

Overcoming the complexity barrier of the dynamic message-passing method in networks with fat-tailed degree distributions

Giuseppe Torrisi ^{*}, Alessia Annibale , and Reimer Kühn *Department of Mathematics, King's College London, Strand, and London WC2R 2LS, United Kingdom* (Received 29 April 2021; revised 24 June 2021; accepted 10 October 2021; published 29 October 2021)

The dynamic cavity method provides the most efficient way to evaluate probabilities of dynamic trajectories in systems of stochastic units with unidirectional sparse interactions. It is closely related to sum-product algorithms widely used to compute marginal functions from complicated global functions of many variables, with applications in disordered systems, combinatorial optimization, and computer science. However, the complexity of the cavity approach grows exponentially with the in-degrees of the interacting units, which creates a defacto barrier for the successful analysis of systems with fat-tailed in-degree distributions. In this paper, we present a dynamic programming algorithm that overcomes this barrier by reducing the computational complexity in the in-degrees from exponential to quadratic, whenever couplings are chosen randomly from (or can be approximated in terms of) discrete, possibly unit-dependent, sets of equidistant values. As a case study, we analyze the dynamics of a random Boolean network with a fat-tailed degree distribution and fully asymmetric binary $\pm J$ couplings, and we use the power of the algorithm to unlock the noise-dependent heterogeneity of stationary node activation patterns in such a system.

DOI: [10.1103/PhysRevE.104.045313](https://doi.org/10.1103/PhysRevE.104.045313)

I. INTRODUCTION

The collective behavior of a broad range of disordered and complex systems can be analyzed in terms of models of interacting binary units, evolving according to stochastic Boolean functions of the neighboring units. They include spin glasses [1–3] in physics, gene-regulatory and immune networks [4–8] in biology, artificial neural networks [9–12] in computer science, agent-based models [13–16] and models of operational or credit risk [17,18] in economics and finance, and a variety of hard combinatorial optimization problems [19–23] to name but a few.

Whenever the system dynamics satisfies detailed balance, equilibrium statistical mechanics provides a powerful arsenal of techniques to analyze their operation at stationarity. However, when the dynamics does not satisfy detailed balance—as in systems that are driven, dissipative or exhibit a degree of asymmetry in the interactions—or when one is interested in genuine dynamical aspects of the problem, one has to resort to tools of nonequilibrium statistical mechanics. These are typically much more cumbersome to use, involving, e.g., non-Markovian dynamics of an “effective” local degree of freedom which requires solving dynamic self-consistency equations for global correlation and response functions in the case of fully connected systems [24,25]. In systems with parallel dynamics, the complexity of the analysis grows exponentially in the number of time steps considered [26] and quickly becomes infeasible. If systems exhibit fully asymmetric interactions though, the analysis simplifies considerably [27] and can often be analyzed in some explicit detail.

Techniques for fully connected systems carry over to systems with diluted interactions if the mean connectivity diverges in the large-system limit [28,29]. More recently, however, there has been considerable interest in the dynamics of genuinely sparse systems, defined on complex networks which remain *finitely coordinated* in the thermodynamic limit [30–33]. In these systems, approximation schemes successfully used for dense systems, such as heterogeneous mean-field and Thouless, Anderson and Palmer (TAP) approaches [34,35], have been shown to be ineffective [36,37]. On the other hand, the dynamic cavity method [38–41] is suitable to analyze sparse systems, where short loops are rare. This method requires us to follow the evolution of a collection of *local* dynamical variables, whose number is proportional to system size, in a *cavity* graph, where, however, the dynamics is non-Markovian whenever there is some degree of symmetry in the interactions. This leads to the aforementioned exponential scaling with the time horizon considered. Only when interactions are fully asymmetric, those retarded self-interactions that make dynamics non-Markovian [29,38,39,42] are absent and the exponential time complexity is removed. For this reason, systems with fully asymmetric interactions, which can be analyzed within a Markovian framework, are a particularly attractive option to use for a first level of analysis. However, for a large class of models (including linear and nonlinear threshold models) widely used in physics, biology, finance, and social science, the evaluation of time-dependent averages of local dynamical variables grows *exponentially* with the in-degree of the node [38,39], even in the Markovian case. For a wide variety of systems of topical interest, which are characterized by fat-tailed degree distributions (see, e.g., Refs. [30–33,43,44] for a general background), this creates an additional and seemingly

*giuseppe.torrisi@kcl.ac.uk

impenetrable complexity barrier that effectively prevents any analytic study of such systems. This barrier is also present in the context of spreading processes; see Ref. [45], where the authors describe an optimization tool that reduces the exponential complexity to polynomial. However, their result relies on the microscopic irreversibility of the dynamics in the sense that a node that becomes active at a certain time will never revert to the inactive state. In contrast, the dynamics investigated in the present paper is not irreversible due to the presence of both positive and negative interaction terms. Hence, the dynamics cannot be solved via the methods used in Refs. [45,47], and a new method is therefore required.

