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

Kossakowski, Jolanda J.; Gordijn, Marijke C. M.; Riese, Harriette; Waldorp, Lourens J.

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

Jolanda J. Kossakowski[1]*, Marijke C. M. Gordijn[2], Harriëtte Riese[3] and Lourens J. Waldorp[1]

[1] Department of Psychology, University of Amsterdam, Amsterdam, Netherlands, [2] Department of Chronobiology, GeLifes, University of Groningen, Groningen, Netherlands, [3] Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion Regulation, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

Mental disorders like major depressive disorder can be modeled as complex dynamical systems. In this study we investigate the dynamic behavior 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 categorized 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.

Keywords: cellular automata, discrete dynamical systems, maximum likelihood estimation, nonlinear dynamics, psychopathology

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 paper, 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 and Cramer, 2018). 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 and 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; Cramer et al., 2016; Wichers et al., 2016; Kossakowski and Cramer, 2018). 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 paper. 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 and Cramer, 2018). Dynamical systems leave "breadcrumbs" behind in these time series that hint toward 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 behavior. There are three reasons why we believe that the dynamics of the 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 neighbors 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 characterizing 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 for a person it is possible that a transition may occur or not.

As an example, we consider a time series of the proportion of active emotions for a single subject, shown in **Figure 1** (left). 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 1**, right). 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 **Figure 1**, right). 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 1**). Based on this, we would classify this individual as someone who may expect a (sudden) increase in the
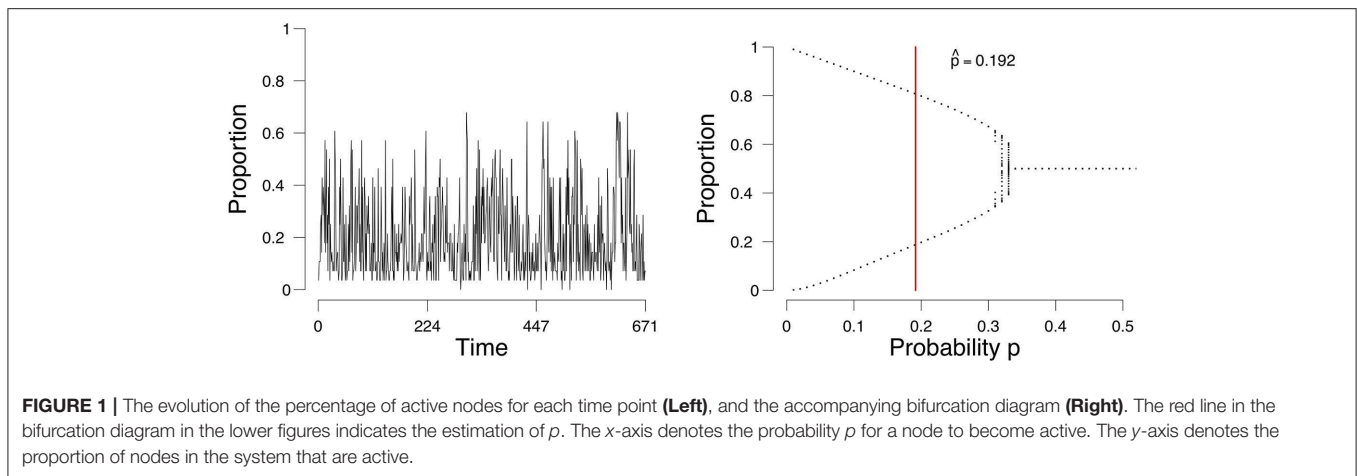
FIGURE 1 | The evolution of the percentage of active nodes for each time point **(Left)**, and the accompanying bifurcation diagram **(Right)**. The red line in the bifurcation diagram in the lower figures indicates the estimation of p. The x-axis denotes the probability p for a node to become active. The y-axis denotes the proportion of nodes in the system that are active.
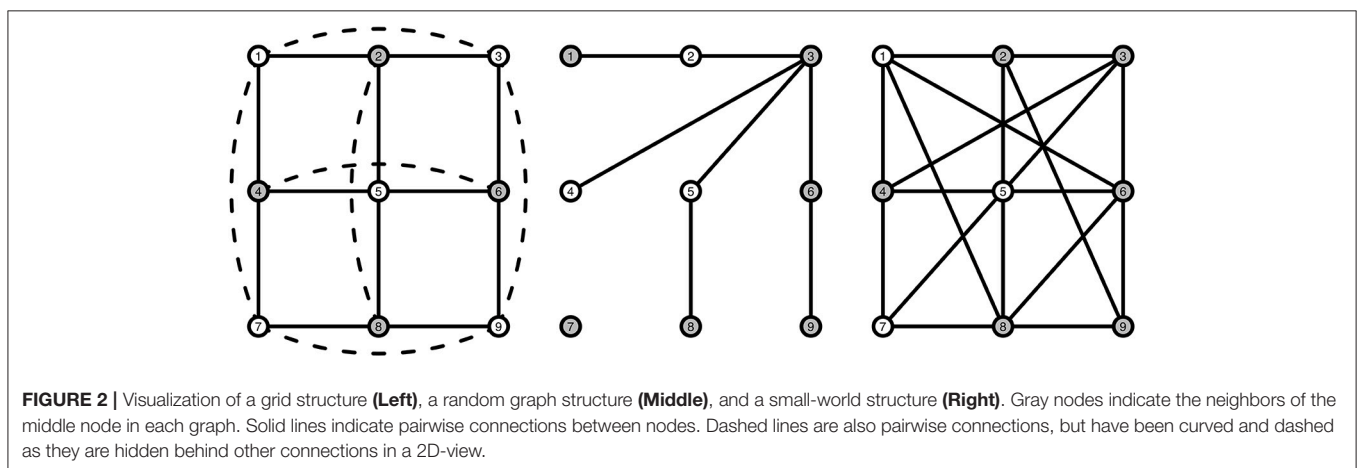

FIGURE 2 | Visualization of a grid structure **(Left)**, a random graph structure **(Middle)**, and a small-world structure **(Right)**. Gray nodes indicate the neighbors of the middle node in each graph. Solid lines indicate pairwise connections between nodes. Dashed lines are also pairwise connections, but have been curved and dashed as they are hidden behind other connections in a 2D-view.

