Journal of Mathematical Psychology 97 (2020) 102380

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

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

Lourens Waldorp *, Jolanda Kossakowski

University of Amsterdam, Nieuwe Achtergracht 129-B, 1001 NP, The Netherlands

ARTICLE INFO

Article history:
Received 13 April 2019
Received in revised form 21 February 2020
Accepted 27 April 2020
Available online xxxx

Keywords:
Cellular automata
Discrete dynamical system
Nonlinear dynamics
Bifurcation
Psychopathology

ABSTRACT

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

1. Introduction

People appear to shift moods, sometimes rather suddenly and in a discrete manner (Heath, 2015; Heath, Heiby, & Pagano, 2007; Hosenfeld et al., 2015). Such 'discontinuities' can be the result of relatively small changes in the environment or person. In some cases such transitions in mood are associated to mental disorders like depression. Recently, mental disorders have been described as a network (graph) of interacting symptoms (Borsboom, Cramer, Schmittmann, Epskamp, & Waldorp, 2011; Bringmann et al., 2013; van Borkulo, Epskamp, Blanken, Boschloo, Schoevers, & Waldorp, 2014). For instance, lack of sleep during the night could lead to poor concentration during the day, which in turn could lead to lack of sleep again by worrying that your job may be on the line. Here we use these ideas and model the dynamics of psychopathology networks as stochastic cellular automata. This connects networks and dynamical systems theory to provide a framework for an explanation for transitions in mood. To analyse the dynamics we use a mean field approximation to stochastic cellular automata where each node is similar in behaviour to all others. We extend known results for the mean field approach in this context to other types of graphs (random and small-world graphs), where the mean field can be interpreted as a weighted average of all nodes in the graph. We provide quantitative results on the error of approximation for the mean field for the different types of graphs. Simulations show that the mean field is accurate in different settings.

Generally, mood changes over time, meaning it is dynamic (Dejonckheere et al., 2019). In healthy individuals changes in the environment or in a person are adapted to in a relatively short period of time, while in individuals with a mood disorder depressed mood could persevere (Heath et al., 2007; Kuppens, Allen, & Sheeber, 2010). These observations seem to imply that healthy individuals obtain symptoms similar to depressed patients, but recover relatively quickly, and so, have a larger variation (Kuppens et al., 2010), and over a short period of time, having a moderate (average) number of symptoms. The state of a moderate or average number of symptoms can therefore be associated with a healthy state. A depressed state, on the other hand, is likely to be associated with a large number of symptoms, which could change suddenly because of the maladaptiveness (Heath et al., 2007).

Describing the dynamics of mood has been the focus of many recent studies. A popular choice has been linear stochastic differential equations (e.g., Ebner-Priemer et al., 2015; Lodewyckx, Tuerlinckx, Kuppens, Allen, & Sheeber, 2011; Oravecz, Tuerlinckx, & Vandekerckhove, 2011, continuous and discretised). In these models a linear, latent state space model explains the changes in observations over time and allows to determine convergence to some stable state. These models use a hierarchical structure

* Corresponding author.
E-mail address: waldorp@uva.nl (L. Waldorp).

https://doi.org/10.1016/j.jmp.2020.102380
0022-2496/

that accommodates individual differences (random effects). Other linear models, that do not assume a latent, underlying structure, have also been suggested (Bringmann et al., 2013; Haslbeck & Waldorp, 2020). These models are linear and as such do not include the possibility of sudden mood changes, which we believe are an essential feature of mood disorders (Heath et al., 2007). Nonlinear modelling of the dynamics of mood changes show that sudden changes are required to be accommodated (Heath et al., 2007). However, most of these nonlinear methods try to establish that some change is at hand, but not to what state (Heath, 2015). Here we try to accommodate both the stable and bistable states a person can be in, associated with a healthy and an unhealthy state, respectively. Additionally, the trajectory obtained from a model could reveal how particular changes could arise. We believe that cellular automata provide behaviour that is rich enough to provide a good description of such systems.

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

Many versions of SCA exist and of particular interest are those that behave similar to the Ising network (a graph with binary nodes where the probability is determined by the nodes and their interactions with other nodes, see, e.g., van Borkulo et al., 2014; Wainwright & Jordan, 2008). The reason is that the Ising network is often used to model realistic phenomena, like magnetisation (Kindermann, Snell, et al., 1980; Sethna, Dahmen, & Perkovic, 2004) or psychopathologies (van Borkulo et al., 2014). We have in mind the application to psychopathology here. In such systems the symptoms of disorders are the nodes in the graph and edges between the symptoms are estimated from data using the Ising model or connectivity is obtained from verbal accounts. Mood shifts can then be associated with (sudden) increases in the number of symptoms (Heath, 2015; Hosenfeld et al., 2015). One basic first question is then whether in the SCA the number of active symptoms will eventually be average (healthy) or large and could possibly change suddenly (unhealthy). And so, here we start with the proportion of active symptoms, which seems a reasonable starting point (Dejonckheere et al., 2019). We will determine for a person whether sudden changes in the proportion of active symptoms is possible (bistability) or not.

Part of the reason we chose the framework of cellular automata is because they are rich in terms of macroscopic behaviour (e.g., proportion of active cells can be bistable or stable) and they allow for analytical treatment to some extent. Watts (1999) showed that a one-dimensional, large-scale cellular automaton (deterministic) where the connectivity between nodes was arranged as a small-world, could perform the density task (end up with a majority of active nodes) and synchronisation task (eventually keep on alternating between all inactive and all active nodes regardless of the initial configuration). Newman and Watts (1999) gave approximations for path length and clustering on a small-world graph, to obtain an analytic solution to the threshold above which a large number of nodes are active. Callaway, Newman, Strogatz, and Watts (2000) studied percolation in different graph topologies in deterministic automata, focussing on the consequences of (randomly) deleting nodes. Here again the objective was to concentrate on stable solutions of the graphs. In a stochastic version, Tomassini, Giacobini, and Darabos (2005) investigated a one-dimensional SCA on a regular and small-world graph in terms of its performance on the density and synchronisation tasks. They determined by using evolutionary algorithms that a small-world topology is most efficient (compared to random graphs and lattices) to solving both tasks, corresponding to the results of Watts (1999) in a deterministic version. Their objective was different from ours in that here we are interested in all types of dynamic behaviour (stable or not), and specifically representing this behaviour for the SCA by the mean field.

Our starting point is the work by Balister, Bollobás, and Kozma (2006) and Kozma et al. (2005) where a two-dimensional grid on an SCA is defined. A one-dimensional grid is a set of nodes on a line where each node is connected to the nodes net to it (see Fig. 1). Similarly, in a two-dimensional grid the nodes are located on the intersecting points of integers in the plane, where at the edges of the grid nodes are connected to the nodes at the opposite end, so that it is in fact a torus (see Fig. 2). The mean field is then used to determine the unconditional probability distribution of the density (relative number of active nodes). Balister et al. (2006) show that the mean field model predicts a bifurcation for small values of the probability of a node switching to another state and they determine its critical point for a neighbourhood of size five (see also Kozma, Puljic, Balister, Bollobas, & Freeman, 2004; Kozma et al., 2005). This is of particular interest in our case as it may explain mood disorders (e.g., depression or manic-depression) from symptoms and their connectivity. To apply these results to random and small-world graphs we determine the marginal distribution across the possible node degree (number of connections of a node) probabilities given the topology of the random or small-world graph. Extending results of homogeneous graphs has been applied to social networks by Barrat, Barthelemy, and Vespignani (2008) and to cellular automata (Janson, Kozma, Ruszinkó, & Sokolov, 2015).

The current study of the accuracy of the mean field approach convinced us that we could apply this approach to empirical data. We have applied the mean field both to a sample of the general population and a group of diagnosed depressed patients (Kossakowski, Gordijn, Harriette, & Waldorp, 2019). First results indicate that the model is reasonable. In the application we tried to determine whether a person is in a position in the state space such that sudden changes are possible (unhealthy, being stuck in a low proportion of symptoms may lead to a sudden transition to a large proportion of active symptoms) or not (healthy, a relatively fast but not too large variation of proportion of symptoms). The mean field model indicates that there appears to be a large

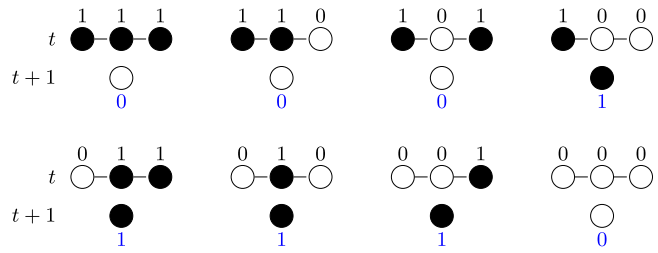

Fig. 1. Illustration of a one-dimensional automaton. The 8 configurations of states $x_{\Gamma(2)}$ (above the three nodes) for the local rule ϕ applied at time t (three nodes at the top) that result in the active or inactive node at time $t+1$ (single node at the bottom with its subsequent state below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

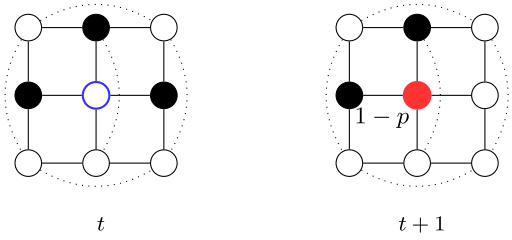

Fig. 2. Illustration of a two-dimensional grid (lattice) where the centre node has four neighbours and itself at the previous time point (not shown). Each node has the same number of neighbours, and the nodes at the periphery are connected to the opposite nodes (dotted lines), making it a torus. At time point t there are three neighbours active, and at time point $t+1$ the centre node has probability $1-p$ of becoming active.

majority in the patient group that indicate sudden changes in the proportion of symptoms are possible, compared to the general population.