The main aim of the present contribution is to provide a way to overcome this complexity barrier. We propose an algorithm inspired by dynamic programming that reduces the dependence of the computational complexity on the in-degrees of the system from exponential to quadratic. The approach can be shown to work for spin systems and Boolean networks, including those with p -node interactions, and is applicable whenever couplings are chosen randomly from a discrete set of equidistant values that may well be site-dependent. This would carry over to systems exhibiting continuous couplings that can be well approximated in this way, e.g., by discretization. We thus crack the complexity barrier in the degree for microscopically reversible dynamics. For the sake of definiteness, we explain it for the case of networks of Boolean linear threshold units, and cover generalizations to other systems in an extended paper. The interested reader can find the code to reproduce the results shown in the present paper at the link given in Ref. [46].

Our paper is organized as follows. In Sec. II we describe the model that is being studied. Section III is devoted to the dynamic programming approach devised to overcome the exponential complexity barrier. In Sec. IV we demonstrate our method on an application using a network with a fat-tailed degree distribution. Section V finally contains a conclusion and discussion. We also included an Appendix which contains details concerning numerics and simulations.

II. THE MODEL

We consider a Boolean network model defined on a finitely connected directed graph of N nodes labeled $i = 1, \dots, N$. For each edge (ij) of the graph, the edge weight $J_{ij} \in \mathbb{R}$ defines the strength of the interaction carried from node j to node i . Here we assume that the nonzero edge weights are binary, and randomly and independently drawn from the distribution $P(J_{ij}) = \eta\delta(J_{ij} - J) + (1 - \eta)\delta(J_{ij} + J)$ with $\eta \in (0, 1)$. We use $\partial_i^{\text{in}} = \{j : J_{ij} \neq 0\}$ to denote the set of predecessors of node i and use $k_i = |\partial_i^{\text{in}}|$ to designate the in-degree of node i . With each node we associate a Boolean state variable $n_i = \{0, 1\}$. Each node i receives a signal given by the local field

$$h_i(\mathbf{n}(t)) = \sum_j J_{ij} n_j(t) \quad (1)$$

and determines its activation state at the next time step following a noisy linear threshold dynamics of the form

$$n_i(t + 1) = \Theta[h_i(\mathbf{n}(t)) - \vartheta_i - z_i(t)]. \quad (2)$$

Here the ϑ_i are local thresholds, and the $z_i(t)$ are independent identically distributed random variables extracted from some cumulative distribution function $\text{Prob}[z \leq x] = \Phi(x)$. A common choice for the noise is “thermal” noise described by the distribution $\Phi(x) = \Phi_\beta(x) = [1 + \tanh(\beta x/2)]/2$ in which $\beta^{-1} = T$ is a measure of the noise strength. Any other noise model could be chosen instead. The local field $h_i(\mathbf{n})$ depends only on the states of those nodes which are predecessors of node i , that we denote with $\mathbf{n}_{\partial_i} = \{n_j, j \in \partial_i^{\text{in}}\}$. In the following, we will thus write the local field in i as $h_i(\mathbf{n}_{\partial_i})$.

The dynamics described by Eqs. (1) and (2) is stochastic. Let us denote by $P_i(t) = \text{Prob}(n_i(t) = 1)$ the activation probability of node i at time t . For finitely coordinated random graphs of the type considered here, the cavity method [48] can be used to analyze the dynamics of the system. It is exact on trees and known to become exact for finitely coordinated random graphs in the thermodynamic limit, as the typical length of any loops in such systems diverges logarithmically in system size N . Using the cavity approach on a fully asymmetric network, see Refs. [37–39], a closed form expression for the activation probability $P_i(t + 1)$ is obtained as

$$P_i(t + 1) = \langle \Phi_\beta(h_i(\mathbf{n}_{\partial_i}) - \vartheta_i) \rangle_{\mathbf{n}_{\partial_i}, t}, \quad (3)$$

in which the angled bracket $\langle \cdot \rangle_{\mathbf{n}_{\partial_i}, t}$ indicates an average over the states of the predecessors of node i , \mathbf{n}_{∂_i} evaluated using their joint node activation probabilities at time t :

$$\langle (\cdot) \rangle_{\mathbf{n}_{\partial_i}, t} = \sum_{\mathbf{n}_{\partial_i}} (\cdot) \prod_{j \in \partial_i^{\text{in}}} P_j(t)^{n_j} [1 - P_j(t)]^{1-n_j}. \quad (4)$$

Within the cavity approximation these joint probabilities can be taken to be independent. However, in order to evaluate the average, one has to sample from a state space containing 2^{k_i} elements for a node with in-degree k_i . This creates the aforementioned exponential complexity barrier for systems exhibiting fat-tailed in-degree distributions. To overcome it, we exploit the independence of joint node activation probabilities of a given node assumed within the cavity formalism to devise a recursive algorithm inspired by dynamic programming that is able to reduce the computational complexity of evaluating local averages from exponential to polynomial in the in-degree of nodes. We note that Eq. (3) is of the form

$$P_i(t + 1) = \sum_{\mathbf{n}_{\partial_i}} \phi_i(h_i(\mathbf{n})) \prod_{j \in \partial_i^{\text{in}}} \psi_j(n_j), \quad (5)$$

with $\phi_i(x) = \Phi_\beta(x - \vartheta_i)$ and $\psi_j(x) = P_j(t)^x [1 - P_j(t)]^{1-x}$. This structure of the cavity equation [Eq. (5)] is shared by belief-propagation involving sum-product algorithms, as broadly used in information theory and machine learning [49]. Our method tackles the exponential complexity barrier also in those contexts.

