Relations between the stability of Boolean networks and percolation

Candidate number: 896065*
University of Oxford

Boolean networks have been extensively studied as models of gene regulatory networks. On the other hand, percolation models have been important in understanding many disordered systems. Here I look at some interesting applications of percolation theory to the understanding of the transition between stable and unstable phases in Boolean networks.

CONTENTS

I. INTRODUCTION

Boolean networks were introduced by Kauffman in 1969 [1] with the main purpose of understanding the small number of cell types observed in biology, which are identified with the possible limit cycles of the discrete dynamical system [2, 3], and they since have become a prominent tool for modeling gene regulatory networks [4]. One crucial feature, if these networks are to describe real biological systems, is their stability against small random damages. Thus, this aspect of the network dynamics has been a focus of study in the field. Kauffman himself observed in his early work that, for Boolean networks with random topology (but fixed number of nodes and in-degree) and random update functions, small perturbation died out if the in-degree was $K = 1$ or $K = 2$, but could grow to macroscopic sizes if the in-degree was $K > 2$.

This result was understood analytically [5, 6] by using the 'annealed approximation', in which the network topology and update functions randomly change at each time step. It was shown [5, 7, 8] that some results of this approximation agree with those of the original 'quenched' system in the thermodynamic limit (when the number of nodes in the network $N \to \infty$). In particular, they looked at the long-time limit of the Hamming distance (also called 'overlap', or 'damage' in this context) between network configurations with slightly different initial conditions, as this is a measure of the stability or instability of small initial perturbations. They found this quantity had the same behavior in the quenched and annealed models, justifying their analysis.

Soon after, this analysis was extended to Boolean networks in the lattice, by exploiting a mapping between the evolution of the Hamming distance on a d-dimensional lattice and a directed percolation process on a particular $d + 1$-dimensional lattice [9]. This mapping generalizes one proposed for 1-dimensional cellular automata before [10]. Other more qualitative analogies between the growth of clusters of "stable" and "oscillatory" sites and percolating clusters were also noted in numerical simulations [2, 11].

More recently, this treatment has been extended to the case where the network topology is arbitrary [12] but fixed, while the update functions change at every time step, which has been called the "semiannealed" approximation [13]. In this case, the evolution of the Hamming distance maps into weighted percolation on a configuration-model directed network [14]. The case where the update functions are restricted to be of a certain class of interest is also studied in that paper.

The reason these analogies can be useful is that they allow to utilize some of the insights and results obtained in the vast field of percolation theory, which has been steadily active ever since its conception six decades ago. For a review of recent advances see [15]. In this letter, I look at a selection of the precise analogies between Boolean networks and percolation that have been proposed.

II. BOOLEAN NETWORK MODEL

A Boolean network is a directed network of N nodes, in which each node i is in one of two states, $x_i(t) = 0$ or $x_i(t) = 1$, at each discrete time t. At each time step, each node i updates it state depending on the states of the nodes that are adjacent to it. This is encoded in a Boolean function F_i, specified for each node:

$$x_i(t) = F_i(\{x_j(t-1) : j \in \mathcal{J}_i\})$$

* REVTeX Support: revtex@aps.org

where $\mathcal{J}_i$ is the set of nodes adjacent to i. A random Boolean network (RBN) refers to a Boolean network with a network topology drawn from some random graph ensemble, and Boolean functions drawn according to some probability distribution. For a quenched RBN, both the network topology and Boolean functions are subsequently fixed during the dynamical evolution.

Commonly used random graph ensembles are those where the number of nodes N is fixed and edges are added uniformly at random while fixing either the in-degree, the in-degree distribution, or the joint in- and out-degree distribution [14, 16]. For Boolean functions, common probability distributions are the uniform distribution or a Bernoulli distribution for the output of each of the possible inputs. In the later case, the probability p_i of the output being a 1 for each input of F_i, is called the "bias probability". A common case is where $p_i = p$ for all i.