We first introduce stochastic cellular automata in Section 2. Then in Section 3 we show how the traditional version of an SCA on a grid can be reduced to a single discrete time dynamical system, called the mean field. In Section 3.2 we show that for the random graph we can use a variation on the formulation for the grid of the dynamical system to describe dynamics. We use these results on the random graph to show in Section 3.3 that we can obtain a similar approximation for the small-world graph, again using the formulation for the grid. Having shown that these approximations are appropriate, we see in Section 4 what the dynamics of the process are for the different graph topologies. We follow these theoretical results by simulations to verify the accuracy of the mean field in Section 5. Proofs can be found in the Appendix.

2. Stochastic cellular automata

A cellular automaton is a dynamical system consisting of nodes in a fixed, finite grid where directly connected nodes (neighbours) determine the state of a node at each subsequent time step (Wolfram, 1984b). Each node i in a node set $V = \{1, 2, \ldots, n\}$ is at time t in one of the states of a finite alphabet A, referred to as $x_i = a \in A$, where, for instance, $A = \{0, 1\}$ (in this paper we only consider binary alphabets). The neighbours in the graph $G = (V, E)$ are given by the edges (i, j) in the edge set E. Often the graph G is a finite subset of the square lattice $\mathbb{Z}^2$ (with nodes at each integer coordinate in the plane, a graph in two-dimensional space), where for each node i the neighbourhood $\Gamma(i)$ are the nodes that have an edge to i, i.e., $\Gamma(i) = \{j : (i, j) \in E\}$. In the original work the neighbourhood included the node itself (at a previous time point), which we differentiate by writing $\Gamma_+(i) = \Gamma(i) \cup \{i\}$. The nodes at the periphery of the grid are connected to nodes at the opposite end, so that the structure becomes a torus where each node has exactly the same number of neighbours (see Fig. 2). In the mean field model described below, we often make use of a generic neighbourhood, where the specific node is irrelevant because each node behaves the same; we call that generic neighbourhood Γ.

A local rule $\phi : A^{\Gamma_+} \to A$ assigns to each configuration of states of the nodes in the neighbourhood $x_{\Gamma_+(i)}$ a value such that $\phi : x_{\Gamma_+(i)} \mapsto x_i$, where $x_i = a \in A$. For instance, consider the one-dimensional graph (a graph on a line with neighbours on each side) in Fig. 1, where in each figure the top row represents time point t and the bottom represents time point $t+1$. The centre node has two neighbours, and application of ϕ is represented by the row below it indicated by $t+1$. Each node has a value from the binary alphabet $A = \{0, 1\}$. There are three nodes and the local rule ϕ determines from $x_{\Gamma_+(2)} = (x_1, x_2, x_3)^\mathsf{T}$ at time point t the value $\phi(x_{\Gamma_+(2)})$ at time point $t+1$. Many different rules for the update function ϕ exist (see e.g., Wolfram, 1984a). In Fig. 1 we show the so-called Rule 30, so named because when the output values are considered as coefficients of the binary expansion, they represent the number 30 in the decimal expansion. This Rule 30 determines, for instance, that if all nodes in $\Gamma_+(2)$ are 1, then at the next time point, the value of the centre node is 0. All 8 possible configurations for three nodes in the neighbourhood $\Gamma_+(2)$ are considered in Rule 30 and are shown in Fig. 1. Cellular automata are completely determined by a particular mapping chosen for ϕ.

Although (deterministic) cellular automata show extremely interesting and complex behaviour such as periodicity and chaos, we require a stochastic cellular automaton (SCA, see, e.g., Paz, 1971) because in psychology we do not have full knowledge of how a symptom changes from inactive to active. Therefore we make the state of a node $X_{i,t}$ a Bernoulli random variable that is 0 or 1 at time point t. We also define $Z_{i,t} = \sum_{j \in \Gamma_+(i)} X_{j,t}$ for the sum of active neighbours of node i. In an SCA one of the most common rules to determine the probability of the state of a node at time point $t+1$ is the majority rule; the majority rule is popular not only because of intuitive appeal, but also because it is stable in the sense that a small number of random flips in the input will not easily lead to different results (O'Donnell, 2014). This seems a reasonable assumption for psychological processes since an unstable rule would lead to dramatic sudden changes all the time, which is not something we observe. We have some empirical evidence that the majority rule is reasonable given the fit of the mean field model to several data sets (see Kossakowski et al., 2019).

The majority rule for any node is defined as the probability p of obtaining state 1 if there are fewer than or equal to half of the neighbours in state 1 at time point t, and probability $1-p$ of obtaining 1 if there are more than half of the neighbours 1 at time point t. That is,

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

where $|\Gamma_+(i)|$ denotes the cardinality of $\Gamma_+(i) = \{(i, j) \in E\} \cup \{i\}$; we refer to this rule as maj_p. The nodes in graph G are updated simultaneously.

As an example, consider a small lattice or grid with boundary conditions (shown in Fig. 2), such that each node has exactly five neighbours including itself at the previous time point t. The centre node has three neighbouring nodes in state 1 at time point t and so we obtain the probability $1-p$ of the centre node being in state 1 at time point $t+1$.

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

3. Mean field approximation on graphs

To learn about long term behaviour of the SCA we may turn to the mean field, which uses the assumption of homogeneous nodes, all having the same update rule and same number of neighbours. In doing so we are able to reduce the complicated SCA to a one-dimensional discrete-time Markov chain. This will allow us to analyse its dynamics more easily (see Section 4). The key ingredient of the mean field approximation, shown by Balister et al. (2006), is that the properties of interest are uniform over the graph. For the (toroidal) grid topology in two dimensions this is intuitive to see: Any node has the same number of neighbours and each node becomes 0 or 1 by the same local rule (majority rule). We extend results of the grid to the random graph and the small-world graph, where, clearly, the assumption that each node has the same number of neighbours is the same is false. We first consider the case for a grid and then move on to the random and small-world graph. For convenience, we have included a table with the most important symbols and their meaning and a reference to the equation number if there is one; see Table 2 in the Appendix.

3.1. Mean field on a grid

Let the graph $G_{\text{grid}}(n, \Gamma_+)$ be a grid with n nodes and boundary conditions such that each node has exactly the same number of neighbours. The mean field approximation for the grid assumes that any set of nodes (of the size of the neighbourhood) could serve as part of the neighbourhood, that is, we have a generic set of nodes Γ_+ from the graph G_{grid} instead of the exact neighbourhood $\Gamma(i)_+$ for node i because each node has the same number of nodes and the same update rule. For instance, in a finite subset of the two-dimensional lattice $\mathbb{Z}^2$ the number of neighbours is 5 (including the node itself). This implies that the number of active nodes in Γ_+, referred to as $Z_t = \sum_{j\in\Gamma_+} X_{j,t}$ (without the subscript i), depends on the number of nodes in state 1. We call the number of active nodes in the entire graph Y_t defined by $\sum_{j\in V} X_{j,t}$, the sum of all $|V| = n$ nodes in the graph. In order to know the evolution of Y_t over time we need to know the transition probability $\mathbb{P}(y_{t+1} \mid y_t)$ in the setting of the SCA with majority rule. Since the $X_{i,t}$ are Bernoulli random variables, Y_t is binomial of size n and with the success probability the probability of $X_{i,t}$, which depends on the number of active nodes in the neighbourhood Γ_+. The probability of obtaining an active node in the generic neighbourhood Γ_+ is determined by Y_t/n, which is often referred to as the density and denoted by ρ_t. So, in the mean field we have the probability $\mathbb{P}(X_{i,t} = 1 \mid z_t, \rho_t)$, and we require the probability $\mathbb{P}(X_{i,t} = 1 \mid \rho_t)$, which demands that we know $\mathbb{P}(Z_t = r \mid \rho_t)$ so that we can marginalise over the possible values $r = 0, 1, \ldots, |\Gamma_+|$ for the number of nodes in state 1. The probability of r nodes in state 1 in the neighbourhood Γ_+ is

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

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

$$\sum_{r=0}^{|\Gamma_+|} \mathbb{P}(x_{i,t+1} \mid r, \rho_t)\mathbb{P}(r \mid \rho_t) = \sum_{r=0}^{|\Gamma_+|} \text{maj}_p(r, |\Gamma_+|)\binom{|\Gamma_+|}{r}\rho_t^r(1-\rho_t)^{|\Gamma_+|-r}. \tag{2}$$

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

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

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

$$p_{\text{grid}}(\rho_t) = \mathbb{P}(x_{i,t+1} \mid \rho_t) \tag{3}$$

where the marginal was over the different number of active neighbours $z_t \in \{0, 1, \ldots, |\Gamma_+|\}$. As Balister et al. (2006) showed, it follows that the number of nodes in state 1 in graph G_{grid} is then a Binomial random variable. Let $B(n, p)$ denote a binomial random variable with n Bernoulli trials each with success probability p.