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 paper 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 data sets, one with patients diagnosed with MDD, and one with subjects from the general population. This paper is set up as follows. First, we will briefly explain the theory of the mean field model and the proposed method. Then we present the simulation to show how the ML estimation performs. Finally, we apply our method to two datasets to show how the method works in different contexts.

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 behavior for stable and bistable behavior 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 and Katok (2003), and Golubitsky and 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, 1984; Sarkar, 2000). A node j that is directly connected to node i is called a neighbor. A grid is a graph $G_{grid}(n, \Gamma)$ with n nodes in the set $V = \{1, 2, \ldots, n\}$ where each node i has the same number of neighbors in its neighborhood $\Gamma = \{j \in V : j \text{ is connected to } i\} \cup \{i\}$ including itself. To ensure that all nodes have exactly the same number of neighbors, 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 2** (left), where the center node is directly connected to its four neighbors, marked in gray. 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

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

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 neighbors for per node at time t.

active in the neighborhood 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 neighbors 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, i.e., for small changes in the number of active nodes the decision does not change (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 applied to the result of the previous time point t times.

To illustrate, say that we have the network presented in **Figure 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 1**. We can then determine how many active neighbors r each node has, by just counting the number of active nodes each node is connected to. As mentioned in **Table 1**, nodes 1, 5, and 7 have three active neighbors, nodes 2, 3, 4, 8, and 9 have two active neighbors, and node 6 has one active neighbor. 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 neighborhood is active. The majority rule uses $r > |\Gamma|/2$ to indicate whether the number of active neighbors is greater than half the size of the neighborhood. Γ here denotes the size of a node's neighborhood. In our example, $|\Gamma| = 5$: each node has exactly four neighbors, 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 **Table 1**, column r_1. We can continue this process for a length T (not shown in **Table 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 neighbors (r). In our application to psychopathology, this uncertainty is required because we cannot

predict the behavior 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 neighbors 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 neighbors is

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

where $|\Gamma|$ is the size of the neighborhood and r the number of active neighbors. 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 behavior of the majority of a node's neighborhood, 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 0 s and 1 s. 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 1** (left).

3. MEAN FIELD MODEL

It is rather difficulty 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 neighbors 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 (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 2**, left) it is easily seen that each node is similar to any other node since each node has the same number of neighbors, and becomes active or inactive in the same way by means of the majority rule that is based on the number of active neighbors. 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 and Kossakowski (2019).

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 neighbors in that node's neighborhood (r) and the probability parameter p, following the majority rule defined in (1). In a grid each node has exactly the same number of neighbors and so the probability of a node changing value depends on the properties of the grid and not on
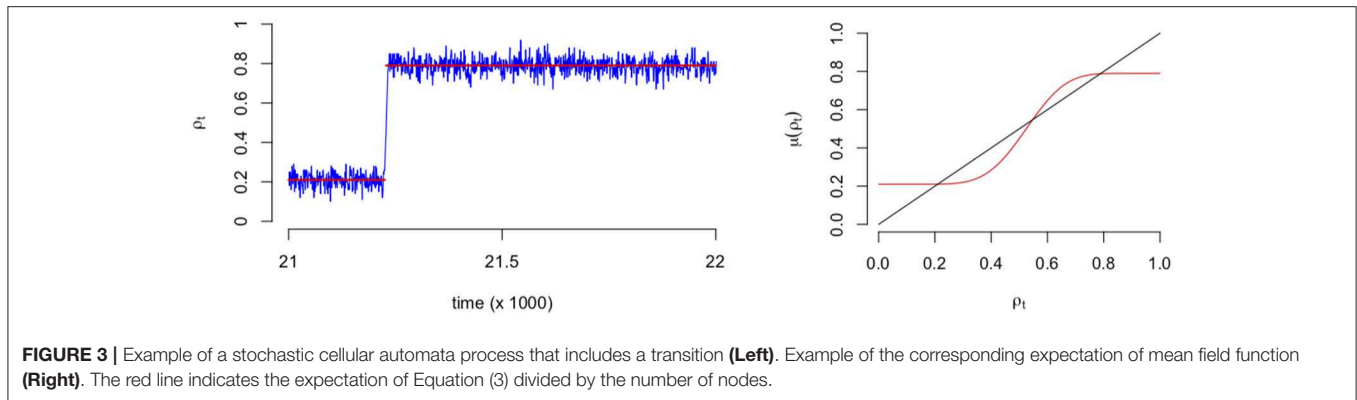
FIGURE 3 | Example of a stochastic cellular automata process that includes a transition **(Left)**. Example of the corresponding expectation of mean field function **(Right)**. The red line indicates the expectation of Equation (3) divided by the number of nodes.

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 neighborhood $|\Gamma|$ particular to the graph. The majority rule in (1) determines for which number of active nodes we obtain p up until 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 neighbors that are 1, we need to consider all possible configurations of $|\Gamma|$ 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 $|\Gamma|$

$$\mathbb{P}(r \mid \rho_t) = \binom{|\Gamma|}{r} \rho_t^r (1 - \rho_t)^{|\Gamma|-r} \qquad (2)$$

Simultaneously, we require the probability p or $1 - p$ from the majority rule in (1), which is assumed to be independent. We need to define the probability for any number of active nodes and therefore marginalize over the number of possible active nodes in the neighborhood r. Putting (1) and (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

$$\rho_{t+1}^{\text{grid}} = \sum_{r=0}^{|\Gamma|} \mathbb{P}(X_{i,t+1} = 1 \mid r) \binom{|\Gamma|}{r} \rho_t^r (1 - \rho_t)^{|\Gamma|-r} \qquad (3)$$

Because the evolution is binomial based on the proportion of active nodes at the previous time point (see Equation 3), it follows from the transition probability that the number of active nodes $X_{t+1} = x_{t+1}$, given that at t is $X_t = x_t$ in the graph G_{grid}, is

$$\mathbb{P}(X_{t+1} = x_{t+1} \mid X_t = x_t) = \binom{n}{x_{t+1}} \rho_t^{\text{grid}}(x_t/n)^{x_{t+1}}$$
$$(1 - \rho_t^{\text{grid}}(x_t/n))^{n-x_{t+1}} \qquad (4)$$

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 emphasize 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 and Kossakowski, 2019), so the mean is a good approximation of the process itself.

