

Large deviations of cascade processes on graphsF. Altarelli,^{1,2} A. Braunstein,^{1,2,3} L. Dall'Asta,^{1,2} and R. Zecchina^{1,2,3}¹*Department of Applied Science and Technology, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy*²*Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri, Italy*³*Human Genetics Foundation, Via Nizza 52, 10126 Torino, Italy*

(Received 11 January 2013; revised manuscript received 22 March 2013; published 11 June 2013)

Simple models of irreversible dynamical processes such as bootstrap percolation have been successfully applied to describe cascade processes in a large variety of different contexts. However, the problem of analyzing nontypical trajectories, which can be crucial for the understanding of out-of-equilibrium phenomena, is still considered to be intractable in most cases. Here we introduce an efficient method to find and analyze optimized trajectories of cascade processes. We show that for a wide class of irreversible dynamical rules, this problem can be solved efficiently on large-scale systems.

DOI: [10.1103/PhysRevE.87.062115](https://doi.org/10.1103/PhysRevE.87.062115)

PACS number(s): 64.60.aq, 05.20.-y, 75.10.Nr, 89.20.Ff

I. INTRODUCTION

Large-scale cascading processes observed in physical and biological systems can be described and understood by means of stylized models of propagation on lattices or graphs. Over the last 40 years, these models have found application to problems arising in a number of different contexts, ranging from competing interactions in dilute magnetic systems [1,2], jamming transitions in glass formers and granular media [3], epidemic spreading [4], and activation cascades in cortical [5] and other biological networks [6] to the spread of information and innovations in social models [7–11] and propagation of liquidity shocks in financial interbank lending networks [12,13]. In all these problems the basic units composing the systems are discrete and undergo irreversible transitions from an “inactive” state to an “active” one depending on the state of their neighbors. Following recent works in the computer science community [14], we refer to this class of dynamical processes as models of *progressive dynamics*.

Theoretical works across several disciplines have focused mostly on the mechanisms responsible for the emergence of some collective behavior, explaining under what conditions on the dynamical rule and the graph (or lattice) structure large-scale propagations can be observed as an outcome of typical realizations of the process, i.e., when starting from random initial conditions. Because of the intrinsic nonlinearity of the dynamics, a *critical (or tipping) point* usually separates a region of parameters in which the dynamics typically occurs only locally from a region of large-scale propagations. This is exactly what occurs in celebrated models of statistical physics, such as bootstrap and k -core percolation [1,15–17], and zero-temperature Ising-like models [2,18], whose critical properties have been extensively studied for several classes of networks, such as d -dimensional lattices and random graphs. Similarly, tipping points are observed in simple models of binary decisions with externalities [7,9,10], providing an explanation for the occurrence of abrupt changes in the collective behavior of socioeconomic systems. These analyses are usually performed either by simulating the evolution of the dynamical rule and averaging over many (randomly drawn) initial conditions, or by resorting to approximate descriptions of the dynamics in the form of differential equations based on mean-field and pair-approximation techniques [19,20].

While the average dynamical properties of these models starting from random initial conditions are rather well understood on general networks, their *large deviations*, describing macroscopic behaviors that deviate considerably from the average ones, are still a largely unexplored domain of research that goes beyond the means of current methods of analysis. Large deviations are of interest for at least two different reasons: because they correspond to *desired* final states (e.g., extraordinarily large propagations of a small set of initially active nodes) or because they correspond to an *observed* final state of an unknown initial one. The application of large-deviation analyses to the nonequilibrium dynamics of interacting particle systems is a subject of intense study in statistical physics [21]. Models of progressive dynamics offer a sufficiently simple, though nontrivial, setting to extend these studies to systems with complex interaction patterns such as random graphs and complex networks.

In this paper we consider the problem of characterizing dynamical trajectories with interesting nontypical statistical properties in deterministic progressive models. In this class of models, the choice of the initial conditions completely determines the dynamical trajectory of the system. However, because of the nonlinearity of the local update rule, even slight differences in the initial conditions can result in completely different collective behaviors. By averaging over all possible initial conditions or drawing them at random, the macroscopic quantities of interest are dominated by their typical behavior, which can be very different from the observed one when a particular choice of the initial conditions is made. On the contrary, we will provide here a method to estimate the statistical properties of rare, but relevant, dynamical trajectories and find the initial conditions that give rise to cascading processes with some desired properties. Understanding under what conditions a rare large-scale propagation may occur and estimating the probability and other statistical properties of such an outcome have remarkable practical applications in a variety of fields beyond physics, such as the study of the spread of information in social networks, the problem of targeted silencing in gene regulatory networks, or the development of systemic risk measures and control techniques in financial and infrastructure systems.

Our approach is based on a static representation of the dynamical rules of deterministic progressive models that

allows one to recast the study of their large deviations into the evaluation of a partition function. In networked systems this can be done by means of the cavity method and derived message-passing algorithms. Even though the method we propose is very general and can be applied to any deterministic progressive dynamics with discrete degrees of freedom, in the following we shall consider explicitly the linear threshold model (LTM), a prototypical model for the analysis of cascade processes on networks.

II. TYPICAL BEHAVIOR OF PROGRESSIVE DYNAMICS ON GRAPHS

In this section we will present the cavity formalism for the analysis of the typical behavior of trajectories in progressive dynamics, from which we will recover previously known results; the formal connection with the main result on large deviations will be discussed in Appendix A. We consider a generic deterministic progressive dynamics in discrete time defined over a graph $G = (V, E)$ and involving discrete state variables $\mathbf{x} = \{x_i, i \in V\}$. For simplicity we shall assume that there are only two states, $x_i = 0$ called *inactive* and $x_i = 1$ called *active*, the generalization to more states being straightforward. A vertex which is active at time t will remain active at all subsequent times, while a vertex which is inactive at time t can get activated at time $t + 1$ if some condition, depending on the state of its neighbors in G at time t and expressing the dynamical rule considered, is satisfied. For instance, in the linear threshold model [8,11,22], the dynamics is defined by the rule

$$x_i^{t+1} = \begin{cases} 1 & \text{if } x_i^t = 1 \text{ or } \sum_{j \in \partial i} w_{ji} x_j^t \geq \theta_i, \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where $w_{ij} \in \mathbb{R}^+$ are weights associated with directed edges $(i, j) \in E$, $\theta_i \in \mathbb{R}^+$ are thresholds associated with $i \in V$, and ∂i denotes the set of neighbors of i in G . The model is strictly related to the zero-temperature limit of the random-field Ising model [2,18] and to the bootstrap percolation process [1,15,16]. The active nodes at time $t = 0$ are called the *seeds* of the progressive dynamics.

A. The direct dynamical problem

A peculiar property of a progressive process defined on a graph G is that any realization of the process is in one-to-one correspondence with a directed acyclic subgraph of G . Let us consider a set of seeds and, for each time step t , draw a directed edge connecting the nodes activated at time t to their neighbors activating at later times. The final result is a directed acyclic graph (DAG) as shown in Fig. 1 for an illustrative case. When the initial conditions are drawn from a distribution, the probability that a node i is active is given by the probability that the node is in the set of seeds plus the probability that it is not a seed but gets activated during the dynamics. The latter is the probability that i is reached by directed paths from the seeds in the ensemble of DAGs associated with the initial distribution. When the underlying graph is a tree and the initial conditions are drawn from a product measure, i.e., with probability $\Pr\{\mathbf{x}^0 = \mathbf{x}\} = \prod_i p_i^{x_i} (1 - p_i)^{1-x_i}$, the probability $\rho_i(t)$ that a node i is active at time t can be computed exactly

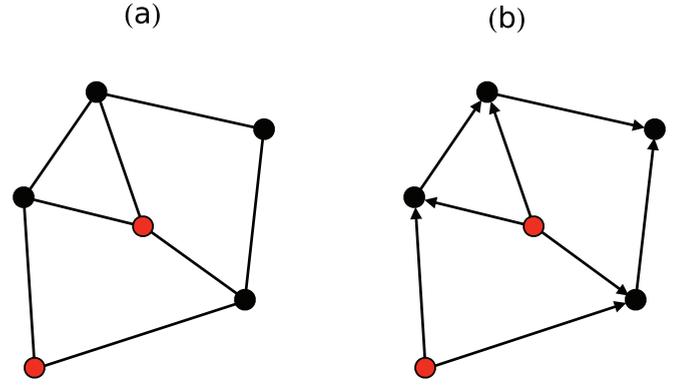


FIG. 1. (Color online) An example of the relation between the progressive models and directed acyclic graphs (DAGs). A graph of six vertices undergoes a LTM with two seeds (vertices marked in red). The weights on all edges are equal to 1 and the threshold is equal to 2 for every node. The result of the dynamics is the DAG on the right. The directions of the edges in the DAG represent the causal relations behind node activations.

by a simple recursive approach. For instance, in the LTM, it is given by

$$\begin{aligned} \rho_i^t &= p_i + (1 - p_i) \langle \Pr\{x_i^t = 1 | x_i^0 = 0\} \rangle \\ &= p_i + (1 - p_i) \sum_{\substack{I \subseteq \partial i \\ \sum_{\ell \in I} w_{\ell i} \geq \theta_i}} \prod_{\ell \in I} \chi_{\ell i}^t \prod_{k \in \partial i \setminus I} (1 - \chi_{ki}^t), \end{aligned} \quad (2)$$

with

$$\chi_{ji}^{t+1} = p_j + (1 - p_j) \sum_{\substack{I \subseteq \partial j \setminus i \\ \sum_{\ell \in I} w_{\ell j} \geq \theta_j}} \prod_{\ell \in I} \chi_{\ell j}^t \prod_{k \in \partial j \setminus (i \cup I)} (1 - \chi_{kj}^t). \quad (3)$$

The quantity χ_{ji}^t is a cavity marginal expressing the probability that node j is active at time t in the absence of node i . Due to the nature of the process, assuming that node i is absent is equivalent to assuming that it is inactive; therefore the causal structure implied by Eqs. (2) and (3) is exact on the tree. When the underlying graph has loops, the recursive equation is not exact. In this case, the DAG corresponding to a single dynamical evolution can present multiple directed paths connecting node i to the same seed (see, for instance, Fig. 1). In this case the decorrelation assumption behind (2) and (3) is not correct because two paths reaching i from different neighbors could originate in the same seed and therefore they might be correlated. This argument shows that the activation probability of a node estimated by (2) and (3) is always an upper bound of the real one. Despite this limitations, the local treelike approximation gives approximately correct results on sufficiently sparse graphs.