Lemma 1 (*Balister et al. (2006, Theorem 2.1)*)**.** *Let $G_{\text{grid}}(n, \Gamma_+)$ be a grid with an SCA as defined above with majority rule maj_p in (1) and $\rho_t = Y_t/n$ be the proportion of nodes in state 1 at time t. Let the marginal probability $p_{\text{grid}}(\rho_t) = \mathbb{P}(x_{i,t+1} \mid \rho_t)$ in (2) be denoted by p_{grid}. Then the evolution of the number of active nodes Y_t on G_{grid} is*

$$Y_{t+1} = B(n, p_{\text{grid}}(\rho_t)). \tag{4}$$

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

In this binomial process at each time step the proportion $\rho_t = Y_t/n$ is determined and then the function p_{grid} in (2) is applied. This provides a new probability for the binomial process at each time point. Because we are considering an averaging process, the mean field, we are interested in the repeated application $p_{\text{grid}} \circ \cdots \circ p_{\text{grid}}(\rho_t)$, where the composition $p_{\text{grid}} \circ p_{\text{grid}}(\rho_t)$ is defined by $p_{\text{grid}}(p_{\text{grid}}(\rho_t))$. If the mean p_{grid} is a good description of the process (deviations from p_{grid} are not too large), then we can use this mean field as an accurate description of the binomial process. We are then interested in the evolution of p_{grid}. First we consider the accuracy of the approximation.

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

Lemma 2. *Accuracy Bound of Density Let $Y_t = \sum_{x\in V} X_{i,t}$ be the sum of n Bernoulli trials given by (4), with mean of the density $p_{\text{grid}}(\rho_t)$.*

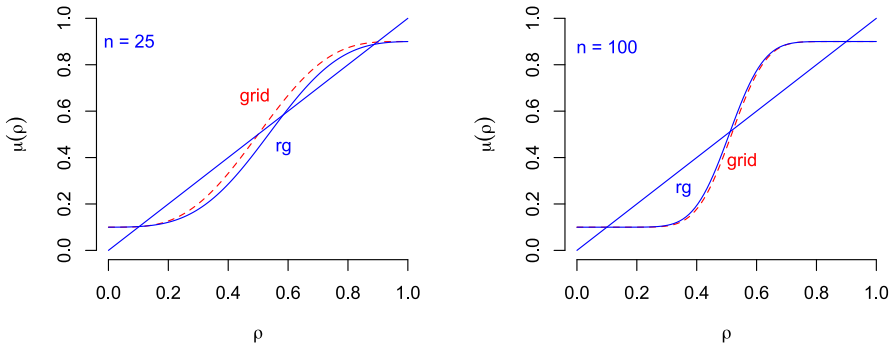

Fig. 3. The expectation $\mu_{\text{rg}} = p_{\text{rg}}$ (blue, solid curve) of Eq. (6) and μ^{ν}_{grid} (red, dotted curve) of Eq. (7) with $p = 0.1$ and $p_e = 0.3$. Left panel shows the curves for a graph of size $n = 25$, showing a clear difference between the curves, and the right panel for graph size $n = 100$. Note that the difference between the curves at the crossings with the 45° line is small.

For every $0 < \varepsilon < \min\{p_{\text{grid}}, 1 - p_{\text{grid}}\}$, *let* $\delta = 2\exp(-\varepsilon^2/2\sigma^2_{\text{grid}})$. *We then have with probability at least* $1 - \delta$

$$|\rho - p_{\text{grid}}| \leq \sqrt{\frac{p_{\text{grid}}(1 - p_{\text{grid}})}{n} 2\log(2/\delta)} \tag{5}$$

For instance, we obtain the interval with probability at least 0.95 of $[\mu_{\text{grid}} - 2.72\sigma_{\text{grid}}, \mu_{\text{grid}} + 2.72\sigma_{\text{grid}}]$. Another interval can be obtained from the DeMoivre–Laplace central limit theorem. This theorem tells us that for large enough n, $W_{\text{grid}} = (\rho - p_{\text{grid}})/\sigma_{\text{grid}}$ is distributed as $N(0, 1)$. In fact, if the third order moment of W_{grid} is $c < \infty$, the Berry–Esseen theorem says that the order of approximation of the distribution of ρ to the normal distribution is $O(3c/\sqrt{n})$ (Venkatesh, 2013). This provides an interval for ρ_{t+1} as a measure of accuracy with $[\mu_{\text{grid}} - 1.96\sigma_{\text{grid}}, \mu_{\text{grid}} + 1.96\sigma_{\text{grid}}]$ with probability 0.95. Clearly, in both limit laws the size of the network n determines the accuracy of the approximation.

3.2. Mean field on a random graph

In the original setting of a grid (with boundary conditions) the number of neighbours is fixed and it was seen that the mean field approximation p_{grid} was accurate for the density because each node is identical with respect to a change depending on its neighbours. Here we introduce the neighbourhood size $|\Gamma|$ as a random variable such that we obtain the probability $\mathbb{P}(X_{i,t+1} \mid |\Gamma| = k, \rho_t)$. Note that we use Γ, the neighbourhood without the node itself. Then we determine the probability of $X_{i,t+1} \mid \rho_t$ by averaging over all possible sizes of neighbourhoods $k = 0, \ldots, n-1$ weighted by its probability for neighbourhood size (marginalising). This is done in a random graph where each node has a binomial number of neighbours. Let $G_{\text{rg}}(n, p_e)$ be a random graph with n nodes and (constant) probability p_e of an edge being present, independently (Bollobás, 2001; Durrett, 2007). Let the size of the neighbourhood $|\Gamma|$ be a binomial random variable with maximal value $n - 1$ neighbours and probability p_e, that is, $B(n - 1, p_e)$. Then the probability of obtaining an active node can be defined conditionally on the event $\{|\Gamma| = k\}$, the neighbourhood having size k. That is, we obtain the probability for each neighbourhood size

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

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

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

Then using (2) obtaining

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

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

$$p_{\text{rg}}(\rho_t) = \sum_{k=0}^{n-1}\sum_{r=0}^{k} \text{maj}_p(r, k)\binom{k}{r}\rho_t^r(1 - \rho_t)^{k-r}\binom{n-1}{k} \times p_e^k(1 - p_e)^{n-k-1}. \tag{6}$$

This is a rather lengthy expression and is not easy to work with. So we aim to approximate this expression and determine the error of approximation. Intuitively we would expect that we could replace the sum over all neighbourhood sizes k in (6) and take the expected neighbourhood size $\lfloor p_e(n-1) \rfloor$ only, for each node, where $\lfloor a \rfloor$ is the integer part of a. Then we obtain a result for the average neighbourhood size, which should be reasonably close. This would result in a simpler formulation of the probability of a node to be active in a random graph and would make determining fixed points easier and simplify computation for large graphs considerably. We next show that such an approximation is a reasonable approach (the proof is in the Appendix).

Proposition 3 (*Probability Random Graph*)**.** *Let the neighbourhood in* $G_{\text{rg}}(n, p_e)$ *of each node in an SCA as defined above be fixed with the expected number of nodes under the random graph* $\nu = \lfloor p_e(n-1) \rfloor$ *such that the neighbourhood size is* ν *in the majority rule* maj_p *in* p_{grid}. *Then the probability* p_{rg} *can be approximated by*

$$p^{\nu}_{\text{grid}}(\rho_t) = \sum_{r=0}^{\nu} \text{maj}_p(r, \nu)\binom{\nu}{r}\rho_t^r(1 - \rho_t)^{\nu-r}. \tag{7}$$

The approximation error is

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

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

This shows that the number of active nodes Y_t in the random graph with probability p_{rg} in (6) and the number of active nodes with probability p^{ν}_{grid} in (7) converge in probability with exponential rate with graph size n. Fig. 3 illustrates the difference between p_{rg} and p^{ν}_{grid} for graph sizes $n = 25$ and $n = 100$.