III. DYNAMIC PROGRAMMING

To introduce the idea behind the recursive dynamic programming approach [50,51] in the present context, let us consider a node i with in-degree k_i , and let us use $\{1, \dots, k_i\}$ to label the indices of the predecessors of node i . One way is to evaluate $P_i(t + 1)$ iteratively. That is, one first evaluates the

average over n_1 and obtains

$$P_i(t+1) = P_1(t) \left\langle \Phi_\beta \left(J_{i1} + \sum_{j=2}^{k_i} J_{ij} n_j - \vartheta_i \right) \right\rangle_{n_2, \dots, k_i, t} + [1 - P_1(t)] \left\langle \Phi_\beta \left(\sum_{j=2}^{k_i} J_{ij} n_j - \vartheta_i \right) \right\rangle_{n_2, \dots, k_i, t}, \quad (6)$$

in which $\langle \cdot \rangle_{n_2, \dots, k_i, t}$ denotes averages over n_2, \dots, n_{k_i} that remain to be performed. The original problem involving k_i predecessors has thus been reduced to *two* smaller problems, each requiring to perform averages over $k_i - 1$ variables. This procedure is repeated recursively on each of the smaller problems until all averages are evaluated, which overall requires to evaluate 2^{k_i} averages for a vertex of degree k_i .

To benefit from the heuristics of dynamic programming, we propose to reformulate the evaluation of the average Eq. (3) through the recursion just described, by *first solving a more general problem*, namely, the problem to evaluate the *family of averages* $\{f_i(\ell, \tilde{h}); 1 \leq \ell \leq k_i; \tilde{h} \in \mathbb{R}\}$ defined by

$$f_i(\ell, \tilde{h}) = \left\langle \Phi_\beta \left(\tilde{h} + \sum_{j=\ell}^{k_i} J_{ij} n_j - \vartheta_i \right) \right\rangle_{n_\ell, \dots, k_i, t}. \quad (7)$$

We refer to \tilde{h} as an auxiliary field. At any given ℓ , the set of $f_i(\ell, \tilde{h})$ of interest would in fact correspond to the set of averages representing the subproblems that remain to be evaluated after $\ell - 1$ levels of the above recursive procedure have been performed. With reference to Eq. (6), the $f_i(\ell, \tilde{h})$ are obtained using the *backward recursion*

$$f_i(\ell, \tilde{h}) = P_\ell(t) f_i(\ell+1, \tilde{h} + J_{i\ell}) + [1 - P_\ell(t)] f_i(\ell+1, \tilde{h}), \quad (8)$$

for $1 \leq \ell \leq k_i$, with the *terminal boundary condition*

$$f_i(k_i+1, \tilde{h}) = \Phi_\beta(\tilde{h} - \vartheta_i). \quad (9)$$

The original average $P_i(t+1)$ of Eq. (3) that we are ultimately interested in is, within this backward iteration scheme, obtained as

$$P_i(t+1) = f_i(1, 0). \quad (10)$$

While this recursive formulation is equivalent to direct evaluation of Eq. (3) and does not provide any speedup per se, the evaluation using dynamic programming implemented in the recursive procedure Eq. (8) leads to a dramatic reduction of computational complexity if the nonzero weights $J_{i\ell}$ of the incoming edges to site i are chosen in a discrete set of equidistant values, which may even be site-dependent, i.e., $J_{ij} \in \{-r_i J_i, \dots, -J_i, 0, J_i, \dots, s_i J_i\}$ for integer r_i, s_i . This will in particular apply to the binary interactions $J_{ij} \in \{0, \pm J\}$ that we are considering in this paper. In this case the recursive procedure (8) requires us to evaluate the $f_i(\ell, \tilde{h})$ only on a discrete grid of (ℓ, \tilde{h}) values as illustrated in Fig. 1. According to Eq. (8), $f_i(\ell, \tilde{h})$ only depends on $f_i(\ell+1, \tilde{h} + J_{i\ell})$ and $f_i(\ell+1, \tilde{h})$. The values of the auxiliary field \tilde{h} at level $\ell+1$ needed to evaluate $f_i(\ell, \tilde{h})$ are thus either the same as at level ℓ , or differ by $\pm J$ depending on the sign of the coupling $J_{i\ell}$. The arrows in Fig. 1 are horizontal in the first case, and