As an illustration of the binomial process, in the left panel of **Figure 3** we see a typical SCA process where it is clear that the fluctuations are around a particular mean (0.2) for time points before $t = 21,000$ approximately. After this point (tipping point) the fluctuations revolve around another mean (0.8) with a higher proportion of emotions. In the right panel of **Figure 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 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 parameter p from (3) in the majority rule, in the interval $(0, 0.5]$, we obtain a *bifurcation diagram*, as shown in **Figure 1** (right). 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 in equilibrium (stable if stable fixed points exist; Hirsch et al., 2004). For each value of p, displayed on the x-axis in **Figure 1**, one sequence is generated, of which the last 50 are displayed in **Figure 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 behavior 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 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 (3) is designed for a grid with a fixed neighborhood 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 $G_{\mathrm{rg}}(n, p(e))$ 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 neighborhood size $|\Gamma|$ is a random variable that is maximally $n - 1$. Each node has a binomial number of neighbors with expected number of neighbors $p(e)(n - 1)$. We extend the idea used for the grid, where we marginalize over all possible configurations of number of active nodes for each neighborhood 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 and Kossakowski, 2019). The difference with the probability on the grid is in the size of the neighborhood (see **Figure 2**, left and middle panel), where in the grid the neighborhood size is fixed to $|\Gamma|$. In the mean field model for the random graph, we fix this to the expected number of neighbors $p(e)(n - 1)$. Let $\nu = \lfloor p(e)(n - 1) \rfloor$ be the integer part of the expected number of neighbors. For the random graph G_{rg} the neighborhood 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 and Kossakowski, 2019):

$$\rho_{t+1}^{\mathrm{rg}} = \sum_{r=0}^{\nu} \mathbb{P}(X_{i,t+1} = 1 \mid r) \binom{\nu}{r} \rho_t^r (1 - \rho_t)^{\nu - r} \quad (5)$$

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 and Strogatz, 1998). A modified version of the small world is the Newman-Watts (NW) small-world (Newman and Watts, 1999). In the NW small-world $G_{\mathrm{sw}}(n, \Gamma, p(w))$ the n nodes each have a neighborhood Γ as in the grid and edges are added to the network following a (constant) wiring probability $p(w)$ (Newman and Watts, 1999). 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 $\rho_t^{\mathrm{grid}}(1 - p(w))^{n - |\Gamma|}$, 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 (5) but the first $|\Gamma|$ edges left out,

because they have already been accounted for by the grid part. We denote this probability by $\rho_t^{\mathrm{rg},\Gamma}$, which denotes the probability as in (5) but with $\rho_t^{\mathrm{rg},\Gamma}$ 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 in the small-world graph G_{sw} is

$$\rho_{t+1}^{\mathrm{sw}} = \frac{|\Gamma|}{n} \rho_t^{\mathrm{grid}}(1 - p(w))^{n - |\Gamma|} + \frac{n - |\Gamma|}{n} \rho_t^{\mathrm{rg},\Gamma} \quad (6)$$

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

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 (1). Here we use maximum likelihood (ML) to obtain an estimate of p (Rajarshi, 2012).

If we take a closer look at Equation (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 (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 (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

$$\log \mathbb{P}(X_t, t \geq 0) = \sum_{t=0}^{T-1} \log \mathbb{P}(X_{t+1} = x_{t+1} \mid X_t = x_t) \quad (7)$$

where T denotes the total duration of the sequence in time points. The transition probability $\mathbb{P}$ is as in (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 t. The data are plugged in the transition probabilities, where we recognize 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 and Kossakowski, 2019). 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 expression (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 and Harrington, 1978). We maximize the log-likelihood function in (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 maximizing the log-likelihood with respect to $p(e)$ and $p(w)$ respectively. Equations (3), (5), and (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, respectively) 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**. Each dot represents a separate run of the mean field equation. Equation (3) is reflected in the top panel of **Figure 4**, (5) in the middel panel, and (6) in the bottom panel of **Figure 4**.

Taking the top panel of **Figure 4** as an example, if we run Equation (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 $p(w)$, respectively, as seen in **Figure 4**. To summarize, in order to categorize 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 personalized 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 have only the estimate of p and for the random graph and the small-world we have the edge probability $p(e)$ and $p(w)$, respectively. Standard errors are obtained from

FIGURE 4 | Examples of bifurcation diagrams for a grid **(Top)**, a random graph (**Middle**; $p(e) = 0.1$) and a small-world graph (**Lower**; $p(w) = 0.1$). 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.
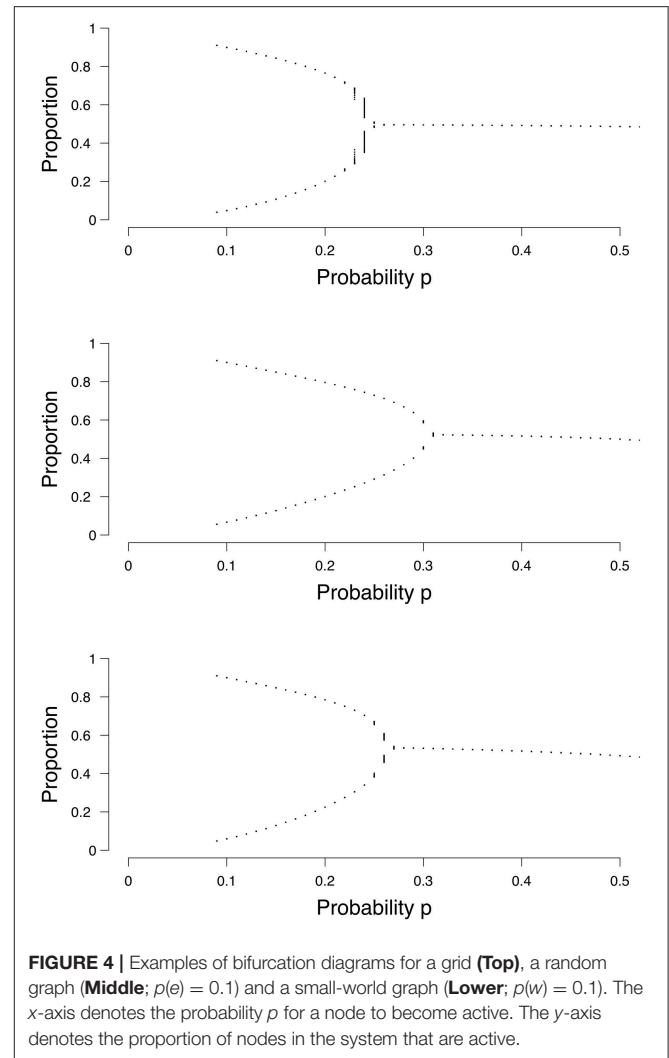
the second order derivatives (Hessian) of the log-likelihood (Rajarshi, 2012). The inverse (matrix) of the Hessian and scaled by $1/T$ results in the variance of the parameter estimate. The square root of the diagonal elements are the standard errors, i.e., $\text{SE}(\hat{p}) = \sqrt{\frac{1}{T}h^{11}}$ 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, where h^{ij} is the ijth element of the inverse Hessian.