Remark 4. To retain a probability of an edge in p^{ν}_{grid} leads to a larger approximation error, i.e., using

$$p^{n-1}_{\text{grid}}(\rho_t) = \sum_{r=0}^{n-1} \text{maj}_p(r, n-1)\binom{n-1}{r}(\rho_t p_e)^r(1 - \rho_t p_e)^{n-r-1} \tag{8}$$

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

We now have expression (7) similar to (2) for a random graph with the probability of an active node at time t determined by both the density ρ_t and an edge being present p_e in the size of the neighbourhood. From (6) and Lemma 1 and Proposition 3 the evolution equation for the random graph can also be described by a binomial process with probability (7)

$$Y_{t+1} = B(n, p^{\nu}_{\text{grid}}(\rho_t)). \tag{9}$$

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

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

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

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

3.3. Small-world graph

More interesting in many real-world applications is the small-world graph, and so we perform a similar analysis for the small-world graph as for the random graph. A small-world graph is one which has high average clustering and low average path length, relative to a random graph with the same number of nodes and edges. These graphs have been shown to model realistic networks like those of working relations between actors and the nerve cells in the worm C. elegans (Watts & Strogatz, 1998), and subsequently the small-world has been shown to apply to many different networks, like the (parcellated) brain (Sporns & Honey, 2006). And most recently, the network of symptoms as defined by the diagnostic statistical manual (a compendium to diagnose patients) has been found to be a small-world. This finding is a possible explanation for the correlations between pairs of symptoms found in different subpopulations (Borsboom et al., 2011). Here we use a slightly modified version of the original small-world graph of Watts and Strogatz (1998), called the Newman–Watts (NW) small-world of Newman and Watts (1999). In this modified version instead of rewiring existing lattice edges, the lattice remains and new edges are added (shortcuts) with some (small) probability (see Fig. 8, bottom row). We use the NW small-world graph rather than the original one because this graph never gets disconnected. In the original small-world graph it may occur that some set of nodes becomes disconnected from the rest of the graph. Here, as before, we require that the graph be connected. Also, the NW small-world is mathematically easier to work with than the original small-world. The main properties of the small-world graph are retained in the NW version (Newman & Watts, 1999).

In the NW small-world graph, where for a given grid structure each node has neighbourhood Γ (without the node itself), a set of $(n-1)p_w$ edges is on average independently added, where p_w is the probability of two nodes being wired. Such a graph is denoted by $G_{\text{sw}}(n, \Gamma, p_w)$. The same idea as with the random graph, where the probability for an active node was corrected by the probability of the degree of a node, averaged over all possible neighbourhood sizes, can be used for the random part in the NW small-world. In the NW small-world we start with a grid with neighbourhood size $|\Gamma|$, which is fixed, and augment the graph randomly with edges according to a binomial variable with probability p_w. We then obtain

$$p_{\text{sw}}(\rho_t) = \sum_{k=|\Gamma|}^{n-1} \sum_{r=0}^{k} \text{maj}_p(r,k) \binom{k}{r} \rho_t^r (1-\rho_t)^{k-r} \binom{n-1}{k} \times p_w^k (1-p_w)^{n-k-1}. \tag{10}$$

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

$$p^{\text{sw}}_{\text{grid}} = p_{\text{grid}}(1-p_w)^{n-|\Gamma|} \tag{11}$$

for the first part of the fixed grid with neighbourhood Γ. Then, in accordance with the random part of the NW small-world, a probability is added to emulate the possible additional neighbours in the random part of the graph, ignoring the first $|\Gamma|$ neighbours from the grid. We obtain this part directly from Proposition 3 for the remaining $n - |\Gamma|$ nodes. Define the probability

$$p_{\text{rg}\setminus\Gamma}(\rho_t) = \sum_{k=|\Gamma|+1}^{n-1} \sum_{r=0}^{k} \text{maj}_p(r,k) \times \binom{k}{r} \rho_t^r (1-\rho_t)^{k-r} \binom{n-1}{k} p_w^k (1-p_w)^{n-k-1}, \tag{12}$$

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

$$p^{\nu}_{\text{grid}\setminus\Gamma}(\rho_t) = \sum_{r=|\Gamma|+1}^{\nu} \text{maj}_p(r,\nu) \binom{\nu}{r} (\rho_t)^r (1-\rho_t)^{\nu-r}, \tag{13}$$

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

Corollary 5 (*Probability on a NW Small-World*). *Let $G_{\text{sw}}(n, \Gamma, p_w)$ be the NW small-world graph of size n with $|\Gamma|$ nodes in the fixed neighbourhood for each node. Furthermore, let $0 < p_w < 1$ be the wiring probability and $\nu = \lfloor p_w(n - |\Gamma|) \rfloor$. Then the approximation error for the probability using the grid structure p^{ν}_{grid} in (13) in the random part is*

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

for any $\varepsilon > 0$.

Eqs. (10) to (13) and Corollary 5 prove that the equation for the evolution on an NW small-world is

$$Y_{t+1} = B(|\Gamma|, p^{\text{sw}}_{\text{grid}}(\rho_t)) + B(n - |\Gamma|, p^{\nu}_{\text{grid}\setminus\Gamma}(\rho_t)) \tag{14}$$

We write

$$p^{\nu}_{\text{sw}} = \frac{|\Gamma|}{n} p^{\text{sw}}_{\text{grid}} + \frac{n-|\Gamma|}{n} p^{\nu}_{\text{grid}\setminus\Gamma} \tag{15}$$

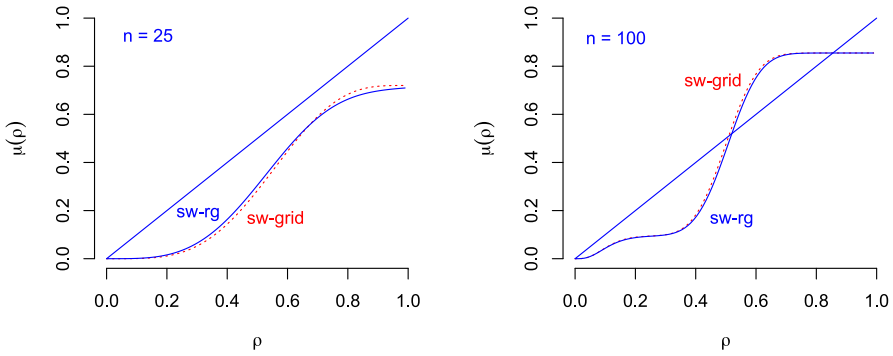

Fig. 4. The probability p_{sw} (blue, solid curve) in Eq. (10) and p^{v}_{sw} (red, dotted curve) in Eq. (15) with $p = 0.1$ and $p_w = 0.3$. The left panel shows the curves for a graph of size $n = 25$, and the right panel for graph size $n = 100$.

for the NW small-world probability based on the approximation with the grid. Fig. 4 shows two examples of the approximation p^{v}_{sw} for the NW small-world. It is clear from Corollary 5 that convergence is a bit slow for small graphs since the difference of nodes in the fixed neighbourhood Γ and in the expected $p_w(n - |\Gamma|)$ neighbours in the random part, determines the rate. But again, we can use p^{v}_{sw} to determine the dynamics of the mean field for large graphs.

4. Dynamical properties

We are interested in the long run behaviour of p_G, where G is a graph (grid, random graph or small-world graph), because this mean field describes what the number of active nodes Y_t, in (4) for the grid, which is a binomial process $B(n, p_G(\rho_t))$, will be. For the process Y_t we know that the transition probabilities are

$$\mathbb{P}(Y_{t+1} = r \mid Y_t = k) = \binom{n}{r} p_G(\rho_t)^r (1 - p_G(\rho_t))^{n-r} \tag{16}$$

where $\rho_t = k/n$. It is easily seen that this is a discrete time Markov process on a finite state space of size $n + 1$, since the probability of Y_{t+1} depends only on Y_t. It is also clear that for each time point t the transition probability is different, i.e., we are dealing with an inhomogeneous Markov process. For graph G of size n we have $n + 1$ states, and because we define the transition probability in (16) using the time dependent value $\rho_t = Y_t/n = \sum_{j=1}^{n} X_{j,t}/n$ in the function p_G, we obtain a different transition probability at each t. Following Saloff-Coste and Zúñiga (2010), we denote the transition probability in (16) by $K_t(k, r) = \mathbb{P}(Y_{t+1} = r \mid Y_t = k)$ (also called a Markov kernel), and denote the $(n + 1) \times (n + 1)$ matrix of transition probabilities by K_t.