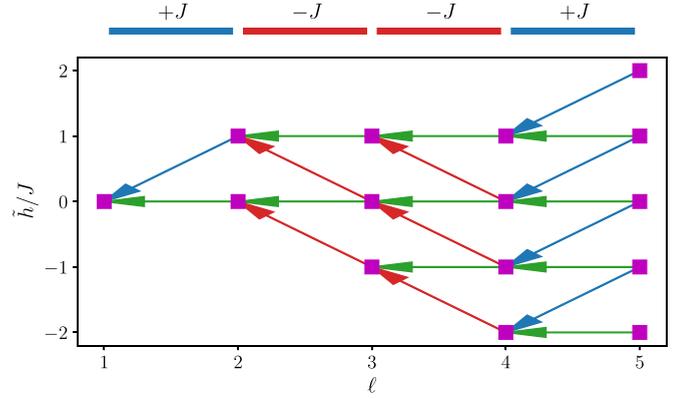


FIG. 1. Set of (ℓ, \tilde{h}) combinations for which $f(\ell, \tilde{h})$ is required in the execution of recursive evaluation Eq. (8) for a particular choice of couplings. In this example, $k_i = 4$, the vector of couplings is $(J_{i,1}, J_{i,2}, J_{i,3}, J_{i,4}) = (+J, -J, -J, +J)$ as shown on the top. Arrows join configurations (ℓ, \tilde{h}) explored through the recursive relation Eq. (8). Green arrow indicates the recursion call from $f_i(\ell+1, \tilde{h})$ to $f_i(\ell, \tilde{h})$. Blue and red arrows indicates the call from $f_i(\ell+1, \tilde{h} + J_{i\ell})$ to $f_i(\ell, \tilde{h})$ through a positive or negative coupling, respectively

diagonally up or down in the second. The total number of values $f_i(\ell, \tilde{h})$ that need to be evaluated for a node with k_i predecessors corresponds to the number of markers in Fig. 1, and is given by

$$\sum_{\ell=1}^{k_i+1} \ell = (k_i+1)(k_i+2)/2. \quad (11)$$

The computational and memory requirement for the evaluation of f_i is thus seen to scale as k_i^2 , for a node with in-degree k_i , rather than 2^{k_i} , as it would in a naive evaluation. This entails that the complexity of our algorithm is $O(\sum_i k_i^2)$. If, on the other hand, the couplings J_{ij} were sampled from a continuous distribution, then the function calls would generally be unique, and no speedup would be obtained. In this latter case, the complexity remains $O(\sum_i 2^{k_i})$, making this problem strongly NP-complete [52].

In the following, we take full advantage of the computational speedup provided by the algorithm discussed above to investigate heterogeneities in Boolean networks with dynamics [Eq. (2)], a fat-tailed degree distribution, and thermal noise.

IV. APPLICATIONS

Literature on dynamical properties of Boolean or spin systems has traditionally focused on averages of extensive quantities in the thermodynamic limit [27]. However, the cavity method has been recently used to assess the heterogeneous behavior of individual nodes [41,53,54]. These arise from variations in node environment, a prominent feature of networks with fat-tailed degree distributions, and they have an important role in network processes, as already known in the context of resilience of collective phenomena in networked systems [55,56]. However, the aforementioned exponential complexity in the in-degrees has so far prevented a detailed