This cavitylike approach to study the time-dependent dynamics of progressive processes on graphs was recently put forward, in slightly different contexts, by several authors. In particular, Ohta and Sasa have used a very similar approach to study the zero-temperature dynamics of the random-field Ising model on the Bethe lattice [18], while Karrer and Newman [23] and Noël *et al.* [24] developed a similar method for the susceptible-infected model of epidemic spreading.

B. Relation to the bootstrap percolation problem

In bootstrap percolation [1], the sites of an empty lattice are first randomly occupied with probability q , and then all occupied sites with less than a given number m of occupied neighbors are successively removed until a stable configuration is reached. As in standard percolation, in the limit of infinitely large graphs, the average properties of the model are characterized by the existence of a critical density q_c of initially occupied sites below which the stable configuration of the system is the empty one. Taking $p = 1 - q$ and interpreting empty (occupied) sites as active (inactive) nodes, the bootstrap percolation process can be mapped exactly on a LTM with uniform weights $w_{ij} = 1, \forall (i, j) \in E$, and thresholds $\theta_i = k_i - m, \forall i \in V$, where k_i is the degree of vertex i . We consider the simple case of a regular random graph with degree K and uniform thresholds equal to θ for all nodes. Because of the homogeneity of the graph, one can assume all marginals χ_{ij}^t to be identical, obtaining for $\chi = \lim_{t \rightarrow \infty} \chi^t$ the

self-consistent equation

$$\chi = p + (1 - p) \sum_{n=\lceil \theta \rceil}^{K-1} \binom{K-1}{n} \chi^n (1 - \chi)^{K-1-n}. \quad (4)$$

The final density of active nodes when each node can be a seed with the same probability p is given by

$$\rho = p + (1 - p) \sum_{n=\lceil \theta \rceil}^K \binom{K}{n} \chi^n (1 - \chi)^{K-n}. \quad (5)$$

As expected, Eqs. (4) and (5) can be mapped exactly on the standard self-consistent equations derived for the bootstrap and k -core percolation processes on regular random graphs [1,15,16]. Figure 2 displays the behavior of the final density ρ_T of active nodes as a function of the density ρ_0 of seeds in two interesting cases with $\theta = 2$ and $K = 3$ [Fig. 2(a)] and $K = 4$ [Fig. 2(b)]. If we stop the dynamics at a finite number of time steps T , e.g., $T = 5$, the activation process is smooth in both cases. On increasing T , the curves for $K = 4$ become much steeper than for $K = 3$. In the limit $T \rightarrow \infty$, corresponding to the static bootstraplike calculation, the two cases have completely different behavior: the activation transition is continuous for $K = 3$ and discontinuous for $K = 4$ (see the insets of Fig. 2). This is exactly the well-known critical phenomenon observed in bootstrap percolation models on Bethe lattices [1,16]. In general, for $\theta = 1, K - 1$ the whole graph is activated ($\rho_\infty = 1$) continuously at a finite density ρ_0^c , whereas for $1 < \theta < K - 1$ the activation is abrupt. The generalization of the calculations to other values of the weights and thresholds as well as to nonregular uncorrelated random graphs is straightforward.

III. LARGE DEVIATIONS OF PROGRESSIVE DYNAMICS ON GRAPHS

In this section we consider the inverse problem of dynamical evolution, i.e., the problem of finding the initial conditions that give rise to a desired final state. If we focus on the behavior of some macroscopic observable, such as the number of activated nodes in the final state as a function of the number of seeds, the inverse problem corresponds to investigating the large-deviation properties of the dynamics.

A. The inverse dynamical problem

Because of irreversibility, the trajectory $\mathbf{x}^T = \{\mathbf{x}^0, \dots, \mathbf{x}^T\}$ representing the time evolution of the system can be fully parametrized by a configuration $\mathbf{t} = \{t_1, \dots, t_N\}$, where $t_i \in \mathcal{T} = \{0, 1, 2, \dots, T, \infty\}$ is the activation time of node i . We conventionally set $t_i = \infty$ if i does not activate within an arbitrarily defined stopping time T . In general, if the number of possible single-node trajectories is n , we can use a discrete variable taking n states. Given a set of seeds $S = \{i : t_i = 0\}$, the solution of the dynamics is fully determined for $i \notin S$ by a set of relations among the activation times of neighboring nodes, which we denote by $t_i = \phi_i(\{t_j\})$ with $j \in \partial i$. In terms of activation times, the dynamical rule for the LTM translates into $t_i = \phi_i(\{t_j\})$ with

$$\phi_i(\{t_j\}) = \min \left\{ t \in \mathcal{T} : \sum_{j \in \partial i} w_{ji} \mathbb{1}[t_j < t] \geq \theta_i \right\}. \quad (6)$$

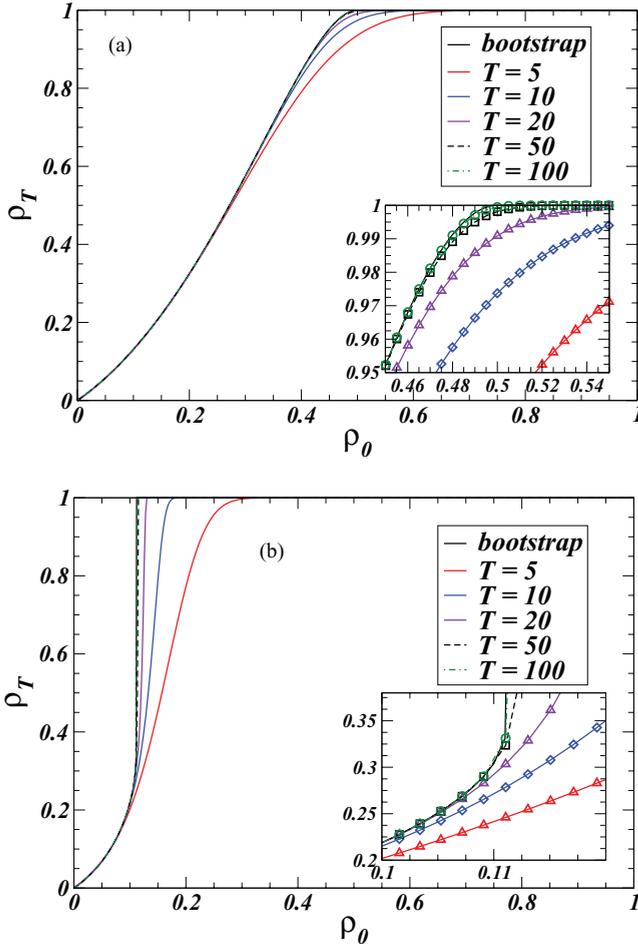


FIG. 2. (Color online) Plot of ρ_T vs ρ_0 obtained by solving the time-dependent equations (2) and (3) and the bootstrap percolation equations (4) and (5) on regular random graphs of degree $K = 3$ (a) and $K = 4$ (b), for threshold $\theta = 2$. On increasing the duration $T = 5, 10, 20, 50, 100$ of the dynamics, the curves obtained using (4) and (5) get closer to the solution (from right to left) of the corresponding bootstrap percolation equations. The insets highlight the regions close to the activation transitions.

Admissible trajectories in this model correspond to vectors \mathbf{t} such that $\Psi_i = \mathbb{1}[t_i = 0] + \mathbb{1}[t_i = \phi_i(\{t_j\})]$ equals 1 for every i .

In this static representation, one can introduce an energetic term $\mathcal{E}(\mathbf{t})$ that gives different probabilistic weights to different trajectories. The path probability associated with a configuration of activation times is

$$P(\mathbf{t}) = \frac{1}{Z} e^{-\beta \mathcal{E}(\mathbf{t})} \prod_{i \in V} \Psi_i(t_i, \{t_j\}_{j \in \partial i}) \quad (7)$$

with $Z = \sum_{\mathbf{t}} e^{-\beta \mathcal{E}(\mathbf{t})} \prod_{i \in V} \Psi_i(t_i, \{t_j\}_{j \in \partial i})$. The large-deviation properties of the dynamical process can be studied by evaluating the static partition function for the dynamic trajectories with an opportunely defined energetic term. Notice that the value chosen for T will affect the “speed” of the propagation: a lower value of T will restrict the optimization to “faster” trajectories, at the (possible) expense of the value of the energy.

The most general form of energy function we consider is $\mathcal{E}(\mathbf{t}) = \sum_i \mathcal{E}_i(t_i)$ where $\mathcal{E}_i(t_i)$ is the “cost” (if positive, or “revenue” if negative) incurred by activating vertex i at time t_i . In the following, we set $\mathcal{E}_i(t_i) = \mu_i \mathbb{1}[t_i = 0] - \epsilon_i \mathbb{1}[t_i < \infty]$, where μ_i is the cost of selecting vertex i as a seed, and ϵ_i is the revenue generated by the activation of vertex i . Variants with arbitrarily signed parameters μ_i and ϵ_i are also possible. Trajectories with small energy will have a good trade-off between the total cost of their seeds and the total revenue of active nodes. For $\epsilon = 0$, the Boltzmann weight reproduces the dynamics of direct propagation from randomly drawn sets of seeds discussed in the previous section. In this case, it can be shown that the equations become equivalent to those presented in (2) and (3) for direct propagation analysis. On the other hand, in the case $\epsilon > 0$, the causal representation of the dynamics as a DAG is not sufficient to solve the optimization problem as this now implies a backward propagation of information from time $t = \infty$ to time $t = 0$.

