

**Generalized percolation in random directed networks**

Marián Boguñá and M. Ángeles Serrano

*Departament de Física Fonamental, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain*

(Received 24 January 2005; published 7 July 2005)

We develop a general theory for percolation in directed random networks with arbitrary two-point correlations and bidirectional edges—that is, edges pointing in both directions simultaneously. These two ingredients alter the previously known scenario and open new views and perspectives on percolation phenomena. Equations for the percolation threshold and the sizes of the giant components are derived in the most general case. We also present simulation results for a particular example of uncorrelated network with bidirectional edges confirming the theoretical predictions.

DOI: [10.1103/PhysRevE.72.016106](https://doi.org/10.1103/PhysRevE.72.016106)

PACS number(s): 89.75.Hc, 64.60.Ak

**I. INTRODUCTION**

A wide class of real systems of many interacting elements can be mapped onto graphs or networks. Under this approach, vertices or nodes of the network represent the elements of the system whereas edges or links among them stand for interactions between different elements. This mapping has triggered a huge number of works and a surge of interest in the field of complex networks that has led to a general framework within which to analyze their topology as well as the dynamical processes running on top of them [1–3]. In many cases, these dynamical processes are directly related to functionality and involve some kind of transport or traffic flow. Furthermore, the very existence of those networks could be naturally explained as a direct consequence of the communication need among its constituents. The Internet or the World Wide Web are clear examples [4]. In order to preserve functionality, networks characterized by transport processes must be connected; that is, a path must exist between any pair of nodes, or, at least, there must exist a macroscopic portion of vertices—or giant component—able to communicate. In this context, percolation theory appears as an indispensable tool to analyze the conditions under which such connected structures emerge in large networks.

The general theory of percolation phenomena for uncorrelated undirected random networks was first developed by Newman and co-workers [5,6] after previous results in [7,8]. The phase transition at which the giant component forms was well characterized, and the size distribution of connected finite components below and above the critical point was calculated as well. Some further refinements were needed in order to approach real nets. Hence, correlations between degrees of neighboring vertices were taken into account in [9–11] and growing networks were treated by Dorogovtsev *et al.* [12] and Krapivsky and Derrida [13].

To go further, directness must be taken into consideration since some of the most interesting real networks present asymmetric interactions. Noticeable examples are the World Wide Web [14], citation networks [15,16], email networks [17], gene regulatory networks [18], or metabolic networks [19]. Percolation theory for purely directed networks was first developed by Newman and co-workers [5,6] and later by Dorogovtsev *et al.* [20]. In the particular case of scale-

free degree distributions a number of interesting specific results were obtained [21]. While allowing for general correlations between the incoming and outgoing numbers of edges of a given vertex, all these studies refer to networks with no degree correlations. Furthermore, the theory is restricted to the class of directed networks with no bidirectional edges, although this class of edges is ubiquitous to all real directed networks (see [22] and references therein). All these limit the applicability of the theory to real networks since bidirectional edges and degree correlations are common to all real directed networks. In this paper we present a general theory for percolation in directed random networks with general two-point correlations and bidirectional links. We will show that the presence of bidirectional edges and degree correlations modify the picture previously drawn [5,6,20,21] in a nontrivial way, opening new scenarios for percolation phenomena.

The paper is organized as follows. In Sec. II, we review concepts, definitions, and the main results previously obtained in the analysis of percolation in random networks. In Sec. III, we develop the general theory for directed networks with bidirectional edges and arbitrary degree correlations and we show how the theories for undirected and purely directed random networks stand as particular cases. Section IV is devoted to the uncorrelated case, which deserves special attention as a null or benchmark model. The relative sizes of the giant components are computed and the explicit expression for the percolation condition is provided. The well-known critical points signaling the phase transition in undirected and purely directed uncorrelated networks are recovered as limiting cases. A practical application of the formalism for uncorrelated networks is presented in Sec. V, where the transformation from a purely directed network to a purely bidirectional one is studied. For power-law degree distributions, the transformation is shown to undergo a nontrivial phase transition. Simulation results support this prediction, finding an excellent agreement with numerical solutions of the theoretical equations. Finally, we conclude with a brief report of results in Sec. VI.

