients with depressive disorder. In this section, we first explain in what respect the application of the contact process model differs when applying it to real patient data and how we adjusted some parameters of the model. Second, we describe the data, explicate how network structures were estimated and how the data was analyzed. Third, we present the result of our application to real data.

9.5.1 Discrepancy between model and real data

As stated earlier, the basis for estimating PI is the contact process model which evolves on a lattice (Fiocco & van Zwet, 2004). This has two important implications when applying it to real data. First, the contact process model assumes a known and fixed network structure. Since the network structure of psychological phenomena such as psychopathology is unknown and likely to differ across individuals, the network structure has to be estimated from empirical data. Since the infection rate parameter is $\hat{\lambda} = U_t / A_t$ (see equation 9.15), where A_t is the number of infected neighbors, $\hat{\lambda}$ strongly depends on the topology of the network, in particular the degree. More densely connected networks will, therefore, tend to have more infected neighbors and, consequently, smaller PIs. This means that PIs of individuals with different networks cannot be compared. Therefore, we correct PI for topology by dividing A_t by the average degree (i.e., $2E/|V|$). Second, the contact process assumes a closed system, meaning that there are no external influences. In such a closed system, a symptom can only be activated by one or more infected neighboring symptoms. In reality, when psychological variables are measured in individuals, this is hardly ever the case. Other variables that are not included in the study, could have had an effect on the measured variables. Therefore, we adjusted the number of upward jumps U_t — this is basically counting all instances in which a variable goes from 0 at time point t to 1 at $t+1$ — by only counting the upward jumps that are preceded by one or more infected neighbors. Violation of the assumption that the system is closed implies for the critical percolation threshold ρ_c that it is unclear whether it is indeed 1. It should be noted that the corrections in PI calculations alleviate the violation of assumptions to some extent, but not entirely.

9.5.2 Description of real data

We used data from a randomized controlled trial (RCT), containing ESM data pertaining to a sample of depressed patients. During a six-week intervention period, 68 patients self-collected ESM data on momentary affective responses for three consecutive days a week for a maximum of 10 measurements per day. Current affect was measured with four positive and six negative affect items (cheerful, satisfied, enthusiastic, relaxed, down, suspicious, guilty, irritated, lonely, and anxious) on a 7-point Likert scale (ranging from 1="not at all" to 7="very"). For a detailed description of the data, see Kramer et al. (2014). In our study, positive affect items were reverse coded such that a low score always means "no problems" and a high score means "problems". For example, when a patient is very relaxed (i.e., originally scored 7 on that item), the score is 1 after reverse coding.

9.5.2.1 Hypothesis testing

As described in Appendix C (Section C.7), we have constructed two tests. First, the one-sample t-test serves to test $\hat{\rho}$ against a critical value ρ_c. Although the value of ρ_c is unknown for systems that are not closed, we use $\rho_c = 1$ to mark the balance between infection and recovery to illustrate how PI can be tested against some critical value. Assuming $\rho_c = 1$, a PI larger than 1 indicates that the symptoms in the network will remain active. Similarly, a PI smaller than 1 indicates the symptoms in the network will eventually die out. Second, the two-sample t-test serves to test whether the PIs of two networks are equal ($PI_1 = PI_2$). PIs of different individuals can be compared since we controled for differences in density by dividing A_t (i.e., the number of infected neighbors) by the average degree of the network.

Tests were performed at significance level $\alpha = .05$. See Figure C.7 in Section C.7 of the Appendix, for the quality of the test statistic.

9.5.3 Results of application to real data

To get reliable network estimates, a minimum of 50 measurements per patient was applied (see Section C.2 Appendix). This resulted in 39 networks of which PI ranged from 0 to 54.53 ($M = 2.84$, $sd = 8.84$). There were 14 patients (36%) with PI > 1.