B. Derivation of the belief-propagation equations

The representation of the dynamics as a high-dimensional static constraint-satisfaction model over discrete variables (i.e., the activation times) defined on the vertices of a graph makes it possible to apply the cavity method [25] and to develop efficient message-passing algorithms, such as belief propagation (BP) and max-sum (MS). As usual in combinatorial optimization, the variables and their constraints can be represented by means of a factor graph. However, in the static representation of the dynamics, every constraint Ψ_i depends on the values of all activation times in the neighborhood of node i ; therefore nearby constraints Ψ_i and Ψ_j share the two variables t_i and t_j , leading to the appearance of short loops in the corresponding factor graph. In order to eliminate these systematic short loops, we employ a dual factor graph in which variable nodes representing the pair of times (t_i, t_j) are associated with edges $(i, j) \in E$, while the factor nodes are associated with the vertices i of the original graph G and enforce the hard constraints Ψ_i and the contribution \mathcal{E}_i from i to the energy. Figure 3 gives an illustrative example of such a dual construction. Whenever the original graph is locally a

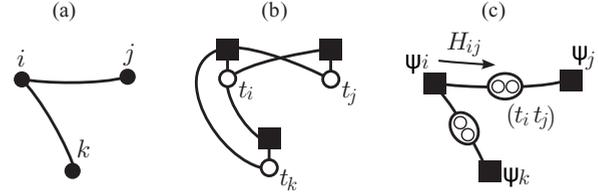


FIG. 3. Dual factor graph representation for the spread optimization problem. (a) Original graph. (b) Naive factor graph formulation, including small loops. (c) Dual factor graph formulation, with variables nodes (t_i, t_j) and (t_i, t_k) and factor nodes Ψ_i, Ψ_j, Ψ_k . The factor Ψ_i must ensure, additionally to the dynamical constraint for vertex i , that t_i components of (t_i, t_j) and (t_i, t_k) coincide.

tree, the dual factor graph is such as well. This property allows one to employ the cavity method. Since the variables appearing in the dual graph are pairs of times (t_i, t_j) , the full distribution can be parametrized in terms of cavity marginals $H_{ij}(t_i, t_j)$ for pairs of times. Let us consider the path probability in (7) and marginalize over all variables but j to compute the probability $P_j(t_j)$ that node j is activated at time t_j . On an infinite tree we have

$$P_j(t_j) \propto \sum_{\{t_i\}_{i \in \partial j}} e^{-\beta \mathcal{E}_j(t_j)} \Psi_j(t_j, \{t_i\}) \prod_{i \in \partial j} H_{ij}(t_i, t_j), \quad (8)$$

where the cavity marginal $H_{ij}(t_i, t_j)$ denotes the probability that nodes i and j are activated at times t_i and t_j in the absence of the constraint Ψ_j and energetic term \mathcal{E}_j . It satisfies the recursive relation

$$H_{ij}(t_i, t_j) \propto e^{-\beta \mathcal{E}_i(t_i)} \sum_{\{t_k\}} \Psi_i(t_i, \{t_k\}) \prod_k H_{ki}(t_k, t_i), \quad (9)$$

where $k \in \partial i \setminus j$. On a general graph, (9) defines the belief-propagation equations which are valid under the hypothesis of fast decay of correlations with the distance or the replica symmetric (RS) assumption [25]. Under this assumption, the statistical properties of the system are described by a unique Gibbs state (i.e., replica symmetry), and the BP equations admit a unique solution.

Given a solution of (9), the marginal probability that neighboring nodes i and j are activated at times t_i and t_j is $P_{ij}(t_i, t_j) \propto H_{ij}(t_i, t_j) H_{ji}(t_j, t_i)$. Equations (9) allow one to access the statistics of atypical dynamical trajectories (e.g., entropies of trajectories or distribution of activation times), but it involves a number of terms which is exponential in the vertex degree. An equivalent but tractable expression can be obtained as follows. For $0 < t_i < \infty$, Eq. (9) can be expressed as

$$H_{ij}(t_i, t_j) \propto e^{-\beta \mathcal{E}_i(t_i)} \sum_{\substack{\theta_1 \geq \theta_i - w_{ji} \mathbb{1}[t_j \leq t_i - 1] \\ \theta_2 < \theta_i - w_{ji} \mathbb{1}[t_j < t_i - 1]}} Q_{ij}^{t_i}(\theta_1, \theta_2), \quad (10)$$

where $Q_{ij}^{t_i}$ is the two-dimensional convolution of functions $f_k^{t_i}(\theta_1, \theta_2) = \sum_{t_k} \delta(\theta_1, w_{ki} \mathbb{1}[t_k < t_i - 1]) \delta(\theta_2, w_{ki} \mathbb{1}[t_k \leq t_i - 1]) H_{ki}(t_k, t_i)$ for $k \in \partial i \setminus j$.

In the limit $\beta \rightarrow \infty$, with a proper rescaling of the messages, (9) gives the max-sum equations and algorithm, which can be used to find explicit solutions at minimum

energy. In this limit, the optimization of the dynamics of the LTM correspond to the *spread maximization problem*, which is computationally hard even to approximate in the worst case [26].

IV. RESULTS ON ENSEMBLES OF RANDOM GRAPHS

On ensembles of (infinitely large) random graphs, the solution of the BP equations (9) can be computed at any finite β using a population dynamics method in the single-link approximation [25].

$$H(0,s) \propto e^{-\beta\mu} p_0^{K-1}, \quad (11a)$$

$$H(t,s) \propto \sum_{\substack{n_- + n_+ + n_0 = K-1 \\ n_- < \theta - \mathbb{1}[s < t-1] \\ \theta - \mathbb{1}[s \leq t-1] \leq n_- + n_0}} \frac{(K-1)!}{n_-! n_+! n_0!} p_t^{K-1-n_-n_0} m_t^{n_-} H(t-1,t)^{n_0} \quad \text{for } 0 < t \leq T, \quad (11b)$$

$$H(\infty,s) \propto e^{-\beta\epsilon} \sum_{n_- \leq \theta - \mathbb{1}[s < T]} \binom{K-1}{n_-} [H(T,\infty) + H(\infty,\infty)]^{K-1-n_-} m_\infty^{n_-}, \quad (11c)$$

where we defined the cumulative messages $p_t = \sum_{t' \geq t} H(t',t)$ and $m_t = \sum_{t' < t-1} H(t',t)$. The normalization constant is just the sum of all messages. The system of equations could be further simplified from $O(T^2)$ messages to $O(T)$ by exploiting the fact that $H(t,s) = H(t, \text{sgn}(t-s+1))$.

The behavior of (11) can be studied by varying μ , ϵ , β , and T for any given assignment of K and θ . We consider the representative cases $K=3, \theta=2$ and $K=4, \theta=2$ in the (ϵ, μ) plane at fixed T and $\beta=1$; then we will comment on the effects of varying T and β . As for the direct dynamics in Sec. II B, we shall consider as observables the density of seeds ρ_0 and the final density (at time T) of active nodes ρ_T . For $\epsilon=0$ and $T \rightarrow \infty$, we recover known results for the static properties of bootstrap percolation [1, 15, 16] in Sec. II B. Although for finite T both cases present a continuous behavior at $\epsilon=0$, the two activation mechanisms are qualitatively different and this difference is amplified in the large-deviation regime.

B. Case $K=3, \theta=2$

For random initial conditions ($\epsilon=0$), the density ρ_T of active nodes in the final state is a continuous function of ρ_0 . Figure 4(a) shows that under optimization the curves develop a gap in the possible values of ρ_0 and ρ_T obtained by varying μ . This means that (for sufficiently large ϵ and β) a value μ^* exists at which both ρ_0 and ρ_T undergo a discontinuous transition, with coexistence and hysteresis phenomena [see Fig. 4(b)]. As β increases the minimum density of seeds admitting full spread ($\rho_T=1$) gradually approaches the values obtained by the MS algorithm (the zero-temperature limit of the BP equations). The total marginal computed from (11) gives the probability $P(t)$ that a node gets activated at time t . The activation time distribution $P(t)$ is displayed in Fig. 4(c) for $T=100$. While

A. Homogeneous solution on random regular graphs

For random regular graphs (RRG) and considering a completely homogeneous setup (i.e., uniform weights $w_{ij}=1 \forall (i,j) \in E$, uniform thresholds $\theta_i=\theta, \forall i \in V$, uniform costs $\mu_i=\mu, \forall i \in V$, and uniform revenues $\epsilon_i=\epsilon, \forall i \in V$), the replica symmetric cavity marginals are expected to be uniform; therefore the population dynamics can be replaced by a self-consistent equation for a single representative BP marginal $H(t,s)$. Since all incoming links are assumed to have the same set of messages, one can group equal messages together, introducing a multinomial distribution and obtaining the following system of nonlinear equations:

for $\epsilon=0$ it always decays exponentially, for $\epsilon>0$ it develops a power-law shape when μ is increased towards the region in which optimization is effective. This means that in order to optimize the dynamics, one can decrease the number of seeds at the cost of generating an activation process that proceeds at a slower pace. The longer the allowed duration T , the smaller

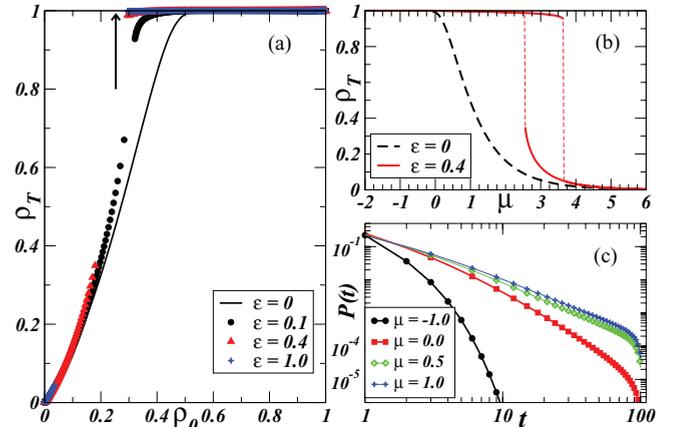


FIG. 4. (Color online) (a) Parametric plot of ρ_T vs ρ_0 obtained by solving the belief-propagation equations in the single-link approximation on regular random graphs of degree $K=3$, for threshold $\theta=2$, duration $T=20$, and $\epsilon=0, 0.1, 0.4, 1$. The vertical arrow indicates the minimum density of seeds ($\rho_0 \approx 0.253$) necessary for the total activation obtained by the max-sum algorithm on finite graphs of size $|V|=10000$. (b) Curves $\rho_T(\mu)$ for $\epsilon=0$ (black dashed line) and 0.4 (red full line). The latter are obtained by following the upper and lower branches of the solution across the transition. (c) Activation time probability $P(t)$ obtained by computing the total BP marginals in a dynamics of duration $T=100$, for $\epsilon=0.4$ and different values of μ .