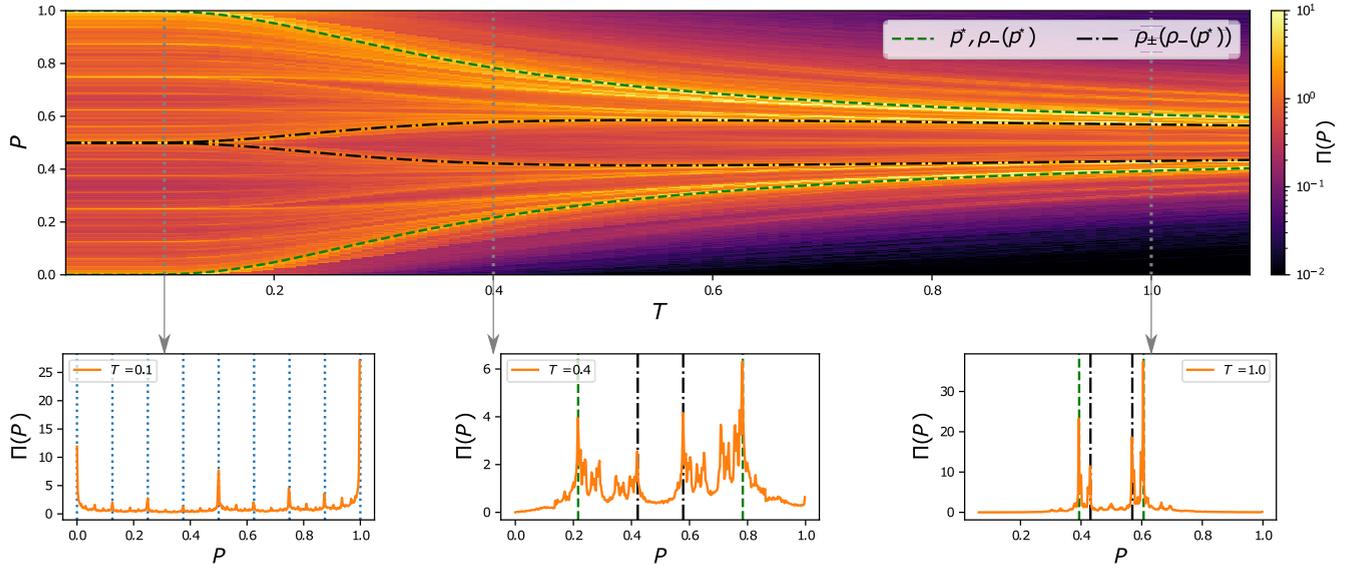


FIG. 2. (top) Heat map of the pdf $\Pi(P)$ of heterogeneous gene activation probabilities at different values of noise parameter T . The dashed lines show the values p^* and $\rho_-(p^*)$, while the dot-dashed lines show the values $\rho_\pm(\rho_-(p^*))$, as a function of the noise level T . The vertical dotted lines indicate the values of T at which the histograms of $\Pi(P)$, in the lower panels, are computed. (bottom) Histograms of $\Pi(p)$ at different noise levels, i.e., $T = 0.1$ (left), $T = 0.4$ (middle), $T = 1$ (right), corresponding to the regimes $T \ll J$, $T \approx J$, and $T \gg J$, respectively. The vertical dotted lines on the left panel are located at $(0, 1/8, 1/4, 3/8, \dots, 1)$. The interactions bias and strength are $\eta = 0.621$ and $J \approx 0.72$, respectively. The network size is $N = 200\,000$. The largest degree of the network is $k = 800$.

study of node heterogeneities, precisely in those systems where they play a bigger role.

Here we show that such an analysis is feasible using our method. As an example, we consider a synthetic network of size $N = 200\,000$ and degree distribution $p(k) \sim k^{-\gamma}$, for $k \gg 1$ with $\gamma = 2.81$,¹ chosen to match that of a gene regulatory network (GRN) prototype [57]. Our analysis uncovers a highly nontrivial distribution of node activation probabilities. By computing the stationary solution of Eq. (3), we obtain the node activation probabilities $\{P_i\}_{i=1}^N$ in stationary conditions. Their distribution $\Pi(P) = N^{-1} \sum_i \delta(P_i - P)$ is shown in Fig. 2 for different values of the noise strength T . For convenience, we fix the threshold at $\vartheta = 0$. Thus, the probability p_0 that a node spontaneously activates when the local field vanishes, is $p_0 = 1/2$. At low noise level, $T \ll J$, nodes with frozen neighbors and positive (negative) local field will be active (inactive) with probability very close to one. This leads to large clusters of nodes locked either in the active or in the inactive state. Correspondingly, the distribution $\Pi(P)$ shows two main peaks at $P = 1$ and $P = 0$, the former higher than the latter due to the presence of a bias $\eta = 0.621$ towards positive interactions (Fig. 2, bottom-left panel). The other peaks correspond to nodes which fluctuate between the active and inactive state, and thus they arise either from nodes with frozen neighbors and zero field or from nodes with fluctuating neighbors. In particular, the peak at $P = 1/2$ arises from nodes that activate spontaneously in a field frozen at zero, while the other peaks arise from nodes with fluctuating neighbors. For example, a node i with unit in-degree and neighbor j that is active with probability p_0 , will be itself active with

probability $p_0(1 - p_0)$, if the coupling J_{ij} is negative, and with probability $p_0 + p_0(1 - p_0)$ if J_{ij} is positive, as the probability of activation in negative (positive) field is zero (one), respectively.

When the noise level is increased to $T \approx J$ there is a finite probability that a node activates in a negative field. Let us denote as $\rho_\pm(p) = \frac{1}{2}(1 \pm p \tanh \frac{\beta J}{2})$ the activation probability of a node with unit in-degree linked to a neighbor that is active with probability p , through the coupling $\pm J$, respectively. Because of the fat-tailed degree distribution and absence of correlations between in-degrees and out-degrees, a significant proportion of nodes are found in unidirectional chains (which may and often will be embedded in branching trees). In any such chain, a node $i + 1$, successor of a node i that is active with probability p_i , will be itself active with probability $p_{i+1} = \rho_\pm(p_i)$. Hence, node activation probabilities are determined by a discrete stochastic map with two branches, and two stable fixed points $p^* = \rho_+(p^*)$ and $p^\circ = \rho_-(p^\circ)$. Along the nodes of a chain connected through positive (negative) interactions the activation will converge to p^* (p°). Due to the bias η , chains of positive interactions are a common occurrence, and yield the rightmost peak of $\Pi(p)$, $P = p^*$ (see bottom mid and right panels). The leftmost peak arises from nodes linked through a negative coupling to a single node with activation probability p^* , which will activate with probability $\rho_-(p^*)$. The other peaks can be interpreted in terms of further compositions of the functions ρ_\pm . This in particular also explain the peaks found in the zero-temperature limit at sums of higher-order powers of $1/2$. While the above line of reasoning explains locations of prominent peaks in $\Pi(P)$, other features of the distribution cannot be explained in terms of chain-like structures alone. This includes, in particular, the height of these peaks and any feature of $\Pi(P)$ outside the