We selected six patients as an example to illustrate the benefits of our approach to analyze data at the individual level: three cases with PI< 1 and three cases with PI> 1. We randomly chose patients with a wide range of PIs, excluding the ones with extreme values. As Figure 9.5 seems to reveal, the three patients with PI< 1 have a lower average sum score (black line) than those with PI> 1. Indeed, when inspecting all 39 patients, those with PI< 1 have a mean of M=28.62 (SD=6.51), while this is M=41.36 (SD=6.98) for those with PI> 1. Interestingly, the sum scores of selected patients appear to be primarily characterized by a lack of positive mood (i.e., high scores on reverse coded positive items; blue line in Figure 9.5), and only secondarily by a surplus of negative mood (red line in Figure 9.5), except for patient 6.

From inspecting the networks in Figure 9.5 and table 9.1, it is hard to tell which characteristics of the networks explain the height of the accompanying PI. To investigate this, we correlated PI with six network properties. We found only very weak correlations between PI and (1) average degree ($r = -.03$), (2) the average shortest path length (spl; the average number of steps on the shortest path between any pair of nodes; $r = .02$), (3) the number of edges within negative mood items (down, irritated, lonely, anxious, and guilty; $r = -.03$), (4) the number of edges within positive mood items (relaxed, satisfied, enthusiastic, and cheerful; $r = -.10$), (5) number of edges between negative and positive mood items ($r = .01$), and (6) global clustering coefficient (based on the number of *triangles*, also called *closed triplets*; $r = .11$).

Zooming in on the *node degree* (i.e., the number of edges of a node), feeling guilty was the only node with a higher, though still weak correlation with PI ($r = .29$); the other nodes were correlated very weakly with PI ($|r| < .16$). A tentative interpretation could be that in patients with more infectious networks, feeling guilty plays an important role; in patients with high PI, feeling guilty affects or is affected by other moods compared to patients with low PI.

Although percolation threshold ρ_c is unknown, we still performed one-sample t-tests as an illustration of how PI can be tested against a fixed value, in this case the value of 1. One-sample t-tests showed that only PI of patients 1 was smaller than the percolation threshold of 1, whereas PIs of other patients did not significantly differ from 1 at the 0.05 level (see diagonal of Table 9.2). Comparison of PIs among patients revealed that patient 4 had a significantly higher PI that that of patient 1 and 2. All other patients did not differ significantly. This could indicate

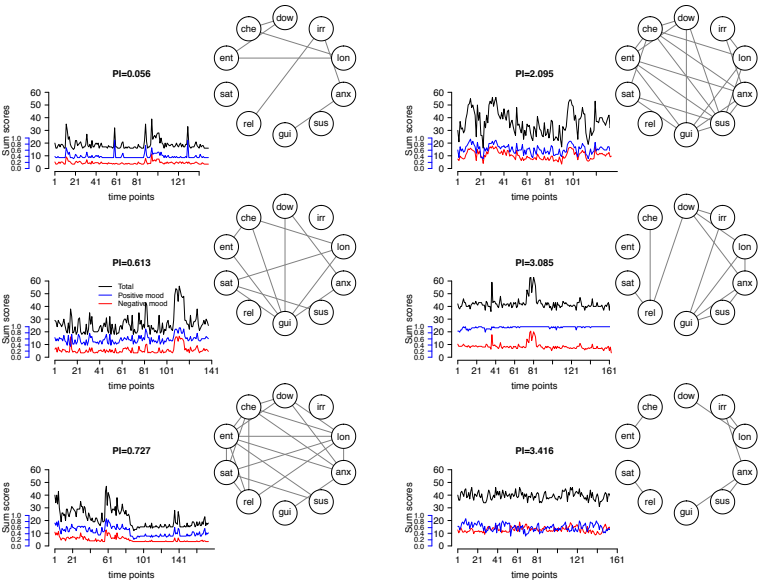

FIGURE 9.5. Real data of six patients. The figure represents results for three patients with PI< 1 (left) and three with PI> 1 (right). Displayed are sum scores on each time point (black) and separate sum scores of positive (blue) and negative (red) mood items. The resulting networks are depicted with the following abbreviations: dow – down, irr – irritated, lon – lonely, anx – anxious, sus – suspicious, gui – guilty, rel – relaxed, sat – satisfied, ent – enthusiastic, che – cheerful.