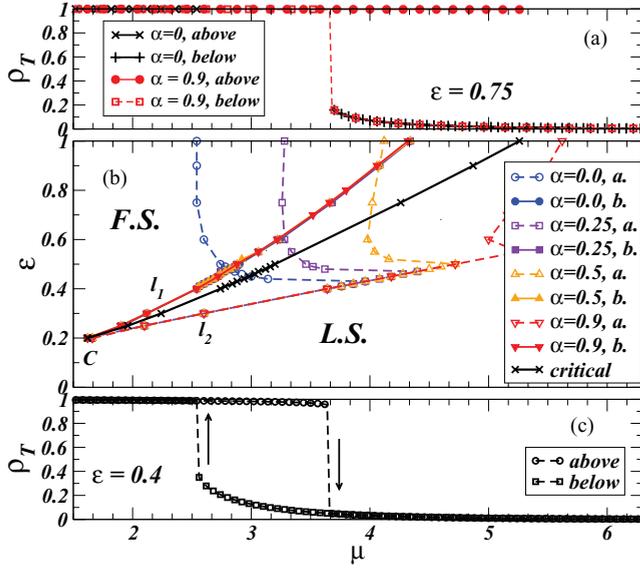


FIG. 5. (Color online) (b) Phase diagram as a function of ϵ and μ for $T = 20$ in random regular graphs of degree $K = 3$ and thresholds $\theta = 2$. The curves are obtained in the single-link approximation using a damping procedure to improve the convergence. Black crosses indicate the location of the thermodynamic phase transition (crossing of the free-energy branches) between a full spread (F.S.) and a low spread (L.S.) phase. The upper and lower curves l_1 and l_2 in the same plot indicate the spinodal lines, obtained for increasing values of the damping factor α from 0 to 0.9. (c) shows the hysteresis phenomenon observed in the density of activated nodes ρ_T as a function of the chemical potential μ for $\epsilon = 0.4$. The same phenomenon for $\epsilon = 0.75$ is shown in (a), where we see that the lack of convergence for $\alpha = 0$ is cured by improving the damping to $\alpha = 0.9$.

the minimum density of seeds required to reach full spread under optimization, but the larger the tail of the distribution.

A tentative phase diagram in the (μ, ϵ) plane, corresponding to the solution of (11) with $T = 20, \beta = 1$, is displayed in Fig. 5. The results are only partially correct because the BP equations do not converge for all values of the parameters. On increasing ϵ from 0, the transition is still continuous, until we encounter a tricritical point $C = (\mu^*, \epsilon^*)$ where the activation transition becomes discontinuous with the appearance of a coexistence phase that grows with $\epsilon > \epsilon^* \approx 0.2$. For moderately small values of ϵ (e.g., $\epsilon = 0.4$ in the bottom panel of Fig. 5), the BP equations converge to their fixed points, and the behavior of the system can be correctly studied for all values of μ . We used a cooling and annealing scheme in μ at fixed ϵ that allowed us to follow the upper (high ρ_T , low μ) and lower (low ρ_T , high μ) branches of the curve $\rho_T(\mu)$ even into the coexistence region. The coexistence phase is limited by two spinodal lines l_1 and l_2 departing from (μ^*, ϵ^*) and indicating the location where the two branches of solutions terminate. It is possible to locate the discontinuous phase transition by comparing the free energy of the two solutions in the coexistence region (black crosses). The meaning of the spinodal lines becomes evident on looking at the bottom panel of Fig. 5, in which we show the behavior of the solutions across the coexistence region for $\epsilon = 0.4$ [see also Fig. 4(b)].

Surprisingly, the spinodal line l_2 (open symbols) seems to present a nonmonotonic behavior with ϵ . This result is just a

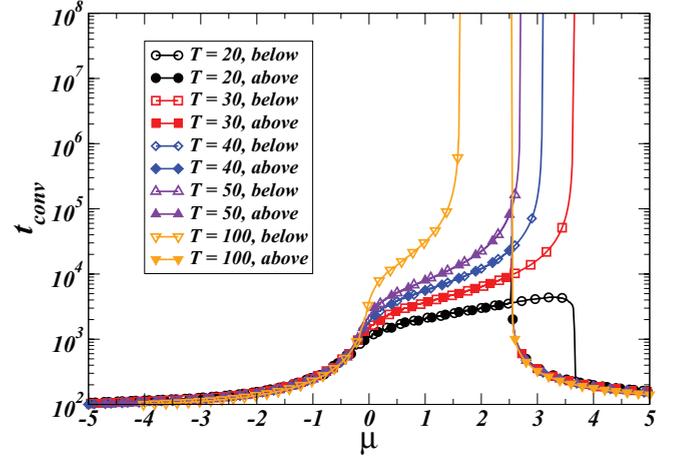


FIG. 6. (Color online) Convergence time (number of iterations) of BP equations for $K = 3, \theta = 2$ in the single-link approximation at $\epsilon = 0.4$ as a function of μ and damping factor $\alpha = 0.9$. Different symbols and colors correspond to different values T of the length of the dynamics. At each value of T we reported two different curves (open and full symbols) corresponding to experiments performed increasing or decreasing the values of μ (in this way following the two branches of solutions).

nonphysical artifact of the lack of convergence of the iteration procedure used to compute the fixed points of (11). In order to improve convergence also for large values of ϵ , we used a “damped” update rule, in which at each iteration, every message is replaced by a linear combination of its old and new values, i.e., $H^{\text{old}}(t, s) \leftarrow \alpha H^{\text{old}}(t, s) + (1 - \alpha) H^{\text{new}}(t, s)$ with $\alpha \in [0, 1]$. On increasing the damping factor α , the convergence properties of (11) are improved and the line l_2 correctly moves smoothly towards larger values of μ (the red line with downward triangles). In the central panel of Fig. 5 we show the effect of nonconvergence on the curves $\rho_T(\mu)$ for $\alpha = 0, 0.25, 0.5, 0.9$. The top panel reports the same plot as the bottom one, namely, $\rho_T(\mu)$, for a larger value of ϵ where the BP equations do not converge without damping. The improvement obtained with a damping factor $\alpha = 0.9$ is evident.

A more sophisticated way of stabilizing the solution scheme for the BP equations (11) is that of using a population of $\mathcal{N}_{\text{pop}} \gg 1$ messages. Solving (11) using population dynamics is very time consuming, but on RRGs the results with $\mathcal{N}_{\text{pop}} \simeq 10^3 - 10^5$ are in agreement with the results obtained using the damped BP equations.

Figure 6 displays the number of iterations t_{conv} necessary to reach the fixed point of (11) as a function of μ for $\epsilon = 0.4$ with damping factor $\alpha = 0.9$. In the region where the optimization of the dynamical process is effective, the convergence time grows continuously until it diverges. For small values of T (e.g., $T = 20$) there is no divergence. At larger T the BP equations stop converging at values of μ that decrease with increasing T . The dependence on T at fixed values of μ is reported in Fig. 7 for $\epsilon = 0.4$. On the contrary, the lack of convergence appears abruptly at the spinodal line (independently of the damping factor α) when μ is decreased from large positive values (Fig. 6).

The above results show that, when the infinite time limit of the direct dynamics presents a continuous activation

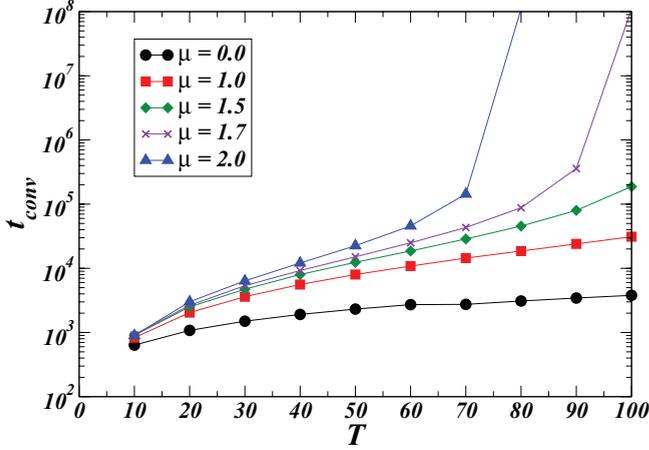


FIG. 7. (Color online) Convergence time of the BP equations as a function of T for different values of μ in random regular graphs with $K = 3$, $\theta = 2$, and $\epsilon = 0.4$.