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 $T \in \{ 50, 100, 200, 500, 5,000 \}$, and the probability $p \in \{ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 \}$ (Equation 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 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 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, 2018). 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 paper less interesting. These and other results can be found online (Kossakowski, 2018). 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**.

Figure 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 Table 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 $T = 5,000$, $n = 100$, $p = 0.4$), 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.
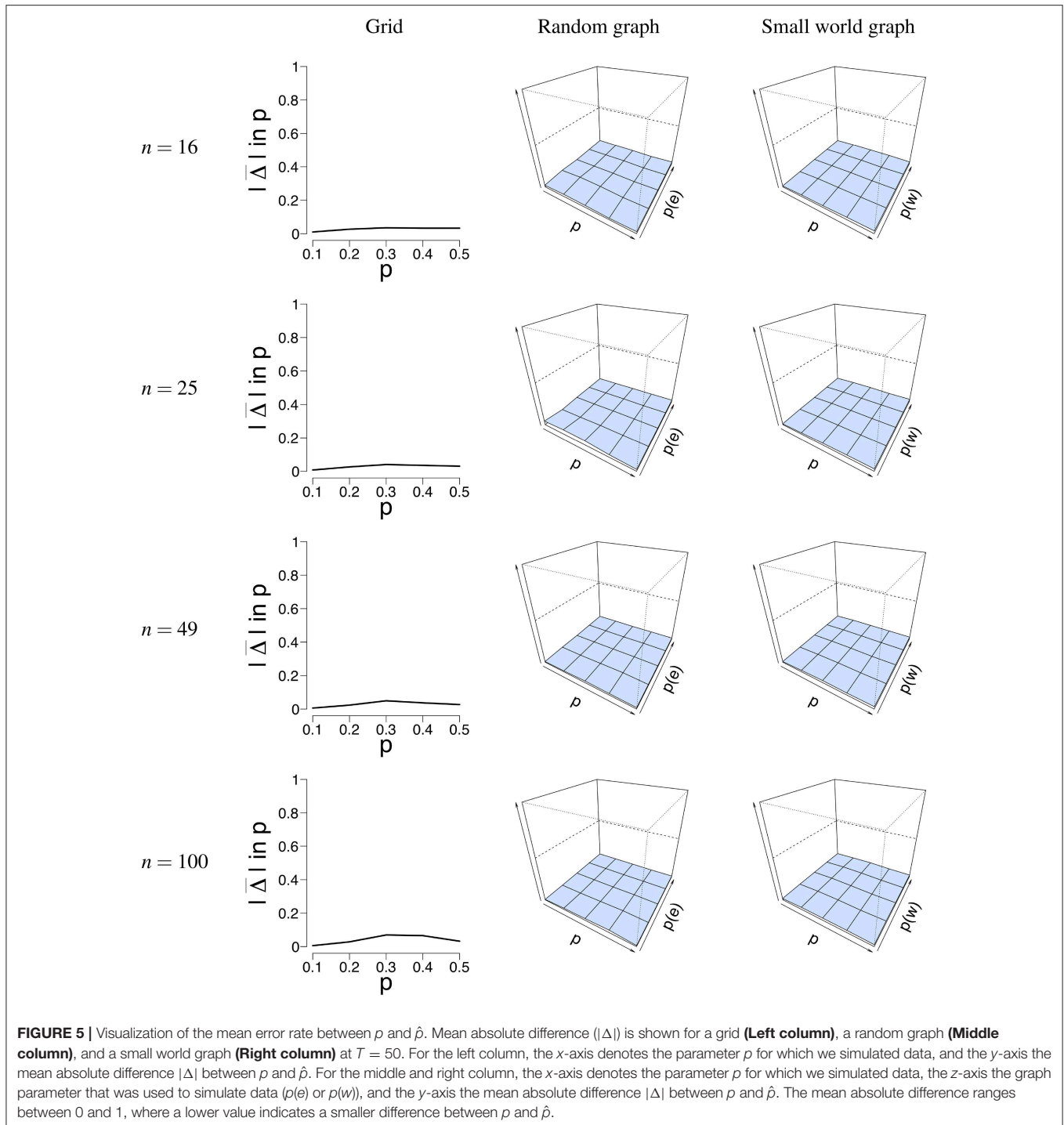
The same conclusion cannot be drawn for graph parameters $p(e)$ and $p(w)$, as seen in **Figure 6**. For a random graph, $\overline{\Delta}(p(e))$ is high when $p(e)$ is low, and decreases as $p(e)$ is increased. This shows that the graph parameter $\hat{p}(e)$ is most accurate when $p(e)$ is high. A 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 $p(e)$, 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 a small-world graph, where $\overline{\Delta}(p(w))$ is high when $p(w)$ is high and p is low, and that it 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** 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 $T = 5,000$ 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 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 $T = 5,000$ nodes for clarity of presentation. We estimated $\hat{p}$ for all three network structures for all datasets. It can be seen in **Figure 7** that $\overline{\Delta}(p)$ is generally low, regardless of the network structure that was used to simulate the data.