III. ANNEALED APPROXIMATION AND WEIGHTED SITE PERCOLATION ON CONFIGURATION-MODEL DIRECTED NETWORKS

The annealed approximation was introduced by Derrida [5] for the case of networks with fixed in-degree K and uniform bias p. The method consists in considering a RBN for which the network topology and Boolean functions are randomly sampled (from some fixed distributions) at every time step. This simplifies the calculation because it removes the correlations that arise between network states and the 'frozen' network topology and Boolean functions. His original treatment has been extended to the case of arbitrary joint in- and out-degree distributions P_{jk} [14, 17].

The question of stability is approached by looking at the spread of "damage". Two initial configurations differing on a fraction ϵ of nodes are evolved in the annealed network, resulting in two different trajectories through the state space of the network, $\mathbf{x}(t)$, and $\tilde{\mathbf{x}}(t)$. At time t, a node i is said to be "damaged" if $x_i(t) \neq \tilde{x}_i(t)$. y_i is defined as the fraction of the time T that node i is damaged, as $T \to \infty$. When averaged over all nodes, and over all initial conditions (differing by a fixed fraction ϵ of nodes), we get $Y = \langle y_i \rangle$, an order parameter that, in the thermodynamic limit, is 0 in the stable or ordered phase, and non-zero in the unstable or chaotic phase.

To find an equation for Y, we approach the problem in a similar way to the generating function approach in random networks [14, 18, 19]. In the annealed approximation, Y is equal to the long-time limit probability that a node is damaged. A node can only become damaged if at least one of its input neighbours is damaged (as its state is a deterministic function of them). Furthermore, the node becomes damaged if F_i produces a different output for the two differing sets of inputs. This happens if $F_i(\{x_j(t-1) : j \in \mathcal{J}_i\}) = 0$ and $F_i(\{\tilde{x}_j(t-1) : j \in \mathcal{J}_i\}) = 1$, or viceversa. Both of these possibilities happen with probability $p_i(1-p_i)$. Therefore, the probability it becomes damaged if at least one of its neighbours is damaged is $q_i = 2p_i(1-p_i)$. This quantity is called the sensitivity [20]. We consider here the case where the bias probability only depends on the in-degree j and out-degree k of the node (which happen with probability P_{jk}), so that $p_i \to p_{jk}$ and $q_i \to q_{jk}$.

To complete the expression for Y, we need the probability that at least one of the input neighbours of a node is damaged. This can be written as $1-(1-E^*)^j$, where E^* is the long-time limit of the probability that an edge originates from a damaged node (which at any finite time is E_t). Putting this together we have,

$$Y = \sum_{j,k} P_{jk} q_{jk} [1 - (1 - E^*)^j] \tag{1}$$

We need to find an equation for E^*. This long-time limit is just a fixed point of a discrete map for the probability E_t that can be constructed as follows. At time $t+1$, E_{t+1} is the probability that the node from which a random edge originates is damaged. If that node has in-degree j, and out-degree k, this is $q_{jk}[1-(1-E_t)^j]$, as above. Furthermore, a node at the end of a randomly chosen edge has in and out degree j and k with probability $\frac{kP_{jk}}{z}$, where $z = \sum_{j,k} jP_{jk} = \sum_{j,k} kP_{jk}$ is the mean degree [19]. Therefore the self-consistency equation for the fixed point E^* is

$$E^* = \sum_{j,k} \frac{kP_{jk}}{z} q_{jk} [1 - (1 - E^*)^j] \tag{2}$$

This derivation is very similar to the derivation of the size the giant component in a random graph [18, 19, 21]. In fact, the analogy can be made precise [14] if we consider weighted site percolation on a configuration-model directed graph, with joint degree distribution P_{jk}. We let the probability of removing a site depend on its degrees and be q_{jk}. Then above the percolation transition, a giant strongly connected component appears. If we look at this component, plus all the nodes that can be reached from it, we have the giant out-component or GOUT.