transition, the optimization of the spreading process is possible and effective. One could also verify how rare the optimal trajectories are by computing their entropy and comparing it with the entropy of random trajectories. We did this in Fig. 8 where we plot the entropy s of the initial conditions that lead to a full spread as a function of the density of seeds. The result for $N = 30$ is obtained by explicit enumeration, whereas for larger systems ($N = 50, 100$) we used a generalization of the cavity method that allows us to fix a global constraint (the number of seeds) by introducing an additional set of messages that flow over a spanning tree superimposed on the original graph (see [27], Appendix B). The curves for the limit of infinite random regular graphs are obtained by computing the entropy, in the cavity approximation, from the fixed-point solution of (11). These quantities for nontypical

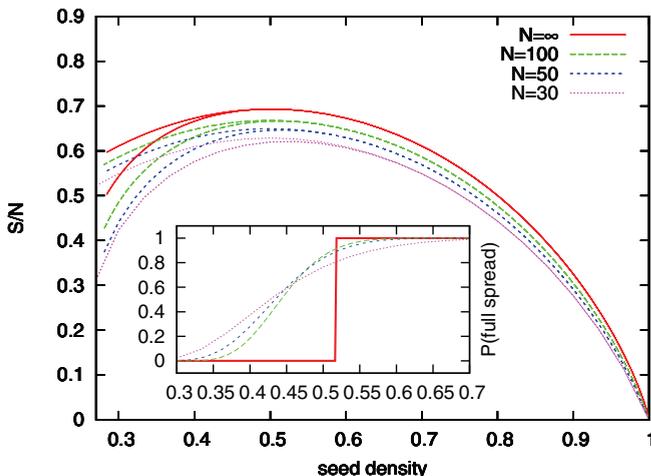


FIG. 8. (Color online) Entropy per site S/N of the solutions of the full spread problem on regular random graphs of degree $K = 3$ and threshold $\theta = 2$, for $N = 30, 50, 100, \infty$ vs seed density ρ_0 . For each N the upper line corresponds to the normalized binomial distribution (i.e., per site entropy of seeds in the absence of optimization) and the lower one to the entropy per site of fully spreading seeds. Inset: Probability P of randomly selecting a fully spreading set of seeds for the same set of parameters.

trajectories are compared with the entropy curves associated with a random choice of initial conditions with fixed density of active nodes ρ_0 , which is given by a binomial sampling of initial seed nodes. When the curves deviate from the binomial, the probability of choosing randomly an optimal set of seeds becomes exponentially small (inset in Fig. 8). In the infinite system, this event is governed by a zero-one law.

C. Case $K = 4, \theta = 2$

In this case, the discontinuous behavior is already present at $\epsilon = 0$ in the limit of large T . For $T = 20$ the curve of ρ_T vs ρ_0 in Fig. 9 is very steep but continuous; however, the underlying dynamics is qualitatively different from that of $K = 3, \theta = 2$ as discussed in Sec. II B. The plot of ρ_T vs ρ_0 in Fig. 9(a) shows that almost no improvement in the density of activated nodes ρ_T is obtained by increasing $\epsilon > 0$. Indeed, for $\epsilon > 0$, the BP equations converge as long as μ is smaller than the critical value corresponding to the abrupt transition for $\epsilon = 0$; then they stop converging in the region where optimization is expected to be effective [see also Fig. 9(b)]. Remarkably, the MS algorithm (supplemented by a reinforcement method [28]) finds full spread solutions that are considerably better than the best BP results (the arrow in Fig. 9).

The activation probability $P(t)$ is very different from that in the previous case. For random seeds ($\epsilon = 0$) the shape of $P(t)$ is not monotonically decreasing, but it develops a second peak that moves towards large times when ρ_0 approaches (from above) the critical value corresponding to the abrupt activation transition. At finite T , the existence of such a peak is a precursor of the discontinuous transition that occurs only for

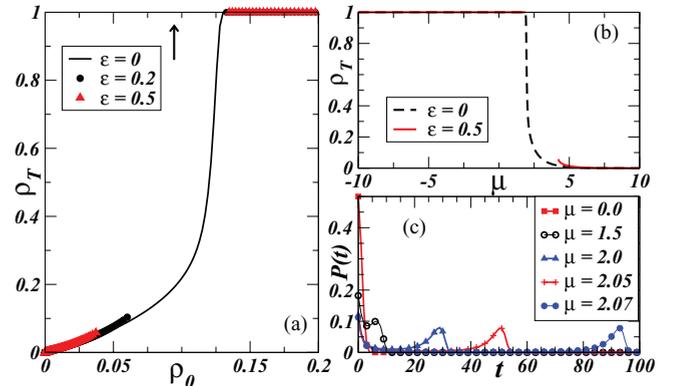


FIG. 9. (Color online) (a) Parametric plot of ρ_T vs ρ_0 obtained by solving (11) in the single-link approximation on regular random graphs of degree $K = 4$, for threshold $\theta = 2$, duration $T = 20$, and $\epsilon = 0, 0.2, 0.5$. No improvement with respect to the random case ($\epsilon = 0$) is obtained by performing optimization ($\epsilon = 0.2, 0.5$). The vertical arrow indicates the minimum density of seeds ($\rho_0 \approx 0.094$) necessary for the total activation obtained by the max-sum algorithm on finite graphs of size $|V| = 10000$. (b) Curves $\rho_T(\mu)$ for $\epsilon = 0$ (black dashed line) and 0.5 (red full line). For $\epsilon = 0.5$ the BP equations do not converge in a region of values of μ close to the discontinuous transition. (c) Activation time probability $P(t)$ obtained by computing the total BP marginals in a dynamics of duration $T = 100$, for $\epsilon = 0$ and different values of μ (for $\epsilon = 0.5$ we obtain exactly the same behavior).

$T \rightarrow \infty$. In this limit, the position of the second peak diverges as ρ_0 approaches the critical point from above. Figure 9(c) shows the behavior of $P(t)$ for $\epsilon = 0$. As long as the BP equations converge, for $\epsilon > 0$ the behavior is the same as that observed for $\epsilon = 0$, with a second peak that appears and gradually moves towards larger times when increasing μ and approaching the threshold of full activation. The peak identifies a “critical mass” of nodes whose dynamical properties are strongly correlated and that are activated almost at the same time. The fact that $P(t)$ does not change for $\epsilon > 0$ is a clue that a large fraction of variables are strongly correlated already at $\epsilon = 0$. The lack of convergence could be due to the onset of long-range correlations responsible for the abrupt activation for $\epsilon = 0$ (and $T = \infty$). It also suggests that there are regions of the parameters in which we expect the space of trajectories to display a complex geometrical structure (e.g., clustering phenomena [29,30]) that cannot be captured by the simple replica symmetric [25] cavity assumption employed here.

Since the effect of the optimization is that of selecting trajectories that allow to postpone as much as possible of this sudden activation, the possibility of controlling the trade-off between the total propagation time T and the number of seeds ρ_0 required to achieve a certain ρ_T is a potentially useful feature of the proposed message-passing algorithms.

The fact that the lack of convergence of the single-link BP equations is here very different from that observed in the case of continuous activation processes can be understood also from the plot of the convergence time of the single-link BP equations in Fig. 10, which turns out to be almost independent of T . The nonconvergence persists if we consider a population of $\mathcal{N}_{\text{pop}} \gg 1$ messages in the single-link approximation and, for the same range of parameter values, it occurs also when solving BP equations on single instances of graphs. In Fig. 11, we show the phase diagram in the (μ, ϵ) plane for $T = 20, \beta = 1$. Already at very small values of ϵ the BP equations stop converging in a region of values of μ that grows with ϵ . The open symbols show that no improvement is obtained using the damping procedure.

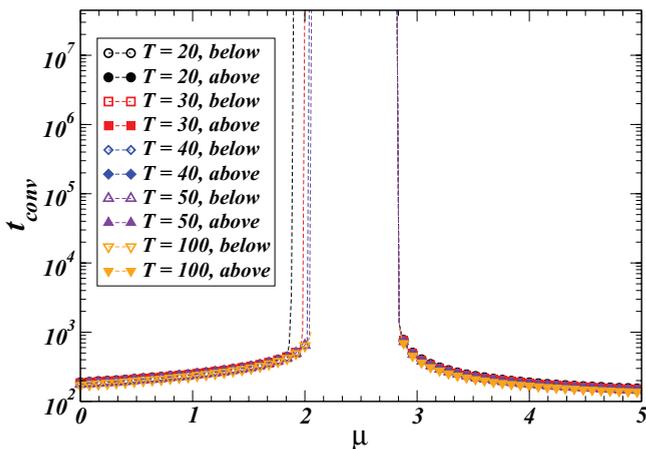


FIG. 10. (Color online) Convergence time (number of iterations) of BP equations for $K = 4, \theta = 2$ in the single-link approximation at $\epsilon = 0.5$ as a function of μ for different values of T .

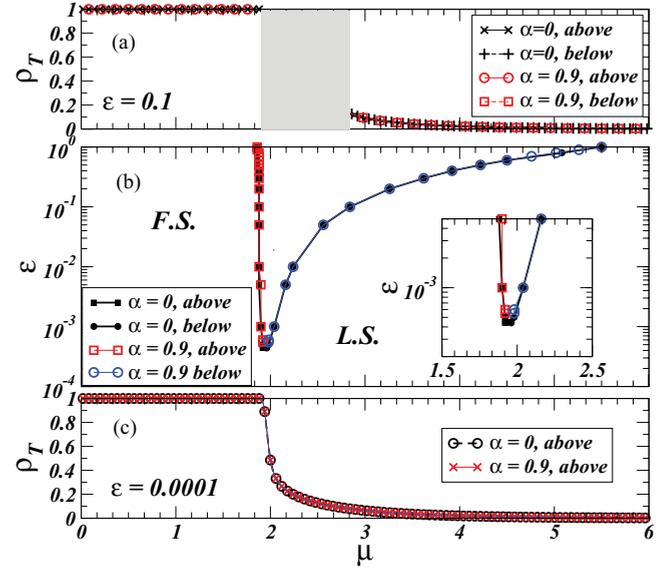


FIG. 11. (Color online) (b) Phase diagram as a function of ϵ and μ for $T = 20$ in random regular graphs of degree $K = 4$ and thresholds $\theta = 2$. Full spread (F.S.) and low spread (L.S.) regions are marked. The curves represent the lines at which BP stops converging in the single-link approximation. The damping procedure (results for $\alpha = 0.9$) does not improve the convergence. The inset highlights the behavior for very small values of ϵ . (c) In this region we can still observe the BP equations converging at some nonzero ϵ . (a) The behavior of the density of activated nodes ρ_T as a function of the chemical potential μ for larger values of ϵ where the BP equations stop converging in a finite interval of values of μ (shaded area).