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

FIGURE 5 | Visualization 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 $\hat{p}$.
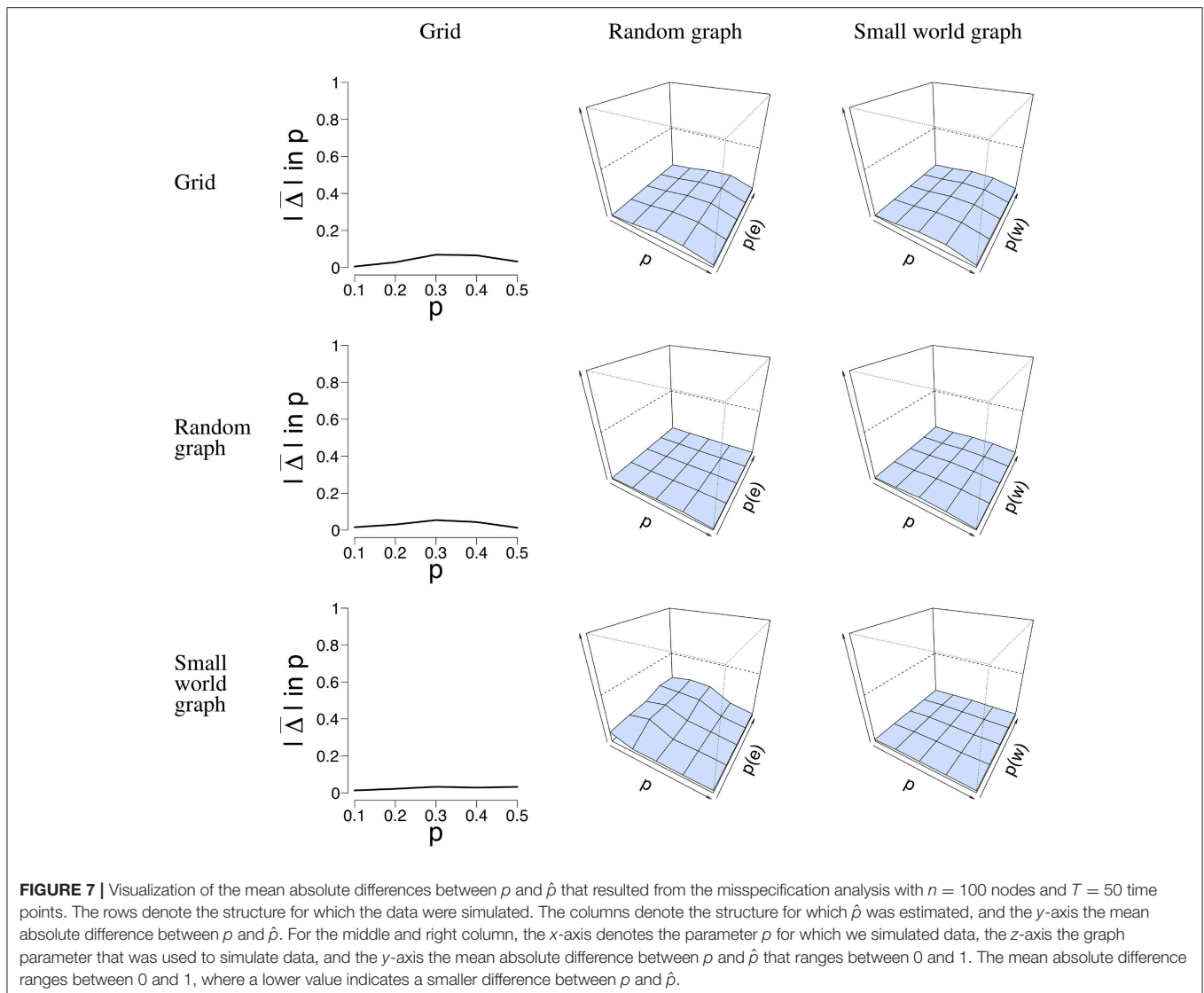
may be difficult to distinguish between the network structures. For data simulated under the small-world graph, the small-world graph itself is most often selected based (69.3% of the cases across conditions) on the BIC. There are no conditions in which the random graph is preferred over the small-world 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**, 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

FIGURE 6 | Visualization 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 7 | Visualization 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, 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 $\hat{p}$.
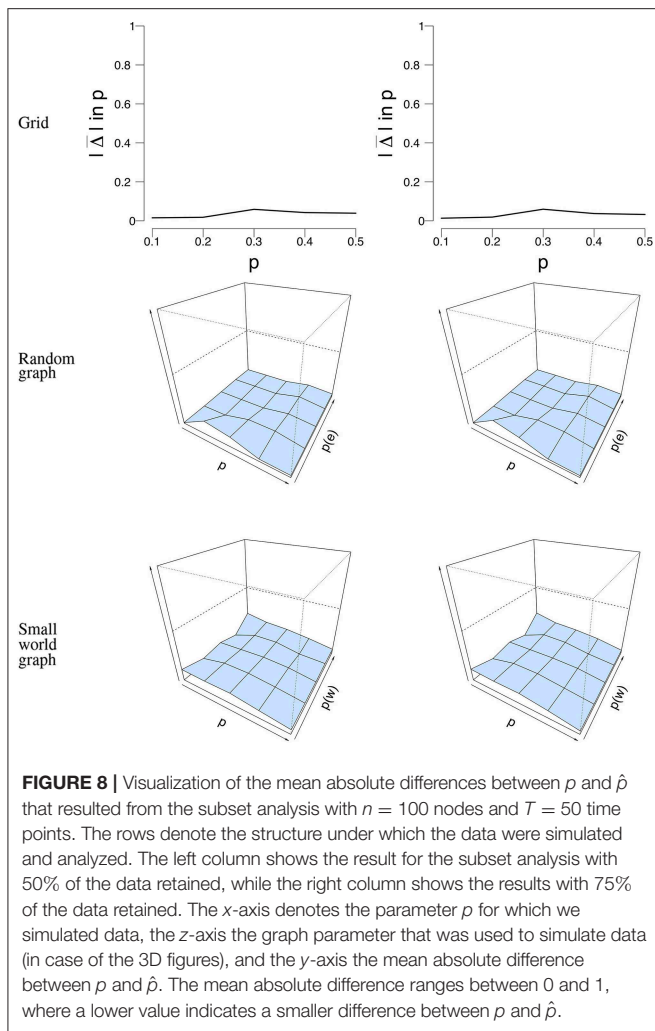
at data with $n = 100$ nodes and $T = 50$ time points, and data with $n = 16$ nodes and $T = 5,000$ time points. For each simulation condition, we randomly selected one simulation and selected a subset of the data, which we repeated 100 times. **Figure 8** shows the mean error rate $\overline{\Delta}(p)$ 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 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.