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

4.1. Stability

For a time-independent (homogeneous) transition Markov kernel $K(r, k) = \mathbb{P}(Y_{t+1} = r \mid Y_t = k)$, for any t, the Markov chain at time t is $\pi_t = \pi_0 K^t$, for initial distribution π_0 over the $n + 1$ states. Recall that if K is irreducible, then π_t is essentially independent of the initial distribution, and there exists a unique invariant distribution π such that $\pi K = \pi$. If in addition K is aperiodic, then for large t the distribution π is well approximated by π_t (Levin, Peres, & Wilmer, 2017; Norris, 1997). Similar (but weaker) results can be obtained with inhomogeneous Markov chains (Saloff-Coste, Zuniga, et al., 2009). For an inhomogeneous Markov chain of finite state, we are concerned with a dissociation from the initial distribution (merging) and whether at some time point the Markov chain becomes stable (the distribution remains within a certain range for each state). Note that probabilities $\pi_t(r)$ need not be stationary, and we do not always expect that such stability will be obtained (see, e.g., Saloff-Coste & Zúñiga, 2010, for examples).

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

$$\pi_t = \pi_0 K_{0,t} = \pi_0 K_0 K_1 \cdots K_t \tag{17}$$

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

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

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

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

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

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

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

where $\pi_0(k)$ is the probability of state k at time point 0 and $V_0 = \{0, 1, \ldots . n\}$. Because $K_{0,1}(k, r) = \mathbb{P}(X_{i,1} = r \mid X_{i,0} = k)$, we have that $\sum_{r \in V_0} K_{0,1}(k, r) = 1$, which gives the result when recursively applying the above inequality. It is clear that if the kernels $K_{t,t+1}(k, r)$ were too small, the inequality is not true, and we would not have a contracting Markov chain. Hence, this result from Saloff-Coste and Zúñiga (2010, Thm. 4.3) about merging, assumes that each of the K_t in the product $K_0 K_1 \cdots K_t$ is irreducible. In fact, we require that for any time point t the elements of the transition kernel $K_{t,t+1}(k, r) = \mathbb{P}(Y_{t+1} = r \mid Y_t = k)$ in (16) are $> \eta_t$, for some $\eta_t > 0$ and states k and r (uniform

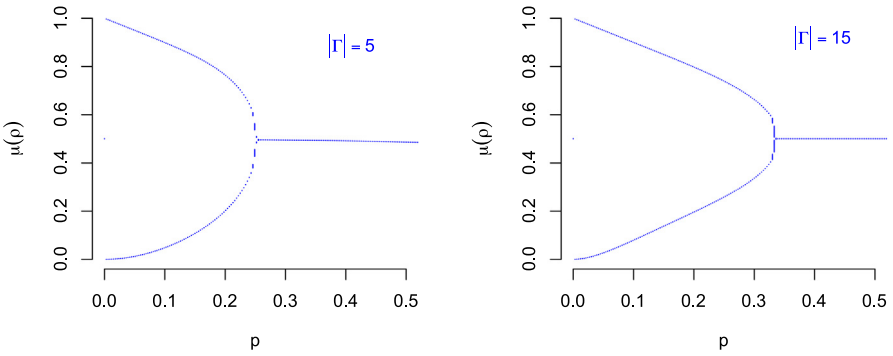

Fig. 5. Bifurcation plots of μ_{grid} in a graph of size $n = 100$ for $|\Gamma| = 5$ (left) and $|\Gamma| = 15$ (right) neighbours.

irreducibility). The transition kernel $K_{t,t+1}(k, r)$ is 0 only if p_G is 0. Hence, we obtain the fact that the Markov kernel $K_{t,t+1}$ on the finite state space $\{0, 1, \ldots, n\}$ is irreducible if the orbit $p_G^t(\rho_0)$ will not become 0 or 1 at some point. This requires that the parameter p in the majority rule maj_p in (1) is such that p_G cannot be in the stable set $S(0) = \{\rho : p_G^t(\rho_0) = 0 \text{ for any } t\}$ or similarly in $S(1) = \{\rho : p_G^t(\rho_0) = 1 \text{ for any } t\}$.

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

Lemma 6. *Let Y_{t+1} be the binomial process in (16) with mean field probability p_G for graph G a grid, random graph or a small-world graph. Suppose that the parameter p in the function p_G is not in the stable set $S(0) \cup S(1)$, and hence that for each t the matrices K_t are uniformly irreducible, i.e., for all states k and r, $K_t(k, r) > \eta_t$, for some $\eta_t > 0$. Then we have the following.*

(1) The chain K_t for large t is merging, i.e., for any initial distribution the chain will approximately be π_t.
(2) The probabilities $\pi_t(r)$ are stable, i.e., within the bounds

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

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

4.2. Dynamics of the mean field

The dynamics of the mean field in the grid G_{grid} from (2) have been described in Balister et al. (2006) and Kozma et al. (2005) for a neighbourhood size of $|\Gamma_+| = 5$. The probability function $p_{\text{grid}} : [0, 1] \to [0, 1]$ defined in (2) and (3) is continuous. And so, since $[0, 1]$ is closed and bounded, we find that p_{grid} has at least one fixed point in $[0, 1]$ (Hirsch, Smale, & Devaney, 2004; Holmgren, 1996). A fixed point is one where we find $p_{\text{grid}}(\rho_t) = \rho_t$. Finding the fixed points for p_{grid} in general is not trivial. Balister et al. (2006) showed that if $|\Gamma_+| = 5$ in the finite grid, then $p = 7/30 \approx 0.233$ is a critical (bifurcation) point, such that if p is in $[7/30, 1/2]$ then there is a stable fixed point at $\rho_t = 0.5$, but when $p < 7/30$ then $\rho_t = 0.5$ is unstable and there are two other stable fixed points. This can be seen in Fig. 5, which shows two bifurcation plots, where for each value of $0 < p \le 0.5$ the function $\mu_{\text{grid}} = p_{\text{grid}}$ is iteratively applied for 1000 steps, and only the last 50 are plotted. Fig. 5 shows that for $|\Gamma_+| = 5$ neighbours the fixed point at is bistable for $p \in [0, 7/30)$ and stable for $p \in [7/30, 0.5]$, as predicted. Since p_{grid} is continuous, stability can be checked by considering the derivative $\partial p_{\text{grid}}/\partial \rho_t = \dot{p}_{\text{grid}}$. If $|\dot{p}_{\text{grid}}|$ is bounded by 1, then the fixed point ρ is attractive, otherwise it is repellent. The derivative with respect to ρ_t is

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

For example, the derivative for $p = 0.15$ is not bounded by 1 for all values of ρ_t. The fixed point $\rho_t = 0.5$ is repellent since at this point $\dot{p}_{\text{grid}}(0.5) \approx 1.359$, and so iteration of p_{grid} will lead away from 0.5. The derivative for $p = 0.35$ is smaller than 1 (≈ 0.672) and so $\rho_t = 0.5$ is an attractive fixed point. It can be seen that for $|\Gamma_+| = 5$ the critical (bifurcation) point is at 0.233, as predicted by theory (Balister et al., 2006). It can also be seen that increasing the neighbourhood size to $|\Gamma_+| = 15$ (right panel) increases the critical point to about 0.32. This increase in critical point corresponds to the simulations in Kozma et al. (2005) where ('long range') edges were added to the nodes, which increased the neighbourhood size.

The dynamics of p^{ν}_{grid} in the random graph G_{rg} are similar to that of the grid. The main difference is that the critical point of the bifurcation is closer to $p = 0.5$. This follows from the previous section where we saw that the critical point increases when the neighbourhood size is increased. As is clear from the definition of p^{ν}_{grid} in (7), the only difference with that of the grid is the neighbourhood size which is increased to $\nu = \lfloor p_e(n - 1) \rfloor$. Fig. 6 shows the result for a graph with $n = 25$ nodes (left panel) and for a graph with $n = 100$ nodes. The approximation of p^{ν}_{grid} is quite accurate, also for the location of the critical point. With a graph of size $n = 100$ the accuracy is such that p_{rg} and p^{ν}_{grid} are nearly indistinguishable, which corresponds to the result in Proposition 3.

The dynamic behaviour of p_{sw} is shown in Fig. 7. Generally, the behaviour is similar to that on the random graph. In Fig. 7 the left panel shows a bifurcation plot of p_{sw} and p^{ν}_{sw} on $G_{\text{sw}}(49, 0.4)$. The accuracy of p^{ν}_{sw} improves greatly for larger n, as seen in the right panel of Fig. 7 for $G_{\text{sw}}(100, 0.4)$. For low values of new edges in the NW small-world p_w, the probability p^{ν}_{sw} is smaller than in the grid. This is because the probability $p^{\text{sw}}_{\text{grid}} = p_{\text{grid}}(1 - p_w)^{n-|\Gamma|}$ is corrected by the number of edges not added to the graph.