D. Erdős-Rényi random graphs

To relax the assumption of complete homogeneity of the graphs, we also considered Erdős-Rényi (ER) random graphs, whose degree distribution is a Poisson distribution of average z . In this case we take a population of $\mathcal{N}_{\text{pop}} = 10^3 - 10^5$ cavity marginals to perform the population dynamics in the single-link approximation. At each update iteration, a degree value k is drawn from the degree distribution of the random graph under study and k messages are chosen randomly from the populations. One of the k messages is replaced by the value computed using the remaining $k - 1$ messages as input of the BP equations. The update rule is iterated until convergence. Since evaluating the convergence of the whole population of messages is computationally demanding, we assumed a convergence criterion based on a global observable. More precisely, we required that the difference between the computed values of the average activation time $\tau = \sum_i t P(t)$ before and after a sweep of updates over the whole population is smaller than a fixed tolerance value (we fixed this value to be 10^{-3}). The results, for average degree $z = 5$ and threshold values $\theta_i = \lfloor (k_i + 1)/2 \rfloor$, are shown in Fig. 12. For $\epsilon = 0.2$, the BP equations converge in the whole range of values of μ and we obtain a slight optimization of the trajectories compared to the random case ($\epsilon = 0$). With increasing ϵ , convergence issues are possible. However, on given instances of ER random graphs, one can successfully use the MS algorithm, which is able to activate the full system with a

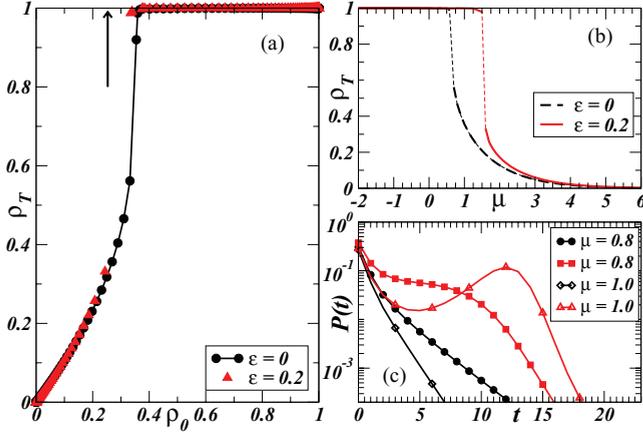


FIG. 12. (Color online) (a) Parametric plot of ρ_T vs ρ_0 obtained by solving the BP equations in the single-link approximation on Erdős-Rényi random graphs of average degree $z = 5$ (minimum degree $k_{\min} = 1$ and maximum degree $k_{\max} = 16$), with thresholds $\theta_i = \lfloor (k_i + 1)/2 \rfloor$, duration $T = 20$, and $\epsilon = 0, 0.2$. (b) Curves $\rho_T(\mu)$ for $\epsilon = 0$ (black dashed line) and 0.4 (red full lines). (c) Activation time probability $P(t)$ obtained by computing the total BP marginals in a dynamics of duration $T = 20$, for $\epsilon = 0$ (black circles and diamonds) and 0.2 (red squares and triangles) and different values of $\mu = 0.8, 1.0$.

density of seeds much lower than the values obtained using the single-link BP approach (see the arrow in Fig. 12).

V. CONCLUSIONS

The study of inverse dynamical problems on large graphs provides an exciting application of message-passing algorithms. The use of single-time cavity marginals for the study of progressive dynamics from randomly distributed seeds is fairly general and it can be straightforwardly applied to models with stochasticity. On the contrary the powerful two-times cavity method developed for the inverse dynamical problem is limited to deterministic settings, such as the LTM, that already include a series of relevant real-world problems. The analysis of large deviations in stochastic progressive models, such as the independent cascades model or the susceptible-infected epidemic model, can be achieved coupling the current representation with the stochastic optimization approach based on a multilevel message-passing construction proposed in [31]. The present analysis paves the way for the derivation of efficient message-passing algorithms for the study of dynamical optimization problems. In fact, the zero-temperature limit of the BP equations provides a max-sum algorithm that can be used to find optimal configurations of seeds ensuring a desired final state, with applications to the design of cost-efficient viral marketing campaigns in social networks and optimal vaccination strategies against epidemic spreading. On the one hand the

fact that a max-sum algorithm takes into account all dynamical constraints makes it more powerful than any centrality-based heuristics usually considered, but its distributed nature makes it much faster than centralized optimization methods based on linear programming and simulated annealing.

ACKNOWLEDGMENTS

The authors acknowledge the European Grants FET Open No. 265496 and ERC No. 267915 and the Italian FIRB Project No. RBF10QUW4.

APPENDIX A: BP EQUATIONS IN THE CAVITY-TIMES REPRESENTATION

Recent works concerning the zero-temperature dynamics of the random-field Ising model [18] and the susceptible-infected model of epidemic spreading [23] have suggested that in the absence of optimization, i.e., when the seeds are randomly drawn, the dynamics can be correctly analyzed using only single-time cavity marginals. In fact, when $\epsilon_i = 0, \forall i \in V$, the two-times formalism that we introduced can be reduced to the set of equations (3) for $\chi_{i\ell}(t) = \Pr_{-\ell}\{x_i^t = 1\}$, i.e., the probability that node i is active at time t in the absence of the neighbor ℓ .

In order to show the equivalence, it is convenient to introduce a different representation of the dynamic rule. For each directed edge (i, ℓ) , the variable $t_{i\ell}$ represents the time at which node i would be activated in the absence of node ℓ , and for $i \notin S$, it satisfies the iterative equation

$$t_{i\ell} = \min_{\sum_{j \in \partial i \setminus \ell} w_{ji} \mathbb{1}[t_{ji} < t] \geq \theta_i} t, \quad (\text{A1})$$

whereas $t_{i\ell} = 0 \forall \ell \in \partial i$ if $i \in S$. As happens for the activation times $\{t_i\}$, also the equations for the cavity times admit a unique solution for a given choice of S , which is in one-to-one correspondence with the solution of the single- t_i model. For convenience, let us define

$$f_i(\{t_k\}_{k \in \partial i}) = \min \left\{ t : \sum_{k \in \partial i} w_{ki} \mathbb{1}[t_k < t] \geq \theta_i \right\}, \quad (\text{A2a})$$

$$f_{ij}(\{t_k\}_{k \in \partial i \setminus j}) = \min \left\{ t : \sum_{k \in \partial i \setminus j} w_{ki} \mathbb{1}[t_k < t] \geq \theta_i \right\}, \quad (\text{A2b})$$

when i is not a seed, $t_{i\ell}$ satisfies the iterative equations $t_{i\ell} = f_{i\ell}(\{t_k\}_{k \in \partial i \setminus \ell})$. In order to optimize the trajectories and average over the initial conditions, we introduce the messages $\hat{H}_{i\ell}(t_{i\ell}, t_{\ell i})$, defined over the cavity times, that represent the joint probability that i and ℓ would be activated at times respectively $t_{i\ell}$ and $t_{\ell i}$ in the absence of the other. The messages satisfy the following BP equations:

$$\begin{aligned} \hat{H}_{i\ell}(t_{i\ell}, t_{\ell i}) &\propto \sum_{\{t_{ki}, t_{ik}\}_{k \in \partial i \setminus \ell}} \prod_{k \in \partial i \setminus \ell} \hat{H}_{ki}(t_{ki}, t_{ik}) \left\{ \prod_{k \in \partial i} \delta(t_{ik}, f_{ik}(\{t_{k'i}\}_{k' \in \partial i \setminus k})) e^{-\epsilon_i \delta(f_i(\{t_{ki}\}_{k \in \partial i}), \infty)} + \prod_{k \in \partial i} \delta(t_{ik}, 0) e^{-\mu_i} \right\} \\ &= \sum_{\{t_{ki}\}_{k \in \partial i \setminus \ell}} \delta(t_{i\ell}, f_{i\ell}(\{t_{ki}\}_{k \in \partial i \setminus \ell})) e^{-\epsilon_i \delta(f_i(\{t_{ki}\}_{k \in \partial i}), \infty)} \prod_{k \in \partial i \setminus \ell} \hat{H}_{ki}(t_{ki}, f_{ik}(\{t_{k'i}\}_{k' \in \partial i \setminus k})) + \delta(t_{i\ell}, 0) \prod_{k \in \partial i \setminus \ell} \hat{H}_{ki}(t_{ki}, 0) e^{-\mu_i}. \end{aligned}$$

When $\epsilon_i = 0 \forall i$, the hypothesis that the messages $\{\hat{H}_{i\ell}(t_{i\ell}, t_{\ell i})\}$ do not depend on the backward cavity times $\{t_{\ell i}\}$ is self-consistently satisfied, and the BP equations can be easily reduced to single-time quantities $\hat{H}_{i\ell}(t_{i\ell})$. To show this, we assume on the right-hand side that $\hat{H}_{ki}(t_{ki}, y) = \hat{H}_{ki}(t_{ki})$ and we get the result that $\hat{H}_{i\ell}(t_{i\ell}, t_{\ell i})$ does not depend on the second argument, i.e.,

$$\hat{H}_{i\ell}(t_{i\ell}) \propto \sum_{\{t_{ki}\}_{k \in \partial i \setminus \ell}} \{\delta(t_{i\ell}, f_{i\ell}(\{t_{ki}\}_{k \in \partial i \setminus \ell})) + \delta(t_{i\ell}, 0)e^{-\mu_i}\} \prod_{k \in \partial i \setminus \ell} \hat{H}_{ki}(t_{ki}). \quad (\text{A3})$$

This implies that the hypothesis is self-consistent and will be verified at every iteration if it is verified at the initial one (e.g., if the messages have uniform initialization).