TABLE 9.1. PI and centrality measures. Average degree, average shortest path length (spl), and the degree of node "guilty" for six selected patients: three with PI < 1 (patient 1 through 3) and three with PI > 1 (patient 4 through 6).

patient	PI	average degree	average spl	degree 'guilty'
1	.06	1.4	7.73	1
2	.61	2.2	4.87	6
3	.73	4.0	3.22	1
4	2.10	4.2	3.18	6
5	3.09	2.4	3.69	4
6	3.42	1.4	6.98	1

that, in general, the variance in $\hat{\rho}$ is very large. Since the variance of $\hat{\rho}$ is based on the sample variance (i.e., based on the estimates of $\hat{\rho}$ per node; see [Van Borkulo et al. (2017)]), this could imply that the separate estimates of ρ are more diverse than would be expected according to contact process model in which there is one $\hat{\rho}$ for the network as a whole.

TABLE 9.2. Results of t-tests. The diagonal contains p values of (one-sided) one-sample t-tests (H_0: $\rho_i = 1$). The off-diagonal contains p values of two-sample t-tests (H_0: $\rho_i = \rho_j$). For comparison between two high or two low PI patients, a two-sided test was performed. Patients 1, 2, and 3 have $PI < 1$, whereas patients 4, 5, and 6 have $PI > 1$.

	1	2	3	4	5	6
1	.008					
2	.107	.116				
3	.478	687	.492			
4	.002	.004	.081	.084		
5	.846	.859	.875	.956	.465	
6	.122	.127	.138	.435	.877	.256

Estimation of PI is based on the estimated network and the data with a certain number of observations that can differ for each individual. As network estimation converges to the *true* underlying network structure with increasing number of observations (Abegaz & Wit, 2013), it is of importance to investigate the effect of network mis-specification on PI estimates. We evaluated stability of PI estimation under conditions similar to our real data, by generating networks (i.e., true networks) and data according to Yin and Li (2011). We generated 1000 temporal and contemporaneous networks with 10 nodes, using a constant of 1.1, 50% negative edges. Density of the contemporaneous network was set to .3, density of the temporal network was set to .1, .3, and .5. Number of simulated time points was 200, resulting in 1000 data sets. Data sets were dichotomized by splitting on the mean per variable, to comply to the method in this chapter to estimate PI. Sensitivity to network mis-specification on PI estimates was assessed by inspecting the difference between PI estimates based on the true network (PI_{true}) and the estimated network (PI_{est}). These simulations concern true contact processes and, therefore, closed processes. Moreover, the same types of networks are used in each condition. Therefore, we used the original estimator of PI without adjustments.

Figure 9.6 shows that differences between PI_{true} and PI_{est} are overall small ($M = .34, SD = .31$). For temporal networks with the lowest density (.1), sensitivity

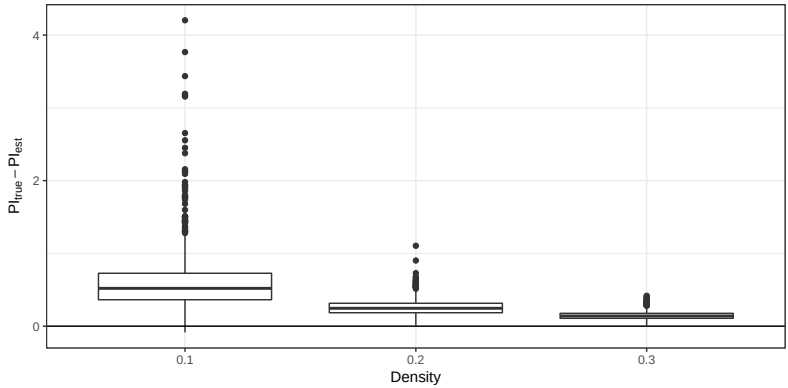

FIGURE 9.6. Boxplots of differences between PI estimates based on the true and estimated networks. Temporal networks were generated with varying densities (.1, .3, and .5).

for network mis-specification is somewhat increased, since the difference between PI_{true} and PI_{est} is rather high ($M = .60$) with a large range of difference scores ($SD = .41$). For higher densities (.3 and .5), the difference between PI_{true} and PI_{est} is low ($M = .15, SD = .06$). To conclude, sensitivity to network mis-specification is small. However, for temporal networks with low density (.1), caution should be exercised when drawing conclusions about real data.