In each case we see that for the interval $\rho_t \in (0, 0.5]$ the proportion is either stable or the proportion is bistable. This implies that we expect the binomial process (16) to have most

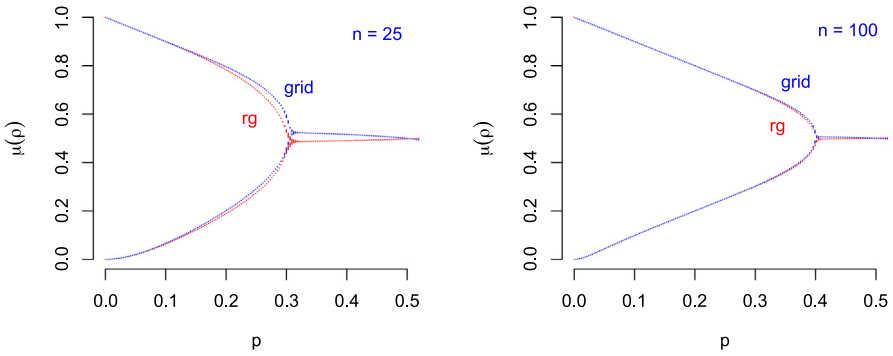

Fig. 6. Plots of μ_{rg} (red) and μ^{v}_{grid} (blue) as a function of p. In the left panel the bifurcation plots are given for a random graph of size $n = 25$ and in the right panel for a random graph of size $n = 100$. All plots are obtained with edge probability $p_e = 0.4$ to guarantee connectedness of the graph.

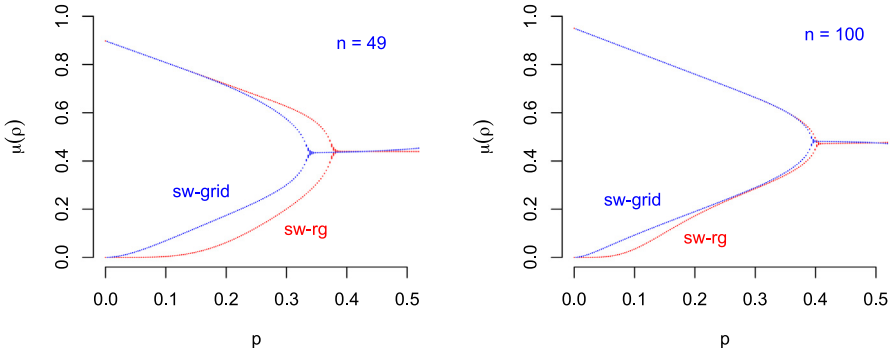

Fig. 7. Bifurcation plots of the small-world mean field $\mu^{v}_{\text{sw}} = p^{v}_{\text{sw}}$ based on the approximation of the random part (in blue) and the mean field $\mu_{\text{sw}} = p_{\text{sw}}$ in (10) based on the random graph (in red). In the left panel a small-world of $n = 49$ nodes and in the right panel a graph of $n = 100$ nodes; all graphs are obtained with the probability of wiring (adding edges) in the NW small-world of $p_w = 0.4$.

probability for $p > p_c$ around a single stable fixed point or when $p < p_c$ most probability is around one of the two stable fixed points. As mentioned above, determining the critical value p_c analytically is difficult and so we turn to numerical estimates in the next section.

5. Numerical evaluation of the mean field

To evaluate the accuracy of the predictions of the mean field in the grid, random, and small-world graph, we simulated networks of different sizes in the topology of a grid, a random graph, and an NW small-world graph. For each combination of parameters, 100 graphs were simulated. In combination with the majority rule the SCA for the grid, random graph or small-world was run for a certain duration T and the states of the last section (10%) of the time series were determined to see if it matches that of the predictions of the mean field. At time point 0, data in $\{0, 1\}$ were generated for each one of the three types of graph, according to the Ising model using the R package *IsingSampler* (Epskamp, 2013). Subsequent values $t > 0$ were obtained for all nodes by the majority rule given the value p for each type of graph. To determine the accuracy we used both 90% and 95% confidence intervals obtained from the central limit theorem (see Section 3.1) for each of the three different graphs. Fig. 8 shows the topology of some simulated random graphs arranged in a square.

We varied the size of the network $n \in \{16, 25, 49, 100\}$, the number of time points $T \in \{50, 100, 200, 500, 5000\}$, and the probability of an active node in the majority rule $p \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$, see (1). We also varied the probability of an edge in the random graph $p_e \in \{0.1, 0.2, \ldots, 0.9\}$, and the probability of wiring in the small-world graph $p_w \in \{0.1, 0.2, \ldots, 0.9\}$. Fig. 9 visualises the evolution of selected simulation conditions. All simulated data, figures, as well as the used R-code are publicly available at the Open Science Framework (Asante et al., 2016).

Results in bifurcation diagrams with 90% and 95% confidence intervals are shown in Fig. 10. The black lines are the bifurcation predictions from the mean field, and the grey area above and below the mean field is the 90% confidence interval, and the grey dotted lines indicate the 95% confidence intervals. The red circles correspond to the last 100 points in the evolution of a particular kind of graph. It can be seen that for each of the different types of graph the densities of the simulated networks (red dots) are mostly within the 95% confidence interval. Except around the critical point (around 0.3) there is some difference between the density of the network and the mean field approximation, especially for smaller time series. Table 1, showing the proportions of the densities corresponding to Fig. 10, confirms the lower proportions in the confidence intervals around the critical point in smaller time series. Note that we did not average the values obtained from the evolution of the graphs, and so the individual fluctuations are also represented by the red circles. All in all, results show that the mean field approximation also performs well when non-regular network structures are considered.

6. Conclusions and discussion

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

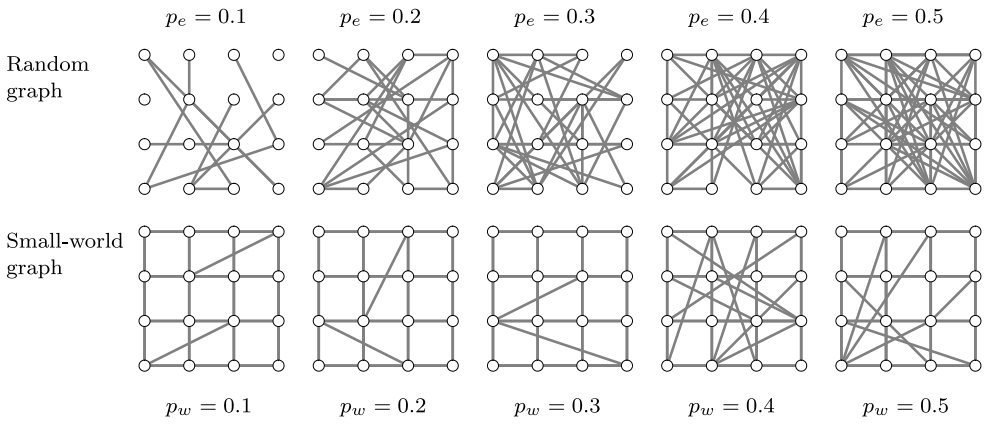

Fig. 8. Examples of simulated random graphs (top row) and small-world graphs (bottom row) with network size $n = 16$. The probabilities of an edge p_e in the random graph and the probability of new edges in the small-world range between 0.1 and 0.5.

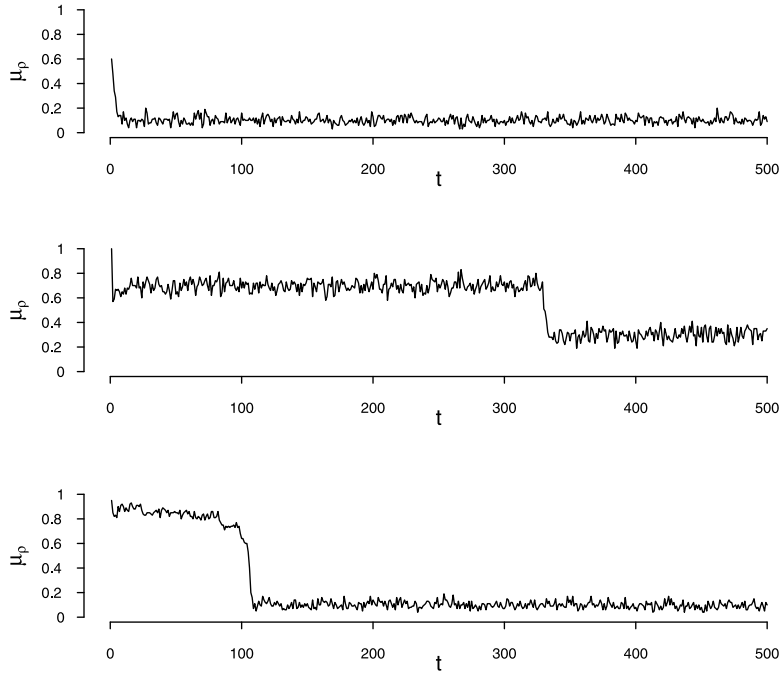

Fig. 9. Examples of the evolution of a torus (upper panel; $p = 0.1$), a random graph (middle panel; $p = 0.3, p_e = 0.6$) and a small world graph (lower panel; $p = 0.1, p_w = 0.9$).