We need now to consider time-cumulative quantities, such as the probability $\chi_{i\ell}(t) = \sum_{t_{i\ell} \leq t} \hat{H}_{i\ell}(t_{i\ell})$ that node i is active at time t in the absence of node ℓ . Let us first define the following sequence of increasing sets U_t :

$$V_t = \left\{ \{t_{ki}\} : \sum_{k \in \partial i \setminus \ell} w_{ki} \mathbb{1}[t_{ki} < t] \geq \theta_i \wedge \sum_{k \in \partial i \setminus \ell} w_{ki} \mathbb{1}[t_{ki} < t-1] < \theta_i \right\}, \quad U_t = \left\{ \{t_{ki}\} : \sum_{k \in \partial i \setminus \ell} w_{ki} \mathbb{1}[t_{ki} < t] \geq \theta_i \right\},$$

$$U_{t+1} = U_t \cup V_{t+1}, \quad U_t \cap V_{t+1} = \emptyset, \quad \mathbb{1}[U_t] = \sum_{0 < t' \leq t} \mathbb{1}[V_{t'}] = \sum_{0 < t' \leq t} \delta(t', f_{i\ell}(\{t_{ki}\}_{k \in \partial i \setminus \ell})).$$

Let us compute the time-cumulative quantities

$$\begin{aligned} \chi_{i\ell}(t) &= \sum_{t_{i\ell} \leq t} \hat{H}_{i\ell}(t_{i\ell}) \propto \sum_{\{t_{ki}\}_{k \in \partial i \setminus \ell}} \left\{ \sum_{0 < t_{i\ell} \leq t} \delta(t_{i\ell}, f_{i\ell}(\{t_{ki}\}_{k \in \partial i \setminus \ell})) + e^{-\mu_i} \right\} \prod_{k \in \partial i \setminus \ell} \hat{H}_{ki}(t_{ki}) \\ &= e^{-\mu_i} + \sum_{\{t_{ki}\}_{k \in \partial i \setminus \ell}} \mathbb{1} \left[\sum_{k \in \partial i \setminus \ell} w_{ki} \mathbb{1}[t_{ki} < t] \geq \theta_i \right] \prod_{k \in \partial i \setminus \ell} \hat{H}_{ki}(t_{ki}) \\ &= e^{-\mu_i} + \sum_{\{t_{ki}\}_{k \in \partial i \setminus \ell}} \sum_{\{x_k=0,1\}} \prod_{k \in \partial i \setminus \ell} \delta(x_k, \mathbb{1}[t_{ki} < t]) \mathbb{1} \left[\sum_{k \in \partial i \setminus \ell} w_{ki} \mathbb{1}[t_{ki} < t] \geq \theta_i \right] \prod_{k \in \partial i \setminus \ell} \hat{H}_{ki}(t_{ki}) \\ &= e^{-\mu_i} + \sum_{\{x_k=0,1\}} \mathbb{1} \left[\sum_{k \in \partial i \setminus \ell} w_{ki} x_k \geq \theta_i \right] \prod_{k \in \partial i \setminus \ell} \sum_{t_{ki}} \hat{H}_{ki}(t_{ki}) \delta(x_k, \mathbb{1}[t_{ki} < t]) \\ &= e^{-\mu_i} + \sum_{\{x_k=0,1\}} \mathbb{1} \left[\sum_{k \in \partial i \setminus \ell} w_{ki} x_k \geq \theta_i \right] \prod_{k \in \partial i \setminus \ell} \left\{ x_k \sum_{t_{ki} < t} \hat{H}_{ki}(t_{ki}) + (1-x_k) \sum_{t_{ki} \geq t} \hat{H}_{ki}(t_{ki}) \right\} \\ &= e^{-\mu_i} + \sum_{\{x_k=0,1\}} \mathbb{1} \left[\sum_{k \in \partial i \setminus \ell} w_{ki} x_k \geq \theta_i \right] \prod_{k \in \partial i \setminus \ell} \{x_k \chi_{ki}(t-1) + (1-x_k)[1-\chi_{ki}(t-1)]\}. \end{aligned}$$

Denoting by $p_i = e^{-\mu_i}/(1+e^{-\mu_i})$ the probability of choosing i as a seed, setting the initial conditions $\chi_{i\ell}(0) = p_i \forall i$ and fixing the normalization factor $(1+e^{-\mu_i})^{-1} = 1-p_i$, we obtain

$$\chi_{i\ell}(t+1) = p_i + (1-p_i) \sum_{\{x_k=0,1\}} \mathbb{1} \left[\sum_{k \in \partial i \setminus \ell} x_k w_{ki} \geq \theta_i \right] \prod_{k \in \partial i \setminus \ell} \{x_k \chi_{ki}(t) + (1-x_k)[1-\chi_{ki}(t)]\}, \quad (\text{A4})$$

from which we can compute finally the density of active nodes $\rho(t) = \frac{1}{N} \sum_i \rho_i(t)$ [see also (2)]. Apart from the obvious differences in the details of the dynamical update, the single-link version of (A4) (assuming all χ_{ki} identical) is substantially equivalent to Eq. (4) in [18].

We remark that the two-times joint probability introduced in our BP formulation is more than a technical artifact; on the contrary, it is quite crucial to allow information to flow backwards in time when optimizing over the final state ($\epsilon_i > 0$).

APPENDIX B: RELATION TO THE DYNAMIC CAVITY EQUATIONS

Inspired by previous works combining dynamical mean-field theories (such as the dynamical replica theory and generating functional approaches) with the cavity method [32–34], several authors have recently introduced a general formalism to study nonequilibrium dynamical processes on sparse graphs under the name of the *dynamic cavity method* [35–37]. The dynamic cavity method has also strong mathematical

similarities with the application of the cavity analysis to the path-integral representation of quantum spin systems [38,39].

It is easy to show that in the absence of optimization ($\epsilon_i = 0, \forall i \in V$), the belief-propagation equations presented in the main text can be viewed as a simplified version, valid only for microscopically irreversible processes, of the dynamic cavity equations. To this end, we adopt a formulation similar to that used by Neri and Bollé [35] by considering the path probability $P(\underline{\mathbf{x}}^t | \underline{\mathbf{h}}^t)$ of a trajectory $\underline{\mathbf{x}}^t = \{\mathbf{x}^0, \dots, \mathbf{x}^t\}$ with $\mathbf{x}^t = \{x_1^t, \dots, x_N^t\}$ in the presence of an external field $\underline{\mathbf{h}}^t = \{\mathbf{h}^0, \dots, \mathbf{h}^t\}$. The path probability can be written in terms of the marginals computed on a cavity graph in which node i is removed as

$$P(\underline{\mathbf{x}}^t | \underline{\mathbf{h}}^t) = P_{(i)}(\underline{\mathbf{x}}^t | \underline{\mathbf{h}}^t + \underline{\mathbf{u}}_{(i)}^t) \prod_{s=1}^t W[x_i^s | \mathbf{x}^{s-1}; h_i^s] p_0(x_i^0), \quad (\text{B1})$$

where $\underline{\mathbf{u}}_{(i)}^t$ is an auxiliary external field acting over the neighbors of i and introduced to keep track of the directed influence of i over them throughout the dynamics. For simplicity we have taken a factorized distribution over the initial conditions. By summing (B1) over all possible trajectories of all nodes $j \neq i$, we get the local marginal $P_i(\underline{\mathbf{x}}_i^t | \underline{\mathbf{h}}^t)$ that satisfies the equation

$$\begin{aligned} P_i(\underline{\mathbf{x}}_i^t | \underline{\mathbf{h}}^t) &= \sum_{\underline{\mathbf{y}}_{\partial i}^t} P_{(i)}(\underline{\mathbf{y}}_{\partial i}^t | \underline{\mathbf{h}}^t + \underline{\mathbf{u}}_{(i)}^t) \prod_{s=1}^t W[x_i^s | \mathbf{y}_{\partial i}^{s-1}; h_i^s] p_0(x_i^0) \\ &= \sum_{\underline{\mathbf{y}}_{\partial i}^{t-1}} P_{(i)}(\underline{\mathbf{y}}_{\partial i}^{t-1} | \underline{\mathbf{h}}^{t-1} + \underline{\mathbf{u}}_{(i)}^{t-1}) \prod_{s=1}^t W[x_i^s | \mathbf{y}_{\partial i}^{s-1}; h_i^s] p_0(x_i^0), \end{aligned} \quad (\text{B2})$$

where $\underline{\mathbf{y}}_{\partial i}^t$ is the joint trajectory of the neighbors of i up to time t , and the last passage comes from the fact that the dynamics is parallel and the transition probability for variable i at time t depends only on the states of neighbors at time $t-1$. Exploiting the treelike assumption, we can express the cavity marginal for the neighbors of i in factorized form,

$$P_{(i)}(\underline{\mathbf{y}}_{\partial i}^{t-1} | \underline{\mathbf{h}}^{t-1} + \underline{\mathbf{u}}_{(i)}^{t-1}) \propto \prod_{j \in \partial i} P_{(i)j}(\underline{\mathbf{y}}_j^{t-1} | \underline{\mathbf{h}}_j^{t-1} + \underline{\mathbf{u}}_{(i)j}^{t-1}). \quad (\text{B3})$$