9.6 Discussion

In this chapter, we presented the contact process model as the first implementation of a method that, according to percolation theory, is able to predict the dynamics of an individual′s mental disorder in the long run. We proposed the Percolation Indicator (PI) as a statistic to assess whether the symptom network is likely to stay infected or will plausibly recover. The present research indicates that the proposed estimation method performs well: the estimate of PI shows little bias and overall low variance across a variety of network architectures and parameterizations. Importantly, the model does not assume stationarity and equidistant measurements, as customary in many alternative frameworks, which makes it highly useful for data collection methods that characteristically lead to

unequally spaced measurement points, like ESM (Bolger et al., 2003; Larson & Csikszentmihalyi, 1983). Furthermore, estimation of PI seems to be little affected by misspecification of the network.

To illustrate the utility of the methodology, we provided an illustrative application to an ESM data set obtained in a clinical trial. To handle real data, we made two adjustments to the model. First, we adjusted for external influences which are not accommodated in the contact process model, since the latter is about closed systems. According to the contact process model, a symptom in a network can only be infected by one of its infected neighboring symptoms. Consequently, the model does not allow for symptoms to get infected, without being preceded by an infected neighboring symptom. In real data, however, it seems unlikely that all relevant nodes are included in the network, and so a node could change to 1 without its neighbors in the obtained network being infected. Therefore, to estimate PI, we only count upward jumps if they are preceded by at least one infected neighbor, as opposed to counting all upward jumps according to the original contact process model. The second adjustment concerns the effect of density of the estimated network on PI. Since the number of infected neighbors — which is in the denominator of PI — is generally higher in more densely connected networks, PI will be lower. This results in incomparable PIs. To control for this effect of density on PI, we divided the number of infected neighbors by the average degree of the network (2 x number of edges / number of nodes), which is related to density. With these adjustments, PIs are more comparable and the assumption that $\rho_c = 1$ is more justified.

The application to real data revealed a substantial amount of participants with PI> 1, which seems a plausible result, given that the data was collected from depressed patients. We developed a one sample t-test for testing whether the percolation indicator is above or below a critical percolation threshold. Applying the method to a subsample of three patients with PI < 1 and three patients with PI > 1 showed that only the first patient with the lowest PI has a value significantly different (lower) than 1. Thus, our model would predict that activity will eventually die out in the network of the first patient, whereas it is inconclusive whether activity will remain or die out in the long run in the network of the other patients.

Whether PI indeed has predictive power, needs to be investigated. To this end, one would need ESM data (to estimate the network structure and PI) and a follow-up measurement (to investigate how well PI predicts how the disorder

evolves). Also, predictive power of PI should be evaluated against existing clinical measures. Examples of known predictors of recovery from depression are prior depression (M. B. Keller, 2003), suicidal ideation and parental report of problem behaviors (Rohde, Seeley, Kaufman, Clarke, & Stice, 2006), and brain connectivity characteristics (Patel et al., 2015). However, many of these clinical measures are unmodifiable. Identifying measures that are modifiable is of most interest, since clinical treatment could benefit from this. Ultimately, if PI turns out to have good predictive quality, investigating alterations of individual network structures could guide efforts to improve clinical treatment.

The model as presented here has some restrictions but could be extended in several ways. Besides the adjustment of two parameters of the model, we will discuss four extensions. First, our model involves only two parameters that apply to the network as a whole, which combine to one PI: infection rate λ and recovery rate μ. A possible alternative model involves more precise specification of parameter values of individual nodes. It is conceivable that each node has its own level of infectiousness. For example, depressive symptom feeling sad is more likely to affect loss of interest and fatigue than suicidal thoughts, whereas suicidal thoughts may be affected by feelings of worthlessness and guilt. That this could be the case in our real data, might be illustrated by the fact that PIs of our real patients do not differ significantly from each other. This must be the result of the relatively large variance of PI, which is based on the variance of the infection rate parameter per node. If PI is based on the infectiousness per node, instead of on the network as a whole, the PI should be modified accordingly, since the correspondence with the classic basic reproduction number (Kolaczyk, 2009) is then lost when the homogeneity across the connections is lost.