¹The distribution used is $p(k) = [k^{-(\gamma-1)} - (k+1)^{-(\gamma-1)}]$.

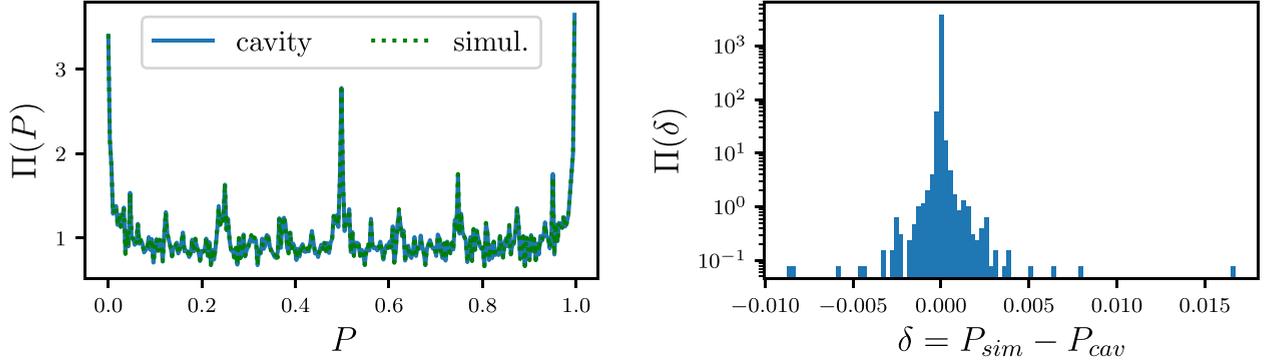


FIG. 3. (left) Histogram of the node activation probability as computed from theoretical calculation (blue solid line) and from direct simulation of the microscopic dynamics (dotted green line). (right) Histogram of the difference $\delta_i = P_{i,\text{sim}} - P_{i,\text{cav}}$, evaluated for each node i of the network. Parameters: $N = 50\,000$, $T = 0.1$, $\gamma = 2.81$, $\eta = 0.5$.

interval $[\rho_-(1), \rho_+(1)]$, as well as properties of the smooth contributions to $\Pi(P)$ observed at finite temperature. Simulation results are provided in the Appendix for a network with $N = 50\,000$ and show excellent agreement with predictions from the cavity equations solved via dynamic programming.

V. CONCLUSIONS

In this paper, we have presented a dynamic programming algorithm that dramatically reduces the computational complexity of the dynamic cavity algorithm for Boolean systems with fully asymmetric interactions. It can be used whenever couplings are randomly sampled from (or can be approximated in terms of) a discrete set of equidistant values. For systems of this type, the dependence of the computational complexity on the local in-degree of nodes that is required for dynamic updates is reduced from exponential to quadratic. Furthermore, evaluations can be carried out *in parallel* for each node. We illustrate the power of the algorithm by studying a Boolean linear threshold model with binary $\pm J$ interactions. However, it can be shown that the present approach can be generalized to nonlinear threshold models involving higher-order interactions and to bipartite systems, which are of particular interest in the context of gene regulatory networks. Thanks to the reduction of the computational complexity, we can fully uncover heterogeneities in node activation patterns in networks with fat-tailed in-degree distribution, that were previously inaccessible. We discuss salient features of the distribution of node activation probabilities and we show that they can in part be rationalized in terms of discrete stochastic maps resulting from the underlying network structure and dynamics. We will report on further aspects of the present model and some of its extensions in a future extended paper.

ACKNOWLEDGMENTS

G.T. is supported by the EPSRC Centre for Doctoral Training in Cross-Disciplinary Approaches to Non-Equilibrium Systems (CANES EP/L015854/1). We thank Professor Franca Fraternali's group for providing access to their computational facilities.