The mapping is based on identifying the nodes in the GOUT with the damaged nodes in the above damage spreading process. This makes intuitive sense because the size of the GOUT can be found by a spreading process, where the spreading property is whether a node belongs to the GOUT or not [19]. Indeed, if S is the expected fraction of nodes in the GOUT, and u is the probability that a random edge leads to a node *not* in the GOUT, Newman *et al.* find [21]

$$S = F_0(1) - F_0(u) \tag{3}$$

$$u = 1 - F_1(1) + F_1(u) \tag{4}$$

where the generating functions are defined as $F_0(w) = \sum_{j,k} P_{j,k} q_{j,k} w^j$ and $F_1(w) = \sum_{j,k} \frac{kP_{j,k}}{z} q_{j,k} w^j$. One can

see that Eq. 3 maps to Eq. 1, and Eq. 4 to Eq. 2, under the substitution $Y = S$, and $u = 1 - E$. Therefore the stability/instability phase transition in this ensemble of RBN is equivalent to the percolation phase transition in the corresponding random graph ensemble (with the same joint degree distribution).

IV. SEMIANNEALED APPROXIMATION AND WEIGHTED SITE PERCOLATION ON A DIRECTED NETWORK

The semiannealed approximation was introduced by Pomerance *et al.* [13] as an extension of the annealed approximation to the case of fixed network topology. In this approximation, only the Boolean functions are randomized at each time step. As a result, one often considers assigning a bias probability to each node: p_i for node i. In this case, we should work out the long-time probability of being damaged for each node, y_i, because in a fixed topology nodes are not statistically equivalent (as they were for the annealed case). Furthermore, as we know the neighbours of node i ($\mathcal{J}_i$), it is straightforward to write the self-consistency equation for y_i as

$$y_i = q_i[1 - \prod_{j \in \mathcal{J}_i} (1 - y_j)] \tag{5}$$

where $q_i = 2p_i(1-p_i)$ is the sensitivity of node i. This equation expresses that a node becomes damaged with probability q_i if at least one of its neighbours is damaged. As noted in [14], this equation also has an analogue in the problem of weighted site percolation on a directed network with the same topology as the RBN. As derived in [22], for a site deletion probability $1 - q_i$, the fraction of site-deleted networks in which node i is not in the GOUT, defined as η_i, obeys the equation:

$$\eta_i = 1 - q_i + q_i \prod_{j \in \mathcal{J}_i} \eta_j \tag{6}$$

This comes from noting that a node is not in the GOUT if it is either deleted or it is not, but none of its neighbours is in the GOUT. Eqs. 6 and 5 are equivalent under the substitution $y_i = 1 - \eta_i$. Furthermore, as the fraction of nodes in the GOUT is $S = \langle 1 - \eta_i \rangle$, and $Y = \langle y_i \rangle$, Y again maps to S. Both Eqs. 6 and 5 are valid under the assumption that the networks are locally tree-like, so that the probabilities appearing in the product could be treated as independent.

V. ANNEALED APPROXIMATION ON A BOOLEAN LATTICE AND DIRECTED PERCOLATION

The mapping between the annealed approximation to Boolean dynamics on a d-dimensional lattice and directed percolation in a corresponding $d+1$-dimensional lattice was discovered much earlier [9]. In fact the treatment corresponds to a semiannealed approximation because the network topology is fixed to a lattice (a square lattice in their case), while the Boolean functions are randomized at each step.

It is interesting to compare their approach to those described above. The main difference is that the square lattice in 2 dimensions (or hypercubic in d dimensions) is very far from locally tree-like, as it is in fact built out of small loops. For this reason, the above approach is not really applicable. However, if one "unfolds" the lattice into an extra dimension corresponding to time, with edges directed forward in time, the network becomes tree-like (there are no loops, taking direction into account). The unfolded network is a $d+1$-dimensional lattice, where each node in the time-$t+1$ layer is connected to its neighbours (from the original lattice) in the time-t layer. For a 1-dimensional chain with only nearest-neighbours connections, the unfolded lattice is composed of two independent square lattices. Similarly, a 2-dimensional square lattice with nearest-neighbours connections gives two independent BCC lattices.