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

FIGURE 8 | Visualization 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 analyzed. The left column 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 $\hat{p}$.

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:

$$BC = \frac{s^2 + 1}{k + C} \tag{8}$$

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 method 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.

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 6 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 2**.
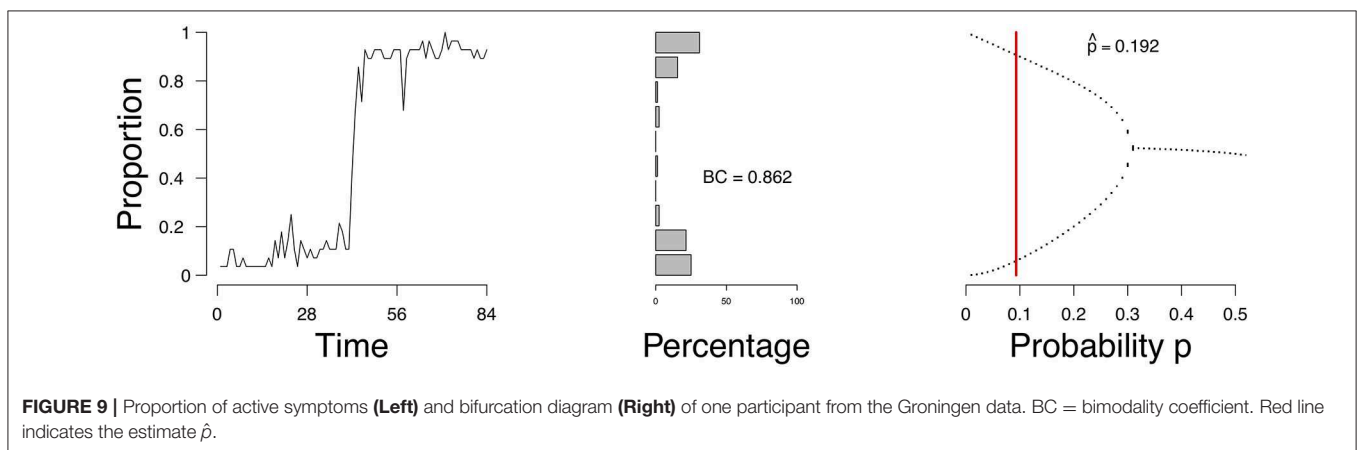
We dichotomized the data by collapsing the "neither" condition with the positive mood state per individual item. We coded the positive mood state as "0" and the negative mood state as "1." We also collapsed the "neither" condition with the negative mood state and ran the analyses with these data, but as these results were very similar to the ones we present, we left it out of this study. After dichotomizing 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

Item	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

*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 te items.*

FIGURE 9 | Proportion of active symptoms **(Left)** and bifurcation diagram **(Right)** of one participant from the Groningen data. BC = bimodality coefficient. Red line indicates the estimate $\hat{p}$.

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, 2018).

TABLE 3 | Items that were included in the analysis, the meaning of each item, and the response range in word and number.

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)

Figure 9 shows the evolution of the density (left), a distribution of the density ρ_t (frequency of the number of active nodes; middle figure), and the estimate of $\hat{p}$ in the bifurcation diagram (right) 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 MFA 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 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.

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 6-h 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 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 dichotomize 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
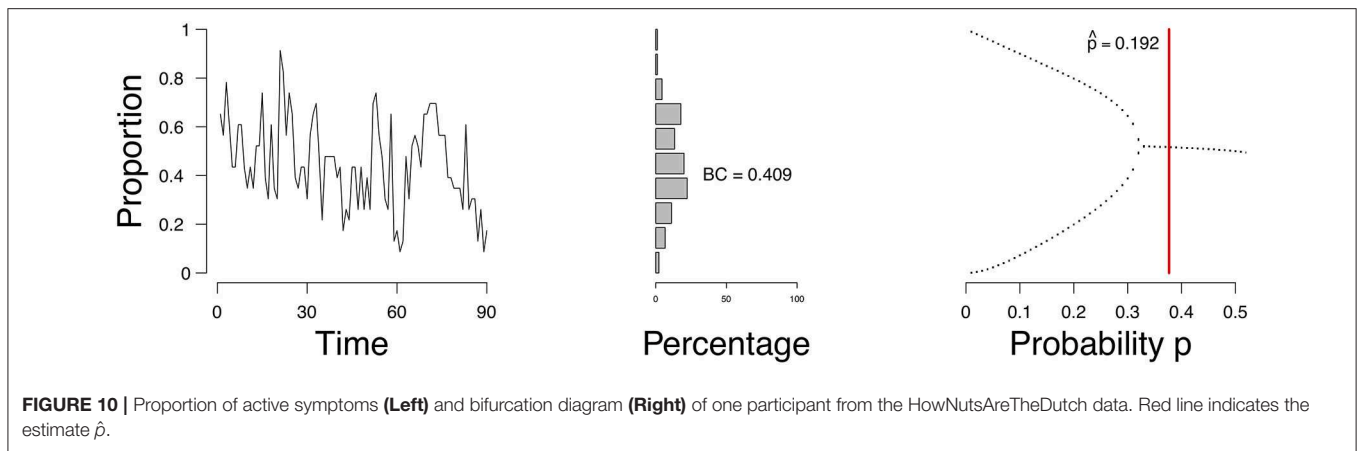
FIGURE 10 | Proportion of active symptoms **(Left)** and bifurcation diagram **(Right)** of one participant from the HowNutsAreTheDutch data. Red line indicates the estimate $\hat{p}$.

also considered removing the missing measurements entirely, but as we found nearly identical results, we chose not to present these results.