A second extension could relax the restriction that a node is equally likely to be infected by any of its neighbors (e.g., feeling sad affects loss of interest and fatigue to the same extent). It is conceivable that the probability of being infected is different for different neighbors (e.g., feeling sad may affect loss of interest more than fatigue). Note, however, that although alternative models could have a better fit — a preliminary analysis with the real data, for example, in which we compared the fit of our original model with the alternative model with separate PIs for each node, revealed that the alternative model fitted best in most of the cases (results not shown here) — it is unclear to what extent this matters for prediction. Estimating more parameters clearly wins one flexibility, but may also

result in estimation problems as more parameters have to be estimated from the same data. Further research is needed to establish the relative merits of more versus less restricted models in practical situations.

Third, the proposed method assumes networks to be stable over time. For disorders characterized by stable patterns of dysfunction, like (untreated) personality disorders, this may not be problematic. However, in less chronic or episodic disorders, such as those related to depression or anxiety, it is likely that psychopathology changes over time and network structures will change accordingly. In such cases, it may be important to include evolution of the model architecture over time in the estimation method; rather than a single value, we may obtain a pattern of PI values as they evolve over time. Such a pattern may yield important information that could be useful in treatment selection: in time frames where the PI is close to its tipping point, a small nudge to one or two elements in the network may be enough to push the system in an alternative stable state.

A fourth extension involves including external factors in the model. As presented in this chapter, the model is about a closed system that does not take external influences (e.g., stressful life events or therapy) into account. One could, for example, measure a variable called *stressful event*. The resulting network will reveal which symptoms or affective responses are influenced by external factors (i.e., stressful events), given that the network structure is stable during the time of measurement. According to the network perspective on psychopathology, however, changing a patient′s network structure could be the aim of an intervention. Including beginning of treatment in the network is not trivial, though, if one is willing to assume that treatment might change network structure. In that case, one could consider measuring before and after treatment. Consequently, two networks can be estimated and accompanying PIs can be calculated and compared. If one is interested in how each session influences the internal system of affective responses or symptoms, one could model *therapy session* as variable. One way to model this external factor is, when taking weekly therapy sessions as an example, to assume that when an individual gets therapy, the dichotomous therapy session variable takes on the value 1 for a certain time. If one is willing to assume that effect of therapy is present for the next three days, than the therapy variable takes on the value 1 during all measurements in the three days following the moment of therapy. After three days the therapy variable is assigned the value 0 (no therapy). Furthermore, for calculating PI, the therapy node can only be

counted as an infected neighbor that is capable of infecting a connected node, but the state of the therapy node itself cannot contribute to the calculation of PI.

To conclude, more knowledge on characteristics of individuals′ networks, such as the PI, might reveal important information about person-specific clinical disorder characteristics. One suggestive hypothesis is that patients with persistent disorders have more infectious network structures and, therefore, do not recover as easily as others. From our real patient data, we could hypothesize that the infectiousness is, at least partly, determined by feeling guilty as this mood state plays a more important role in persisters; the number of temporal associations of feeling guilty with other mood states is related to PI and, therefore, may be predictive of future course. Information on PI and the specific structure of a patient′s symptom dynamics may also guide clinical therapy, since important symptoms in the infectious network could be targeted using micro-interventions. Intervening on an important symptom or connection could alter the network structure, thereby lowering the infectiousness of the system as a whole. From our real data application, feeling guilty seems a plausible candidate. However, additional research with ESM data at symptom level is required to substantiate these hypotheses. Currently, most (if not all) ESM data concern affective responses or mood states, as the present empirical example does. Although mood and affect dynamics are considered to be indicative of future course of a mental disorder (Wichers, 2014; Wichers et al., 2009; Wichers, Wigman, & Myin-Germeys, 2015), more research is needed to demonstrate this. Thus, in the future, applying our model to DSM criteria (American Psychiatric Association, 2013) might steer us to new ideas pertaining to clinical definitions of disorders.