In directed site percolation, one considers the spread of a property through a directed network, after a fraction of nodes have been removed. In our case, we can consider the spread of damage. A node is never damaged at time $t+1$ if its neighbours at time t aren't damaged. On the other hand, if any of its neighbours at time t are damaged, the node is damaged at time $t+1$ with probability $q = 2p(1-p)$, the sensitivity. The spread of damage (in the semiannealed approximation) is thus equivalent to directed site percolation on the $d+1$-dimensional lattice, with occupation probability q. The long-time average damage Y corresponds to the fraction of nodes in the giant component.

For the chain, with only nearest neighbours interactions (in-degree $K = 2$), these results imply that there is no chaotic phase. This is because q ranges between 0 and 1/2, and cannot reach the square lattice percolation threshold, $q_c \approx 0.705$ [23]. However, if one includes next nearest neighbours ($K = 4$), the percolation threshold is lowered. Intuitively, one expects the percolation to be about half of that of the $K = 2$ case. Indeed, it is numerically found [9] that $p_c \approx 0.25$, corresponding to $q_c \approx 0.375$, not far from 0.705/2. The observed percolation threshold in Fig. 2 is consistent with that result.

VI. NUMERICAL SIMULATION

Numerical simulations were performed to test the above relations. Figure 1 compares the computed mean values of S and Y for a particular simple class of configuration model networks. The network size was $N = 1000$. For two identical random initial configurations (each site with an state chosen uniformly and independently at random), 10 random sites were chosen to be damaged. The Boolean network was then evolved for 50 time steps, and

the last 20 were used to calculate the average Y. Y was then averaged over 50 initial conditions, and over 10 random Boolean networks (uniformly random Boolean functions, and a random graph ensemble described in Fig. 1). S is computed from a single site percolation trial per random Boolean network. With these parameters finite-size effects still produce noticeable deviations after the transition. Simulations for large networks, showing better agreement, are found in [14].

Figure 2 shows the results for the 1-dimensional Boolean lattice, with nearest and next-nearest-neighbours interactions. The network size is $N = 100$, and was run for 50 time steps, where the last 30 were used to find Y, which was then averaged over 100 random initial conditions, and over 10 RBNs with random Boolean functions with bias probability p.

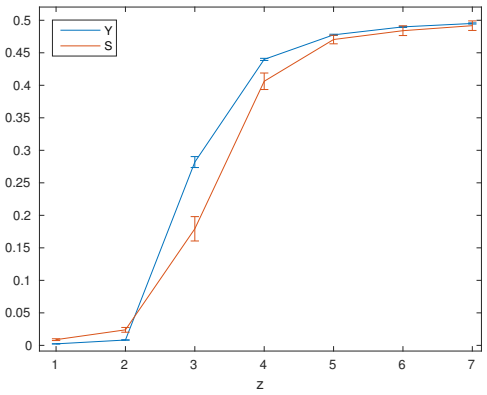

FIG. 1. Mean values for S, as defined in [14], and long-time average damage Y, versus mean degree z for undirected configuration model networks, with degree distributed uniformly between $z-1$ and $z+1$. Error bars represent standard errors of the means, calculated as in [14].

There are several modeling issues to note, mostly related to finite size effects. The expressions derived above rely are only valid on the asymptotic limit of large networks (number of nodes $N \to \infty$). To begin with, the annealing approximation is only valid in the $N \to \infty$ limit [7] (this result actually applies for Kauffman's original RBNs where the topology is random, with fixed in-degree. Whether the annealed approximation is valid, even in the large N limit for the more general networks considered here is not obvious, given that it does fail for some classes of RBNs [24]). Indeed, in [14], they find that the results from the quenched disorder caused observable deviations, for the network sizes used, at least.