**II. PERCOLATION IN UNCORRELATED PURELY DIRECTED NETWORKS**

The topological structure of directed networks is more complex than that associated with undirected graphs. The

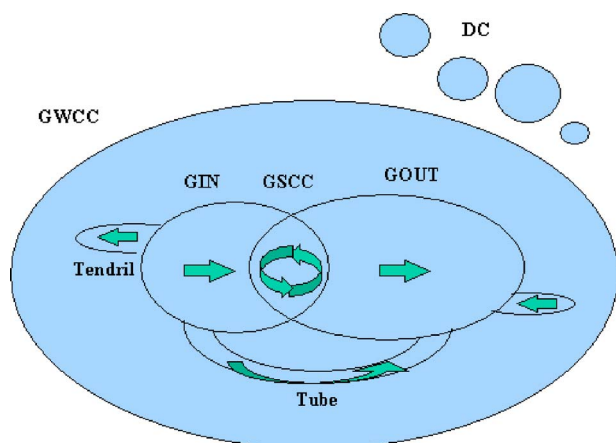


FIG. 1. (Color online) Schematic representation of the component structure of a directed network.

edges associated with each node in a directed net are usually differentiated into incoming and outgoing. Usually, no bidirectional links are considered so that each vertex has two coexisting degrees  $k_i$  and  $k_o$ , which sum up to the total degree  $k = k_{in} + k_{out}$ . Hence, the degree distribution for a directed network is a joint degree distribution  $P(k_i, k_o)$  of in- and out-degrees, which in general may be correlated. The bidirectional edge symmetry of undirected networks is thus completely broken in purely directed ones, with implications down to the level of percolation properties. The giant connected component in undirected graphs becomes internally structured in the case of directed networks so that four different types of giant components may arise (see Fig. 1). Whether giant or not, these components are characterized as follows (according to definitions in [20]).

(i) The weakly connected component (WCC), the percolative cluster in undirected graphs. In the WCC, every vertex is reachable from every other, provided that the directed nature of the edges is ignored.

(ii) The strongly connected component (SCC), the set of vertices reachable from its every vertex by a directed path.

(iii) The in-component (IN), all vertices from which the SCC is reachable by a directed path.

(iv) The out-component (OUT), all vertices which are reachable from the SCC by a directed path.

Notice that, with these definitions, the SCC is included in both the IN and OUT components. The percolation theory developed by Newman and co-workers [5,6] and Dorogovtsev *et al.* [20] for directed graphs with arbitrary degree distribution and statistically uncorrelated vertices has shown that there are two phase transitions: the one at which the giant weakly connected component (GWCC) appears and the one at which the other three giant components appear simultaneously: the giant in-component (GIN), the giant out-component (GOUT), and the giant strongly connected component (GSCC) as the intersection of the other two. The first phase transition corresponds in fact to the standard phase transition in an undirected random graph with arbitrary degree sequence and statistically uncorrelated vertices. The condition for this phase transition was first given by Molloy and Reed [7] and reads

$$\sum_k k(k-2)P(k) \geq 0. \quad (1)$$

When this condition is fulfilled and although some disconnected finite components may remain, the GWCC emerges. It contains a macroscopic portion of the vertices in the network capable of communicating with each other regardless of the orientation of their links. The second phase transition is characteristic of directed networks. The critical point [6]

$$\sum_{k_i, k_o} k_o(k_i - 1)P(k_i, k_o) = 0 \quad (2)$$

marks the first simultaneous appearance of the other three giant components: the GSCC, the GIN, and the GOUT, as well as other secondary structures such as tubes or tendrils [14]. In this case, the different giant components can be quantified using their relative sizes  $g_{in}$ ,  $g_{out}$ , and  $g_{sc}$ , which measure the number of vertices in each set divided by the total number of vertices.