APPENDIX: SIMULATIONS

1. Parameter definition

In the present paper, we consider a model of a synthetic network in the configuration model class, i.e., we study ensembles of maximally random directed graphs \mathbf{J} with constrained in-degree and out-degree sequences. The sequence of in-degree and out-degree are chosen to be the same for simplicity, but the values associated with nodes are shuffled, therefore there is no correlation between the in- and out-degree. The sequence of in-degree of nodes k_i^{in} is drawn from the distribution $\rho^{\text{in}}(k) = N^{-1} \sum_i \delta_{k,k_i}$. We use $\langle k \rangle$ to denote the mean in-degree. The parameters of our synthetic model are chosen in such a way to match properties of gene regulatory networks, specifically the network described in the dataset TRUST [57]. The in-degree node distribution is well described by $P_\gamma(k) = k^{-\gamma} - (k+1)^{-\gamma}$ with power-law tail $P_\gamma(k) \sim \gamma k^{-\gamma-1}$ and $\gamma = 2.81$ [57]. For each nonzero entry of the adjacency matrix A_{ij} , the corresponding interaction term J_{ij} can be positive or negative. We sample the sign of the interaction as an independent random variable with bias $\eta = 0.621$ reflecting the bias of regulatory interactions (excitatory vs inhibitory) in TRUST [57]. The absolute value of the interaction term is $J = 1/(\langle k^{\text{in}} \rangle)^{1/2}$. The size of the network used in the simulation of Fig. 2 is $N = 200\,000$.

2. Comparison between theory and simulation

We compute iteratively the node activation probability at time $t+1$ given the knowledge of node activation probabilities at time t ,

$$P_i(t+1) = \frac{1}{2} \left(1 + \left\langle \tanh \left[\frac{\beta}{2} (h_i(\mathbf{n}_{\partial_i}) - \vartheta_i) \right] \right\rangle_{\mathbf{n}_{\partial_i}, t} \right). \quad (\text{A1})$$

We inspect the stationary solution of the node activation probability $P_i = \lim_{t \rightarrow \infty} P_i(t) \forall i \in \{1, \dots, N\}$. The stationary activation probabilities are obtained by running the iterative procedure of Eq. (A1) until convergence. The convergence is controlled by measuring the error between consecutive steps $\epsilon_{t+1} = \max_i |P_i(t+1) - P_i(t)|$. Let us call $P_{i,\text{cav}} = P_i(t)$ corresponding to the smallest t satisfying the exit condition $\epsilon_t < 10^{-4}$. We test cavity predictions against a simulation of the microscopic dynamics Eq. (2). We evaluate the frequency of

node activation from dynamical trajectories taking the sample average of node activation

$$P_{i,\text{sim}} = \frac{1}{t_m} \sum_{t=t_0}^{t_0+t_m} n_i(t), \quad (\text{A2})$$

for $t_m \gg 1$, and t_0 a sufficiently long time to allow the dynamics to relax to stationarity. In Fig. 3 we show the distribution of the node activation probability as computed by the cavity

method and by estimation from simulation. These are found to be in perfect agreement, confirming the validity of the cavity method to investigate the stationary state, see Fig. 3. Moreover, in order to reach the resolution imposed by cavity $\epsilon_t < 10^{-4}$, we need to simulate long trajectories in Eq. (A2) with $t_m \approx 10^8$, which makes the evaluation from simulation more computationally expensive. For the same network realization and parameter choice, the cavity implementation took 50 s, while simulations took more than 11 days on the same machine.