In the present case, we have $h_j^s = -\theta_j$ and $u_{(i)j}^s = w_{ij} x_i^{s-1}$ for all times $s > 0$, while the transition probability is given by the deterministic update rule as follows:

$$W[x_i^s = 0 | \mathbf{y}_{\partial i}^{s-1}; h_i^s] = \mathbb{1} \left[\sum_{k \in \partial i} w_{ki} y_k^{s-1} < \theta_i \right], \quad (\text{B4})$$

$$W[x_i^s = 1 | \mathbf{y}_{\partial i}^{s-1}; h_i^s] = \mathbb{1} \left[\sum_{k \in \partial i} w_{ki} y_k^{s-1} \geq \theta_i \right]. \quad (\text{B5})$$

For a given directed edge (i, j) , thus the local field can be univocally parametrized in terms of the variable in the removed node, leading to the following set of recursive equations for the cavity marginals $P_{(\ell)i}(\underline{\mathbf{x}}_i^t | \underline{\mathbf{h}}_i^t + \underline{\mathbf{u}}_{(\ell)i}^t) \equiv P_{i\ell}(\underline{\mathbf{x}}_i^t, \underline{\mathbf{x}}_\ell^t)$:

$$\begin{aligned} P_{i\ell}(\underline{\mathbf{x}}_i^t, \underline{\mathbf{x}}_\ell^t) &\propto p_0(x_i^0) \sum_{\{\underline{\mathbf{x}}_j^{t-1}\}_{j \in \partial i \setminus \ell}} \prod_{s=1}^t W[x_i^s | \mathbf{x}_{\partial i}^{s-1}; -\theta_i] \\ &\times \prod_{j \in \partial i \setminus \ell} P_{ji}(\underline{\mathbf{x}}_j^{t-1}, \underline{\mathbf{x}}_i^{t-1}). \end{aligned} \quad (\text{B6})$$

A path probability is now a vector of $O(4^{t+1})$ variables, but further reduction is possible because the dynamics is microscopically irreversible. The full sequence of $t+1$ binary values taken by $\underline{\mathbf{x}}_i^t = \{x_i^0, \dots, x_i^t\}$ can be encoded in a single integer $t_i = \{0, \dots, t\}$ representing the time at which the variable i flips from 0 to 1 (with the convention $t_i = 0$ for a seed). With this parametrization, $P_{ij}(\underline{\mathbf{x}}_i^t, \underline{\mathbf{x}}_j^t)$ becomes $Q_{ij}(t_i, t_j)$ and the dynamic cavity equations (B6) take the form

$$\begin{aligned} Q_{i\ell}(t_i, t_\ell) &\propto p_0(\delta_{t_i, 0}) \sum_{\{t_j\}_{j \in \partial i \setminus \ell}} \mathbb{1} \left[\sum_{j \in \partial i} w_{ji} \mathbb{1}[t_j < t_i - 1] < \theta_i \right] \mathbb{1} \\ &\times \left[\sum_{j \in \partial i} w_{ji} \mathbb{1}[t_j < t_i] \geq \theta_i \right] \prod_{j \in \partial i \setminus \ell} Q_{ji}(t_j, t_i) \\ &= e^{-c_i \delta_{t_i, 0}} \sum_{\{t_j\}_{j \in \partial i \setminus \ell}} \mathbb{1} \left[\sum_{j \in \partial i} w_{ji} \mathbb{1}[t_j < t_i - 1] < \theta_i \right] \mathbb{1} \\ &\times \left[\sum_{j \in \partial i} w_{ji} \mathbb{1}[t_j < t_i] \geq \theta_i \right] \prod_{j \in \partial i \setminus \ell} Q_{ji}(t_j, t_i), \end{aligned} \quad (\text{B7})$$

where we represented the initial factorized distribution in terms of weights over the seeds, i.e., $p_0(\delta_{t_i, 0}) = e^{-\mu_i \delta_{t_i, 0}} / (1 + e^{-\mu_i \delta_{t_i, 0}})$. It is easy to check that (B7) corresponds to the BP equations derived in the main text for the cavity messages $H_{i\ell}(t_i, t_\ell)$ in the case in which all $\epsilon_i = 0$. From the above derivation it is also evident that optimization could be included by introducing in the dynamic cavity equation an additional local energy term $-\sum_i \sum_{s \leq t} \epsilon_i \delta_{x_i^s, 1}$ in order to account for optimization over the trajectories, which in the activation time representation becomes $\sum_i \epsilon_i \delta_{t_i, \infty}$.

It could be in principle possible to relax the assumption of complete irreversibility, using probability distributions over time-dependent paths, but the optimization of fully reversible dynamics is currently numerically infeasible and it requires additional separate ideas to overcome current computational limitations.

[1] J. Chalupa, P. L. Leath, and G. R. Reich, *J. Phys. C* **12**, L31 (1979).

[2] D. Dhar, P. Shukla, and J. P. Sethna, *J. Phys. A* **30**, 5259 (1997).

[3] M. Sellitto, G. Biroli, and C. Toninelli, *Europhys. Lett.* **69**, 496 (2005); C. Toninelli, G. Biroli, and D. S. Fisher, *Phys. Rev. Lett.*

96, 035702 (2006).

- [4] M. E. J. Newman, *Phys. Rev. E* **66**, 016128 (2002).
- [5] R. O'Dea, J. J. Crofts, and M. Kaiser, *J. R. Soc. Interface* **10**, 1742 (2013).
- [6] B. Kholodenko, M. B. Yaffe, and W. Kolch, *Sci. Signal.* **5**, re1 (2012).
- [7] T. C. Schelling, *J. Conflict Resolution* **17**, 381 (1973).
- [8] M. Granovetter, *Am. J. Sociol.* **83**, 1420 (1978).
- [9] D. J. Watts, *Proc. Natl. Acad. Sci. USA* **99**, 5766 (2002).
- [10] M. Jackson and L. Yariv, *Econ. Publique* **16**, 69 (2005).
- [11] D. Acemoglu, A. Ozdaglar, and E. Yildiz, in *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC), 12–15 Dec. 2011, Orlando FL* (IEEE, Piscataway, NJ, 2011), pp. 2329–2334.
- [12] L. Eisenberg and T. H. Noe, *Manage. Sci.* **47**, 236 (2001).
- [13] A. G. Haldane and R. M. May, *Nature (London)* **469**, 351 (2011).
- [14] J. Kleinberg, in *Algorithmic Game Theory*, edited by N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani (Cambridge University Press, Cambridge, 2007).
- [15] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Phys. Rev. Lett.* **96**, 040601 (2006).
- [16] J. Balogh and B. Pittel, *Random Struct. Alg.* **30**, 257 (2007).
- [17] C. L. Farrow, P. Shukla, and P. M. Duxbury, *J. Phys. A* **40**, F581 (2007).
- [18] H. Ohta and S.-I. Sasa, *Europhys. Lett.* **90**, 27008 (2010).
- [19] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008).
- [20] J. P. Gleeson, *Phys. Rev. Lett.* **107**, 068701 (2011).
- [21] H. Touchette, *Phys. Rep.* **478**, 1 (2009).
- [22] D. Kempe, J. Kleinberg, and É. Tardos, in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC* (ACM, New York, NY, USA, 2003), pp. 137–146.
- [23] B. Karrer and M. E. J. Newman, *Phys. Rev. E* **82**, 016101 (2010).
- [24] P. A. Noël, A. Allard, L. Hébert-Dufresne, V. Marceau, and L. J. Dubé, *Phys. Rev. E* **85**, 031118 (2012).
- [25] M. Mézard and G. Parisi, *Eur. Phys. J. B* **20**, 217 (2001); *J. Stat. Phys.* **111**, 1 (2003).
- [26] Z. Lu, W. Zhang, W. Wu, J. Kim, and B. Fu, *J. Comb. Optim.* **24**, 374 (2011).
- [27] A. Braunstein, A. Ramezanpour, R. Zecchina, and P. Zhang, *Phys. Rev. E* **83**, 056114 (2011).
- [28] M. Bailly-Bechet, C. Borgs, A. Braunstein, J. Chayes, A. Dagkessamanskaia, J. M. François, and R. Zecchina, *Proc. Natl. Acad. Sci. USA* **108**, 882 (2011).
- [29] G. Parisi, M. Mézard, and M. A. Virasoro, *Spin Glass Theory and Beyond*, Lecture Notes in Physics (World Scientific, Singapore, 1986).
- [30] M. Mézard and A. Montanari, *Information, Physics and Computation*, Oxford Graduate Texts (Oxford University Press, Oxford, 2009).
- [31] F. Altarelli, A. Braunstein, A. Ramezanpour, and R. Zecchina, *J. Stat. Mech.: Theory Exp.* (2011) P11009; *Phys. Rev. Lett.* **106**, 190601 (2011).
- [32] J. P. L. Hatchett, B. Wemmenhove, I. Pérez Castillo, T. Nikoletopoulos, N. S. Skantzos, and A. C. C. Coolen, *J. Phys. A* **37**, 6201 (2004).
- [33] J. P. L. Hatchett, I. Pérez Castillo, A. C. C. Coolen, and N. S. Skantzos, *Phys. Rev. Lett.* **95**, 117204 (2005).
- [34] A. Mozeika and A. C. C. Coolen, *J. Phys. A* **41**, 115003 (2008).
- [35] I. Neri and D. Bollé, *J. Stat. Mech.: Theory Exp.* (2009) P08009.
- [36] Y. Kanoria and A. Montanari, *Ann. Appl. Probab.* **21**, 1694 (2011).
- [37] E. Aurell and H. Mahmoudi, *J. Stat. Mech.: Theory Exp.* (2011) P04014.
- [38] C. Laumann, A. Scardicchio, and S. L. Sondhi, *Phys. Rev. B* **78**, 134424 (2008).
- [39] F. Krzakala, A. Rosso, G. Semerjian, and F. Zamponi, *Phys. Rev. B* **78**, 134428 (2008).