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 ($W = 75,248$, $p = 0.353$). We also looked at the mean scores of the *Depression and Anxiety Stress Scale* (DASS; Lovibond and Lovibond, 1995a,b), 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 ($W = 10,093$, $p = 0.194$), the QIDS ($W = 74,275$, $p = 0.127$), the positive items of the PANAS ($W = 67,314$, $p = 0.557$) or the negative items of the PANAS ($W = 72,253$, $p = 0.366$). 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, 2018).

Figure 10 shows the evolution of the density (left figure), a distribution of the density (frequency of the number of active nodes; middle figure), and the estimate of $\hat{p}$ in the bifurcation diagram (right figure) 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 31.9% of the participants as being bimodal. When we compare the results from the MFA 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 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.

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 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 hypothesized 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 toward 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 30 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 2** and **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. 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 analyzed, 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 analyses.

The current study only allows for binary and no missing data. We applied different techniques for dichotomizing 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 mean 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 paper has three assumptions: (1) we assume that each node in a graph has the same neighborhood 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 behavior. Waldorp and Kossakowski (2019) 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 operationalized 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 paper 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 operationalize 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 behaviors that are beyond the scope of this paper. 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 emphasize 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.

DATA AVAILABILITY

The datasets for this study will not be made publicly available because Under the EU General Data Protection Regulation, we are not allowed to publish raw results for all dataset that are discussed in the manuscript. Result figures for all participants are available online.

ETHICS STATEMENT

This study involves a secondary data analysis. Informed consent was obtained during the data collection for the primary data analysis. See the following two references for more information about the ethics approval: Gordijn et al. (1994, 1998) and van der Krieke et al. (2015).

AUTHOR'S NOTE

A pre-print of this manuscript is published at http://www.jolandakossakowski.eu/wp-content/uploads/2019/07/20190603MFA.pdf.

REFERENCES

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5), 5th Edn*. Washington, DC: American Psychiatric Association.

Balister, P., Bollobás, B., and Kozma, R. (2006). Large deviations for mean field models of probabilistic cellular automata. *Rand. Struct. Algor.* 29, 399–415. doi: 10.1002/rsa.20126

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

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry* 16, 5–13. doi: 10.1002/wps.20375

Borsboom, D., Epskamp, S., Kievit, R. A., Cramer, A. O., and Schmittmann, V. D. (2011). Transdiagnostic networks commentary on Nolen-Hoeksema and

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01762/full#supplementary-material

Watkins (2011). *Perspect. Psychol. Sci.* 6, 610–614. doi: 10.1177/1745691611425012

Cramer, A. O. J., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L. J., Kendler, K. S., Scheffer, M., et al. (2016). Major depression as a complex dynamical system. *PLoS ONE* 11:e0167490. doi: 10.1371/journal.pone.0167490

Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., et al. (2012). Dimensions of normal personality as networks in search of equilibrium: you can't like parties if you don't like people. *Eur. J. Pers.* 26, 414–431. doi: 10.1002/per.1866

Csikszentmihalyi, M., and Larson, R. (1987). Validity and reliability of the experience-sampling method. *J. Nerv. Ment. Dis.* 175, 526–536. doi: 10.1097/00005053-198709000-00004

Durett, R. (2007). *Random Graph Dynamics*. Cambridge: Cambridge University Press.

Epskamp, S. (2015). IsingSampler: Sampling methods and distribution functions for the Ising model. Available online at: https://CRAN.R-project.org/package=IsingSampler

Fleming, T. R., and Harrington, D. P. (1978). Estimation for discrete time nonhomogeneous Markov chains. *Stochast. Process. Appl.* 7, 131–139. doi: 10.1016/0304-4149(78)90012-1

Golubitsky, M., and Stewart, I. (2003). *The Symmetry Perspective: From Equilibrium to Chaos in Phase Space and Physical Space.* Basel: Springer Science & Business Media.

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

Gordijn, M. C., Beersma, D. G., Bouhuys, A. L., Reinink, E., and van den Hoofdakker, R. H. (1994). A longitudinal study of diurnal mood variation in depression: characteristics and significance. *J. Affect. Disord.* 31, 261–273. doi: 10.1016/0165-0327(94)90102-3

Guloksuz, S., Pries, L. K., and Van Os, J. (2017). Application of network methods for understanding mental disorders: pitfalls and promise. *Psychol. Med.* 47, 2743–2752. doi: 10.1017/S0033291717001350

Gulyás, L., Kampis, G., and Legendi, R. O. (2013). Elementary models of dynamic networks. *Eur. Phys. J. Spec. Top.* 222, 1311–1333. doi: 10.1140/epjst/e2013-01928-6

Hasselblatt, B., and Katok, A. (2003). *A First Course in Dynamics.* Cambridge: Cambridge University Press.

Hirsch, M. W., Smale, S., and Devaney, R. L. (2004). *Differential Equations, Dynamical Systems, and an Introduction to Chaos.* Oxford: Academic press.

Holmgren, R. (1996). *A First Course in Discrete Dynamical Systems.* New York, NY: Springer Science & Business Media.

Hosenfeld, B., Bos, E. H., Wardenaar, K. J., Conradi, H. J., van der Maas, H. L., Visser, I., et al. (2015). Major depressive disorder as a nonlinear dynamic system: bimodality in the frequency distribution of depressive symptoms over time. *BMC Psychiatry* 15:222. doi: 10.1186/s12888-015-0596-5

Janssens, K. A. M., Bos, E. H., Rosmalen, J. G. M., Wichers, M. C., and Riese, H. (2018). A qualitative approach to guide choices for designing a diary study. *BMC Med. Res. Methodol.* 18:140. doi: 10.1186/s12874-018-0579-6

Kossakowski, J. J. (2018). *Results "Applying a Dynamical Systems Model and Network Theory to Major Depressive Disorder".* Available online at: from https://osf.io/edyzp/

Kossakowski, J. J., and Cramer, A. O. J. (2018). "Complexity, chaos and catastrophe: modeling psychopathology as a dynamic system," in *Frontiers of Cognitive Psychology* ed M. Vitevitch (London: Taylor & Francis), 1–29.