Regarding percolation, finite size effects are also very important, given that the giant component is only unambiguously defined in the $N \to \infty$ limit. One has to choose a definition for the finite case that agrees with the definition in the asymptotic limit. In [14], they find that defining the quantity S as the fraction of nodes in any SCC (where an SCC is defined to be a mutually path-connected set of nodes, with at least one loop) works well. In my simulations, for the case of directed percolation in

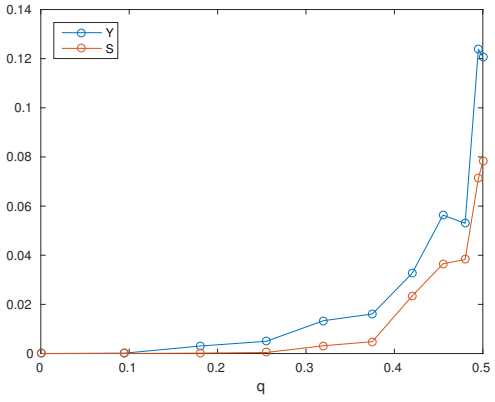

FIG. 2. Long-time average damage Y (red), versus the sensitivity $q = 2p(1-p)$, for the 1-dimensional Boolean lattice, with $K = 4$; and mean fraction of nodes in an SCC (as defined in the text), S (blue), for the associated $1+1$-dimensional lattice. Error bars are smaller than the circles.

lattices, I found that defining S as the fraction of nodes in an SCC, averaged over the SCCs to which nodes in in the $t = 0$ layer belong, gave better results than the size of the largest SCC, for the lattice with $K = 4$ nearest neighbours.

The locally tree-like assumption also boils down to a finite-size issue, because the density of loops in most random graph ensembles considered goes to 0 as $N \to \infty$, making the equations derived above exact in that limit. Small loops, can make a finite amount of damage persist in them even below the threshold. This contribution to Y is not predicted by the equations above. Another finite-size effect that turns out to be more important [14] is a version of the "gambler's ruin" effect, by which fluctuations sometimes cause individual runs to have 0 damage (for all nodes), and necessarily stay so for all subsequent time. This can happen even if the expected value Y from the equations above is non-zero, and causes a systematic error between the predicted and actual Y for some fraction of networks near the critical point, specially for smaller networks.

VII. CONCLUSION

We have looked at some ways in which the problem of stability in different classes of random Boolean networks can be mapped to corresponding percolation problems, under some approximations. Boolean networks are used to model, among other things, gene regulatory networks. However, it is important to note that they do hide many details, which can be important in certain real systems [4].

These mappings allow to use known results from percolation theory to find quantities of interest for RBNs. For instance, estimates of the critical bias probability in Boolean lattices can be obtained from directed percolation thresholds in higher dimensional lattices. Other

quantities like critical exponents may also be found by exploiting this analogy, although one must note that quenched Boolean networks appear to belong to a different universality class than annealed ones [25].

Finally, an extension of the method used by Stauffer *et al.* to study Boolean lattices by converting time into an extra spatial dimension could be possible to study more general Boolean networks that strongly violate the locally tree-like assumption. The corresponding percolation problem will in general have no analytical solution. However, one may be able to predict the stability of the network efficiently using existing algorithms to simulate percolation, as has already been done for locally tree-like networks above and in [14].

[1] Stuart A Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," Journal of theoretical biology **22**, 437–467 (1969).

[2] D Stauffer, "Random boolean networks: Analogy with percolation," Philosophical Magazine Part B **56**, 901–916 (1987).

[3] Stuart A Kauffman, "Emergent properties in random complex automata," Physica D: Nonlinear Phenomena **10**, 145–156 (1984).