Table 1
Proportion of the densities of active nodes ending up in the confidence interval (90% or 95%). The setup is similar to Fig. 10, where each row corresponds to a type of network and the columns correspond to the size of the network and length of the time series. In each cell the probability p of a 1 is varied, as in Fig. 10.

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

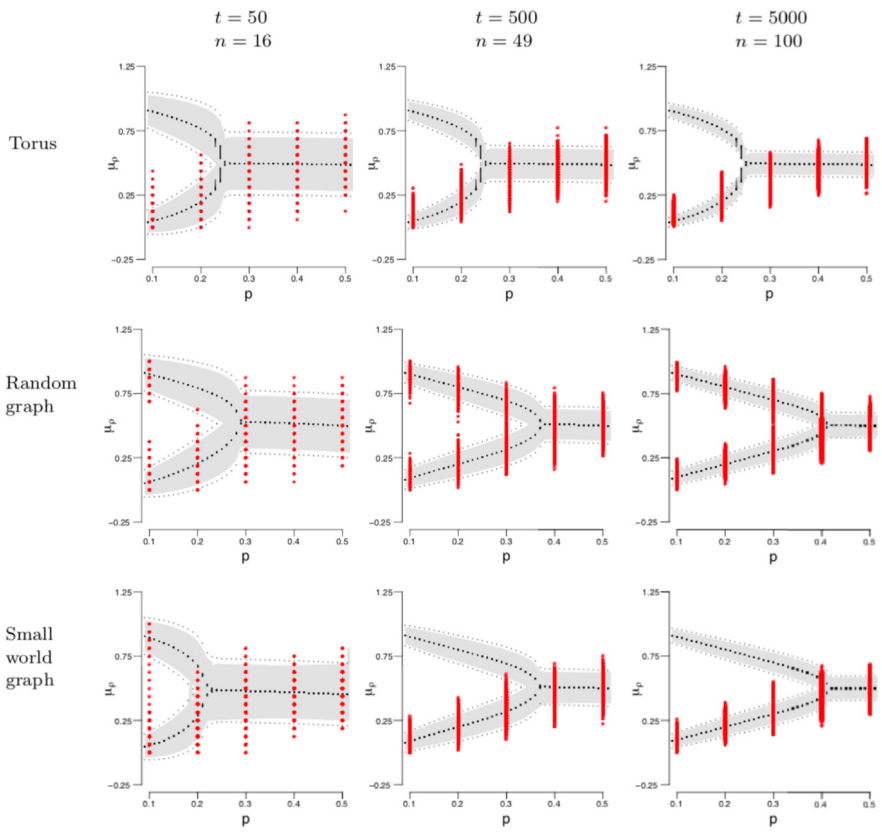

Fig. 10. Bifurcation diagrams of a torus (upper panel), a random graph (middle panel; $p_e = 0.5$) and a small world graph (lower panel; $p_w = 0.6$). Grey solid area = 90% confidence interval around bifurcation. Dashed grey lines = 95% confidence interval around bifurcation. Red dots = mean density estimates at different values of p.

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

Our approximation is based on the formulation of the grid (torus) where a relatively simple sum over possible active nodes determines the probability of a randomly selected node in the graph being active. We showed that for large graphs this approximation is accurate. This simplification could serve to obtain a more extensive analysis of the dynamics such as that presented in Janson et al. (2015). There the majority rule (the stochastic element) was removed from the model, to obtain exact fixed points for the model. Here we chose not to remove the stochastic element since we aim to introduce different rules for updates than the majority rule, like a conditional Ising probability.

The model we used is based on the similarity between the proportion of symptoms in the long run of the model and the empirical data obtained so far for patient and non-patient groups. The 'inertia' of the patient groups (Heath et al., 2007; Kuppens et al., 2010) seems to suggest sudden changes are possible (bistability), while in the general population a 'healthy' variation of symptoms is expected (Kuppens et al., 2010), suggesting a medium level of active symptoms. This corresponds to the mean field model presented here, where, for different a priori probabilities of a node changing value, the proportion of active symptoms is either stable or bistable. In Kossakowski et al. (2019) this idea seems to be confirmed by empirical data, obtained from both a group of diagnosed depressed patients and the general population.

We chose the majority rule in the mean field approach to determine the probability of changing a node to 0 or 1, which induced stochasticity in the cellular automaton. Our motivation for the majority rule was that (i) we do not know what exactly influences each node, and (ii) small changes to a node should not lead to large changes because this would lead to many large fluctuations of the graph. Additionally, we have some evidence that the mean field model including the majority rule is appropriate in several samples (Kossakowski et al., 2019). It would be interesting to see what behaviour other rules would reveal and whether the fit to real data would improve.

Our initial motivation for these results was to obtain a model where we could assess the possibility of a person changing mood suddenly. Based on the estimate of the graph and the corresponding probability of an active node p and its corresponding bifurcation diagram, we would then be able to determine the possibility of that person suddenly changing from few to many active symptoms (low to high proportion) to vice versa. This assessment might be useful in a clinical setting where a decision on a particular type of intervention is required. This idea is pursued in another paper (Kossakowski et al., 2019). The results therein indicate that the mean field model seems reasonable.

Acknowledgments

JK is partly funded by the Research Priority Area Yield, part of the Research Institute of Child Development and Education, University of Amsterdam, the Netherlands, and JK and LW are partly funded by the European Research Council Consolidator Grant received by Prof. Denny Borsboom (grant no. 647209).

Table 2
Most important symbols used in the paper with reference to the associated equation.

p	(1)	Probability for a node to be in state 1 if ≤ 0.5 neighbours are active
$1-p$	(1)	Probability for a node to be in state 1 if > 0.5 neighbours are active
maj_p	(1)	Majority rule as a function of $z_{i,t}$ and Γ
$X_{i,t}$		Random variable of node i at time t, takes values 0 or 1
$\Gamma(i)$		Set of nodes in neighbourhood of node i
Γ		Generic neighbourhood for any node
$Z_{i,t}$		Sum of nodes $X_{i,t}$ at time t in neighbourhood $\Gamma(i)$
p_{grid}	(3)	Probability in a grid of $x_{i,t+1}$ given ρ_t
Y_t	(4)	Sum of all nodes $X_{i,t-1}$ in a graph
ρ_t		proportion of active symptoms
p_{rg}	(6)	Probability in a random graph of $x_{i,t+1}$ given ρ_t
p^{ν}_{grid}	(7)	Approximate probability in a random graph of $x_{i,t+1}$ given ρ_t
p_{sw}	(6)	Probability in a small-world of $x_{i,t+1}$ given ρ_t
$p^{\text{sw}}_{\text{grid}}$	(11)	Probability for a small-world, only the part of the grid
$p_{\text{rg}\setminus\Gamma}(\rho_t)$	(12)	Probability for a small world, only the random graph part
p^{ν}_{sw}	(13)	Approximate probability in a small-world of $x_{i,t+1}$ given ρ_t
K_t		Markov kernel, $(n+1)\times(n+1)$ matrix with transition probabilities
$K_{0,t}$	(17)	Markov kernel product of K_t from 0 up to t
π_t	(17)	$n+1$ probability vector of a Markov chain

Appendix

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

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

and similarly, define $h_-(\varepsilon) = h_+(-\varepsilon)$. Then Chernov's bound (Lesigne, 2005; Venkatesh, 2013) for the density ρ_t of a grid with n nodes and its mean at time t, $p_{\text{grid}}(\rho_t)$ defined in (2), for $0 < \varepsilon < \min\{p_{\text{grid}}, 1-p_{\text{grid}}\}$ immediately gives

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

The Kullback–Leibler divergence can be approximated quadratically by

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

Similarly for $h_-(\varepsilon)$ gives

$$\mathbb{P}(|\rho_t - p_{\text{grid}}(\rho_t)| > \varepsilon) \leq 2\exp(-\varepsilon^2/2\sigma^2_{\text{grid}}(\rho_t)) \tag{18}$$

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

Proof (*Eq.* (8))**.** To obtain (8) assume a fixed value ν for all neighbourhood sizes k, maj_p uses the value $k=\nu$. Then maj_p only depends on the number of active neighbours r. First note that

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

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

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

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

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

which leads to (8).