Kossakowski, J. J., Groot, P. C., Haslbeck, J. M. B., Borsboom, D., and Wichers, M. (2017). Data from 'Critical slowing down as a personalized early warning signal for depression'. *J. Open Psychol. Data* 5, 1–3. doi: 10.5334/jopd.29

Kozma, R., Puljic, M., Balister, P., Bollobás, B., and Freeman, W. J. (2004). "Neuropercolation: a random cellular automata approach to spatio-temporal neurodynamics," in *International Conference on Cellular Automata* eds P. M. A. Sloot, B. Chopard, and A. G. Hoekstra (Berlin: Springer), 435–443. doi: 10.1007/978-3-540-30479-1_45

Kozma, R., Puljic, M., Balister, P., Bollobás, B., and Freeman, W. J. (2005). Phase transitions in the neuropercolation model of neural populations with mixed local and non-local interactions. *Biol. Cybern.* 92, 367–379. doi: 10.1007/s00422-005-0565-z

Krieke L. V., Jeronimus, B. F., Blaauw, F. J., Wanders, R. B., Emerencia, A. C., Schenk, H. M., et al. (2015). HowNutsAreTheDutch (HoeGekIsNL): a crowdsourcing study of mental symptoms and strengths. *Int. J. Methods Psychiatr. Res.* 25, 123–144. doi: 10.1002/mpr.1495

Kuznetsov, Y. A. (2013). *Elements of Applied Bifurcation Theory.* New York, NY: Springer-Verlag.

Lebowitz, J. L., Maes, C., and Speer, E. R. (1990). Statistical mechanics of probabilistic cellular automata. *J. Stat. Phys.* 59, 117–170. doi: 10.1007/BF01015566

Lovibond, P., and Lovibond, S. (1995a). The structure of negative emotional states - comparison of the depression anxiety stress scale (DASS) with the beck depression and anxiety inventories. *Behav. Res. Ther.* 33, 335–343. doi: 10.1016/0005-7967(94)00075-U

Lovibond, S., and Lovibond, P. (1995b). *Manual for the Depression Anxiety Stress Scales, 2nd Edn.* Sydney, NSW: Psychology Foundation.

Molenaar, P. C. (2007). On the implications of classic ergodic theorems: analysis of developmental processes has to focus on intra-individual variation. *Dev. Psychobiol.* 50, 60–69. doi: 10.1002/dev.20262

Newman, M. E. J., and Watts, D. J. (1999). Renormalization group analysis of the small-world network model. *Phys. Lett. A* 263, 4–6. doi: 10.1016/S0375-9601(99)00757-4

O'Donnell, R. (2014). *Analysis of Boolean Functions.* Cambridge: Cambridge University Press.

Peeters, F., Ponds, R., and Vermeeren, M. (1996). Affectiviteit en zelfbeoordeling van depressie en angst. *Tijdschrift voor de Psychiatrie* 38, 240–250.

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* Vienna: R Core Team. Available online at: https://www.R-project.org

Raes, F., Daems, K., Feldman, L., Johnson, S., and van Gucht, D. (2009). A psychometric evaluation of the Dutch version of the responses to positive affect questionnaire. *Psychol. Belgica* 49, 293–310. doi: 10.5334/pb-49-4-293

Rajarshi, M. B. (2012). *Statistical Inference for Discrete Time Stochastic Processes.* New Delhi: Springer.

Rush, A. J., Bernstein, I. H., Trivedi, M. H., Camody, T. J., Wisniewski, S., Mundt, J. C., et al. (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. *Biol. Psychiatry* 59, 493–501. doi: 10.1016/j.biopsych.2005.08.022

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Camody, T. J., Arnow, B., Klein, D. N., et al. (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. *Biol. Psychiatry* 54, 573?-583. doi: 10.1016/S0006-3223(02)01866-8

Sarkar, P. (2000). A brief history of cellular automata. *ACM Comput. Surv.* 32, 80–107. doi: 10.1145/349194.349202

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., et al. (2014). Early-warning signals for critical transitions. *Nature* 46, 53–59. doi: 10.1038/nature08227

van de Leemput, I. A., Wichers, M., Cramer, A. O., Borsboom, D., Tuerlinckx, F., Kuppens, P., et al. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proc. Natl. Acad. Sci. U.S.A.* 111, 87–92. doi: 10.1073/pnas.1312114110

von Zerssen, D. (1986). "Clinical self-rating scales (CSRS) of the Munich psychiatric information system (PSYCHIS München)," in *Assessment of Depression* eds N. Sartorius and T. A. Ban (Berlin: Springer), 270–303.

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). *Psychol. Methods* 17, 228–243. doi: 10.1037/a0027127

Waldorp, L. J., and Kossakowski, J. J. (2019). *Mean Field Dynamics of Stochastic Cellular Automata for Random and Small-World Graphs.* Available online at: https://arxiv.org/pdf/1610.05105.pdf

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of small-world' networks. *Nature* 393, 440–442. doi: 10.1038/30918

Wichers, M., Groot, P. C., Psychosystems, ESM Group, and ESW Group (2016). Critical slowing down as a personalized early warning signal for depression. *Psychother. Psychosomat.* 85, 114–116. doi: 10.1159/000441458

Wit, E., van den Heuvel, E., and Romeijn, J.-W. (2012). All models are wrong…': an introduction to model uncertainty. *Stat. Neerlandica* 66, 217–236. doi: 10.1111/j.1467-9574.2012.00530.x

Wolfram, S. (1984). Computation theory of cellular automata. *Commun. Math. Phys.* 96, 15–57. doi: 10.1007/BF01217347

World Health Organization (2012). *Depression, A Hidden Burden.* Available online at: http://www.who.int