[4] Hidde De Jong, "Modeling and simulation of genetic regulatory systems: a literature review," Journal of computational biology **9**, 67–103 (2002).

[5] Bernard Derrida and Yves Pomeau, "Random networks of automata: a simple annealed approximation," EPL (Europhysics Letters) **1**, 45 (1986).

[6] B Derrida and G Weisbuch, "Evolution of overlaps between configurations in random boolean networks," Journal de physique **47**, 1297–1303 (1986).

[7] HJ Hilhorst and M Nijmeijer, "On the approach of the stationary state in kauffman's random boolean network," Journal de Physique **48**, 185–191 (1987).

[8] U Bastolla and Giorgio Parisi, "Closing probabilities in the kauffman model: an annealed computation," Physica D: Nonlinear Phenomena **98**, 1–25 (1996).

[9] Bernard Derrida and Dietrich Stauffer, "Phase transitions in two-dimensional kauffman cellular automata," EPL (Europhysics Letters) **2**, 739 (1986).

[10] Eytan Domany and Wolfgang Kinzel, "Equivalence of cellular automata to ising models and directed percolation," Physical review letters **53**, 311 (1984).

[11] G Weisbuch and D Stauffer, "Phase transition in cellular random boolean nets," Journal de physique **48**, 11–18 (1987).

[12] The networks are assumed to be locally tree-like, although this requirement may not be as strong in practice [26].

[13] Andrew Pomerance, Edward Ott, Michelle Girvan, and Wolfgang Losert, "The effect of network topology on the stability of discrete state models of genetic control," Proceedings of the National Academy of Sciences **106**, 8209–8214 (2009).

[14] Shane Squires, Edward Ott, and Michelle Girvan, "Dynamical instability in boolean networks as a percolation problem," Physical review letters **109**, 085701 (2012).

[15] Abbas Ali Saberi, "Recent advances in percolation theory and its applications," Physics Reports **578**, 1–32 (2015).

[16] Bartolo Luque and Ricard V Solé, "Phase transitions in random networks: simple analytic determination of critical points," Physical Review E **55**, 257 (1997).

[17] Deok-Sun Lee and Heiko Rieger, "Comparative study of the transcriptional regulatory networks of e. coli and yeast: Structural characteristics leading to marginal dynamic stability," Journal of theoretical biology **248**, 618–626 (2007).

[18] Mark EJ Newman, Steven H Strogatz, and Duncan J Watts, "Random graphs with arbitrary degree distributions and their applications," Physical review E **64**, 026118 (2001).

[19] Mark Newman, *Networks: an introduction* (OUP Oxford, 2010).

[20] Ilya Shmulevich and Stuart A Kauffman, "Activities and sensitivities in boolean network models," Physical review letters **93**, 048701 (2004).

[21] Duncan S Callaway, Mark EJ Newman, Steven H Strogatz, and Duncan J Watts, "Network robustness and fragility: Percolation on random graphs," Physical review letters **85**, 5468 (2000).

[22] Juan G Restrepo, Edward Ott, and Brian R Hunt, "Weighted percolation on directed networks," Physical review letters **100**, 058701 (2008).

[23] Iwan Jensen, "Low-density series expansions for directed percolation: Iii. some two-dimensional lattices," Journal of Physics A: Mathematical and General **37**, 6899 (2004).

[24] Agnes Szejka, Tamara Mihaljev, and Barbara Drossel, "The phase diagram of random threshold networks," New Journal of Physics **10**, 063009 (2008).

[25] SP Obukhov and D Stauffer, "Upper critical dimension of kauffman cellular automata," Journal of Physics A: Mathematical and General **22**, 1715 (1989).

[26] Sergey Melnik, Adam Hackett, Mason A Porter, Peter J Mucha, and James P Gleeson, "The unreasonable effectiveness of tree-based theory for networks with clustering," Physical Review E **83**, 036112 (2011).