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

$$p_{\text{rg}}(\rho_t) - p^{\nu}_{\text{grid}}(\rho_t) = \sum_{k=0}^{n-1}\sum_{r=0}^{k}\binom{k}{r}\rho_t^r(1-\rho_t)^{k-r}\binom{n-1}{k}p_e^k \times (1-p_e)^{n-k-1}[\text{maj}_p(r,k) - \text{maj}_p(r,\nu)]$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

completing the proof. □

Proof (*Lemma* 6)**.** We showed already in Section 4.1 that the transition kernel is a contraction and so with t increasing we will decrease the distance in total variation to 0 between $\pi_0 K_{0,t}$ and $\pi_0' K_{0,t}$. For this to work we need that for each t, $K_{t,t+1}$, the single step transition probability, is irreducible. This implies that the transition probability $\mathbb{P}(Y_{t+1} = m \mid Y_t = k)$ from state k to state m

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

with p_G the probability determined by the type of graph G, a grid, random or small-world graph. We consider each type of graph in turn.

For the grid we have

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

in (16), which is bounded by p and $1-p$. And so if $p \in (0,1)$, then $K_{t,t+1}$ is irreducible.

For the random graph, where we used the approximation by increasing the neighbourhood, i.e., we use p^{ν}_{grid}, where $\nu = \lfloor p_e(n-1) \rfloor$, we have a similar situation since there instead of $|\Gamma|$ we use ν.

For the small-world graph it is clear from Section 3.3 that p_{sw} converges to 0 as $t \to \infty$ when ρ_t is in the stable set $S(0)$ of p^{ν}_{sw}. Hence we require that ρ_t is outside of $S(0)$. Likewise, we require that ρ_t is not in the stable set $S(1)$. So, only if p_{grid} converges to 0 will $K_{t,t+1}$ converge to 0. For any p in the majority function maj_p such that it is not in $S(0) \cup S(1)$ we define $\eta_t = \min_{t,k,m} K_{t,t+1}(k,m)$, which implies that for each time point t and any states k and m there is an η_t such that $K_{t,t+1}(k,m) \geq \eta_t$, which is uniform (over states) irreducibility. We can therefore conclude that whenever ρ_t is not in $S(0) \cup S(1)$, then we obtain merging. Since

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

and

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

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

References

Asante, K., Barbour, E., Barker, L., Benjamin, M., Bowman, S. D., Boughton, A. P., et al. (2016). Open science framework. https://osf.io/ewf2g/.

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

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

Bollobás, B. (2001). *Random graphs*. Springer.

Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PLoS One*, *6*, e27407.

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

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

Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., et al. (2019). Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature Human Behaviour*, *3*, 478.

Durrett, R. (2007). *Random graph dynamics*. Cambridge University Press.

Ebner-Priemer, U. W., Houben, M., Santangelo, P., Kleindienst, N., Tuerlinckx, F., Oravecz, Z., et al. (2015). Unraveling affective dysregulation in borderline personality disorder: a theoretical model and empirical evidence. *Journal of Abnormal Psychology*, *124*, 186.

Epskamp, S. (2013). IsingSampler: Sampling methods and distribution functions for the Ising model. URL github.com/SachaEpskamp/IsingSampler.

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

Haslbeck, J., & Waldorp, L. (2020). mgm: Estimating Time-Varying Mixed Graphical Models in High-Dimensional Data. *Journal of Statistical Software*, *93*(8).

Heath, R. (2015). Detecting nonlinearity and edge-of-chaos phenomena in ordinal data. *Nonlinear Dynamics, Psychology, and Life Sciences*, *19*, 229–248.

Heath, R. A., Heiby, E. M., & Pagano, I. S. (2007). Complex dynamics in depression: an application to long-term, mood-rating time series. In W. J. J. Neufeld (Ed.), *Advances in clinical cognitive science: Formal modeling of processes and symptoms* (pp. 263–291). Washington: American Psychological Association.

Hirsch, M. W., Smale, S., & Devaney, R. L. (2004). *Differential equations, dynamical systems, and an introduction to chaos* (2nd ed.). Academic Press.

Holmgren, R. (1996). *A first course in discrete dynamical systems*. Springer Science & Business Media.

Hosenfeld, B., Bos, E. H., Wardenaar, K. J., Conradi, H. J. J., 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.

Janson, S., Kozma, R., Ruszinkó, M., & Sokolov, Y. (2015). Bootstrap percolation on a random graph coupled with a lattice. arXiv preprint arXiv:1507.07997.

Janson, S., Łuczak, T., Vallier, T., et al. (2012). Bootstrap percolation on the random graph $G_{n,p}$. *Annals of Applied Probability*, *22*, 1989–2047.

Kindermann, R., Snell, J. L., et al. (1980). *Markov random fields and their applications*: *Vol. 1*, American Mathematical Society Providence, RI.

Kleczkowski, A., & Grenfell, B. T. (1999). Mean-field-type equations for spread of epidemics: The 'small world'model. *Physica A. Statistical Mechanics and its Applications*, *274*, 355–360.

Kossakowski, J. J., Gordijn, M. C. M., Harriette, R., & Waldorp, L. (2019). Applying a dynamical systems model and network theory to major depressive disorder. *Frontiers in Psychology: Quantitative Psychology and Measurement*.

Kozma, R., Puljic, M., Balister, P., Bollobas, B., & Freeman, W. (2004). Neuropercolation: A random cellular automata approach to spatio-temporal neurodynamics. In P. M. A. Sloot, B. Chopard, & A. G. G.Hoekstra (Eds.), *Lecture Notes in Computer Science*: *Vol. 3305*, *Cellular Automata* (pp. 435–443). Springer Berlin Heidelberg.

Kozma, R., Puljic, M., Balister, P., Bollobás, B., & Freeman, W. J. (2005). Phase transitions in the neuropercolation model of neural populations with mixed local and non-local interactions. *Biological Cybernetics*, *92*, 367–379.

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, *21*, 984–991.
Lesigne, E. (2005). *Heads or tails: An introduction to limit theorems in probability*. American Mathematical Society.
Levin, D. A., Peres, Y., & Wilmer, E. L. (2017). *Markov chains and mixing times*. American Mathematical Association.
Lodewyckx, T., Tuerlinckx, F., Kuppens, P., Allen, N. B., & Sheeber, L. (2011). A hierarchical state space approach to affective dynamics. *Journal of Mathematical Psychology*, *55*, 68–83.
Mott, J., & Schneider, H. (1957). Matrix norms applied to weakly ergodic Markov chains. *Archiv der Mathematik*, *8*, 331–333.
Newman, M. E., & Watts, D. J. (1999). Scaling and percolation in the small-world network model. *Physical Review E*, *60*, 7332.
Norris, J. R. (1997). *Markov chains*. Cambridge University Press.
O'Donnell, R. (2014). *Analysis of boolean functions*. Cambridge University Press.
Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2011). A hierarchical latent stochastic differential equation model for affective dynamics. *Psychological Methods*, *16*, 468.
Paz, A. (1971). *Introduction to probabilistic automata*. Academic Press.
Saloff-Coste, L., & Zúñiga, J. (2010). Merging and stability for time inhomogeneous finite Markov chains. arXiv preprint arXiv:1004.2296.
Saloff-Coste, L., Zuniga, J., et al. (2009). Merging for time inhomogeneous finite Markov chains, Part I: Singular values and stability. *Electronic Journal of Probability*, *14*, 1456–1494.
Sarkar, P. (2000). A brief history of cellular automata. *ACM Computing Surveys*, *32*, 80–107.
Sethna, J. P. P., Dahmen, K. A., & Perkovic, O. (2004). Random-field ising models of hysteresis. arXiv preprint cond-mat/0406320.
Sporns, O., & Honey, C. J. J. (2006). Small worlds inside big brains. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 19219–19220.
Tomassini, M., Giacobini, M., & Darabos, C. (2005). Evolution and dynamics of small-world cellular automata. *Complex Systems*, *15*, 261–284.
van Borkulo, D., Epskamp, S., Blanken, T. F., Boschloo, L., Schoevers, R. A., & Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Scientific Reports*, *4*.
Venkatesh, S. (2013). *The theory of probability*. Cambridge University Press.
Von Neumann, J. (1951). The general and logical theory of automata. *Cerebral Mechanisms in Behavior*, 1–41.
Wainwright, M. J., & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, *1*, 1–305.
Watts, D. J. (1999). *Small-worlds: The dynamics of networks between order and randomness*. Princeton University Press.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440–442.
Wolfram, S. (1984a). Cellular automata as models of complexity. *Nature*, *311*, 419–424.
Wolfram, S. S. (1984b). Computation theory of cellular automata. *Communications in Mathematical Physics*, *96*, 15–57.