-
- [1] S. F. Edwards and P. W. Anderson, *J. Phys. F: Met. Phys.* **5**, 965 (1975).
- [2] D. Sherrington and S. Kirkpatrick, *Phys. Rev. Lett.* **35**, 1792 (1975).
- [3] L. Viana and A. J. Bray, *J. Phys. C: Solid State Phys.* **18**, 3037 (1985).
- [4] S. A. Kauffman, *J. Theor. Biol.* **22**, 437 (1969).
- [5] B. Derrida and Y. Pomeau, *Europhys. Lett.* **1**, 45 (1986).
- [6] G. Parisi, *Proc. Natl. Acad. Sci. USA* **87**, 429 (1990).
- [7] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, and F. Moauro, *Phys. Rev. Lett.* **109**, 268101 (2012).
- [8] E. Agliari, A. Annibale, A. Barra, A. C. C. Coolen, and D. Tantari, *J. Phys. A: Math. Theor.* **46**, 415003 (2013).
- [9] J. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982), reprinted in Ref. [58].
- [10] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (CRC Press, Boca Raton, 2018).
- [11] D. Amit, H. Gutfreund, and H. Sompolinsky, *Ann. Phys. (NY)* **173**, 30 (1987).
- [12] P. Sollich, D. Tantari, A. Annibale, and A. Barra, *Phys. Rev. Lett.* **113**, 238106 (2014).
- [13] D. Challet and Y. C. Zhang, *Phys. A (Amsterdam, Neth.)* **246**, 407 (1997).
- [14] A. C. C. Coolen, *The Mathematical Theory of Minority Games—Statistical Mechanics of Interacting Agents* (Oxford University Press, Oxford, 2005).
- [15] G. Iori, *Int. J. Mod. Phys. C* **10**, 1149 (1999).
- [16] S. Bornholdt, *Int. J. Mod. Phys. C* **12**, 667 (2001).
- [17] K. Anand and R. Kühn, *Phys. Rev. E* **75**, 016111 (2007).
- [18] J. P. L. Hatchett and R. Kühn, *J. Phys. A: Math. Gen.* **39**, 2231 (2006).
- [19] M. Weigt and A. K. Hartmann, *Phys. Rev. Lett.* **84**, 6118 (2000).
- [20] S. Cocco and R. Monasson, *Eur. Phys. J. B* **22**, 505 (2001).
- [21] O. C. Martin, R. Monasson, and R. Zecchina, *Theor. Comput. Sci.* **265**, 3 (2001).
- [22] S. Franz, M. Leone, F. Ricci-Tersenghi, and R. Zecchina, *Phys. Rev. Lett.* **87**, 127209 (2001).
- [23] M. Mézard, G. Parisi, and R. Zecchina, *Science* **297**, 812 (2002).
- [24] C. De Dominicis, *Phys. Rev. B* **18**, 4913 (1978).
- [25] H. J. Sommers, *Phys. Rev. Lett.* **58**, 1268 (1987).
- [26] E. Gardner, B. Derrida, and P. Mottishaw, *J. Phys. (Paris)* **48**, 741 (1987).
- [27] A. Crisanti and H. Sompolinsky, *Phys. Rev. A* **37**, 4865 (1988).
- [28] R. Kree and A. Zippelius, in *Models of Neural Networks*, edited by E. Domany, J. van Hemmen, and K. Schulten (Springer, Berlin, 1991), pp. 193–212.
- [29] B. Derrida, E. Gardner, and A. Zippelius, *Europhys. Lett.* **4**, 167 (1987).
- [30] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: from Biological Networks to the Internet and WWW* (Oxford University Press, Oxford, 2003).
- [31] M. E. J. Newman, *Networks: An Introduction*, 2nd ed. (Oxford Univ. Press, Oxford, 2018).
- [32] E. Estrada, *The Structure of Complex Networks: Theory and Applications* (Oxford University Press, Oxford, 2011).
- [33] V. Latora, V. Nicosia, and G. Russo, *Complex Networks: Principles, Methods and Applications* (Cambridge University Press, Cambridge, 2017).
- [34] M. Mézard and J. Sakellariou, *J. Stat. Mech.* (2011) L07001.
- [35] Y. Roudi and J. Hertz, *J. Stat. Mech.: Theory Exp.* (2011) P03031.
- [36] P. Zhang, *J. Stat. Phys.* **148**, 502 (2012).
- [37] E. Aurell and H. Mahmoudi, *Phys. Rev. E* **85**, 031119 (2012).
- [38] K. Mimura and A. C. C. Coolen, *J. Phys. A: Math. Theor.* **42**, 415001 (2009).
- [39] I. Neri and D. Bollé, *J. Stat. Mech.: Theory Exp.* (2009) P08009.
- [40] P. Paga and R. Kühn, *J. Stat. Mech.: Theory Exp.* (2015) P03008.
- [41] A. Y. Lokhov, M. Mézard, H. Ohta, and L. Zdeborová, *Phys. Rev. E* **90**, 012801 (2014).
- [42] B. Derrida, *J. Phys. A: Math. Gen.* **20**, L721 (1987).
- [43] J. Leskovec and A. Krevl, SNAP Datasets: Stanford Large Network Dataset Collection, <http://snap.stanford.edu/data> (2014).
- [44] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- [45] F. Altaelli, A. Braunstein, L. Dall’Asta, and R. Zecchina, *J. Stat. Mech.: Theory Exp.* (2013) P09011.
- [46] See, <https://zenodo.org/badge/latestdoi/355104547>.
- [47] A. Y. Lokhov, M. Mézard, and L. Zdeborová, *Phys. Rev. E* **91**, 012811 (2015).
- [48] M. Mézard and G. Parisi, *Eur. Phys. J. B* **20**, 217 (2001).
- [49] M. Mézard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, 2009).
- [50] R. Bellman, *Science* **153**, 34 (1966).
- [51] A. D. Scott and G. B. Sorkin, *ACM Trans. Algorithms* **5**, 1 (2009).
- [52] D. Wojtczak, in *International Computer Science Symposium in Russia* (Springer, 2018), pp. 308–320.
- [53] T. Kobayashi and T. Onaga, [arXiv:2103.09417](https://arxiv.org/abs/2103.09417).
- [54] R. Kühn and T. Rogers, *Europhys. Lett.* **118**, 68003 (2017).

- [55] R. Albert, H. Jeong, and A. L. Barabási, *Nature (London)* **406**, 378 (2000).
- [56] A. Annibale, A. C. C. Coolen, and G. Bianconi, *J. Phys. A: Math. Theor.* **43**, 395001 (2010).
- [57] H. Han, J. Cho, S. Lee, A. Yun, H. Kim *et al.*, *Nucleic Acids Res.* **46**, D380 (2018).
- [58] *Neurocomputing: Foundations of Research*, edited by J. A. Anderson and E. Rosenfeld (MIT Press, Cambridge, 1988).