The efforts to understand how this landscape is modified by the consideration of correlations between degrees of neighboring vertices has been exclusively focused on the percolation analysis of undirected networks [9–11]. Assortative mixing by degree, observed in the vast majority of social real networks, has been found to favor percolation in the sense that the giant component appears at lower edge density. On the contrary, disassortative correlations, characteristic of technological and biological networks, make difficult the formation of the giant component even if the second moment of the degree distribution diverges.

### III. GENERALIZED PERCOLATION

In this section we develop a general theory for percolation in directed random networks with arbitrary two-point degree correlations and bidirectional edges. We term our networks “random”—or Markovian—because, apart from purely local properties and two-point correlations being fixed, networks are maximally random. This implies that the whole topology is encoded into the degree distribution  $P(\mathbf{k}) \equiv P(k_i, k_o, k_b)$  and the transition probabilities  $P_i(\mathbf{k}'|\mathbf{k})$ ,  $P_o(\mathbf{k}'|\mathbf{k})$ , and  $P_b(\mathbf{k}'|\mathbf{k})$  measuring the probability to reach a vertex of degree  $\mathbf{k}'$  leaving from a vertex of degree  $\mathbf{k}$  using an incoming, outgoing, or bidirectional edge, respectively. Notice that we have used the notation  $\mathbf{k} \equiv (k_i, k_o, k_b)$  and that we consider all three kind of edges as independent entities. These transition probabilities are related through the following degree detailed balance conditions [23]:

$$k_b P(\mathbf{k}) P_b(\mathbf{k}'|\mathbf{k}) = k'_b P(\mathbf{k}') P_b(\mathbf{k}|\mathbf{k}') \quad (3)$$

and

$$k_o P(\mathbf{k}) P_o(\mathbf{k}'|\mathbf{k}) = k'_o P(\mathbf{k}') P_o(\mathbf{k}|\mathbf{k}'). \quad (4)$$

These conditions assure that any edge leaving a vertex points to another vertex or, in other words, that the network is closed. Notice that this may not be the case in situations where information is incomplete and the out-degree of a vertex is known but not the neighbors at the end of these edges.

To analyze the percolation properties for this class of networks, it is necessary to calculate the joint distribution  $G(s, s')$  of the number of vertices (plus itself),  $s$ , that are reachable from a given vertex and the number of vertices (plus itself),  $s'$ , that can reach that vertex. Notice that we can leave a vertex using an outgoing or bidirectional edge and we can arrive at a vertex through an incoming or bidirectional edge. These sets of vertices are called the out- and in-components of a given vertex, respectively. Analogously, we can define the marginal probabilities  $G_o(s)$  and  $G_i(s)$  for the number of reachable vertices from a given one and the number of vertices that can reach it, respectively. These three functions contain the information of the sizes of the different giant components of the network. Notice that, if the network is above the percolation threshold,  $G_o(s) \neq \sum_{s'} G(s, s')$  and  $G_i(s) \neq \sum_{s'} G(s', s)$ . The probability to belong to a finite component is smaller than 1 in this situation, which implies that the remainder corresponds to the probability to belong to an infinite component—that is, the giant component. The function  $G_o(s)$ , for instance, measures the probability that a given vertex has a finite out-component of size  $s$  regardless of the size of its in-component, which can be finite or not. On the other hand, the function  $\sum_{s'} G(s, s')$  only accounts for finite components. Therefore, the relative sizes of the different giant components of the network can be written as

$$g_{out} = 1 - \sum_s G_i(s), \quad g_{in} = 1 - \sum_s G_o(s) \quad (5)$$

and

$$g_{scc} = 1 - \sum_s G_o(s) - \sum_s G_i(s) + \sum_{s,s'} G(s, s'), \quad (6)$$

where the GOUT is thought of as the set of vertices with an infinite in-component, the GIN as the set of vertices with an infinite out-component, and the GSCC as the set of vertices with infinite in- and out-components simultaneously.

In heterogeneous networks, this set of probabilities depends on the degree of the vertex from where we start the count. Therefore, the functions  $G(s, s')$ ,  $G_o(s)$ , and  $G_i(s)$  are to be expressed as

$$G_i(s) = \sum_{\mathbf{k}} P(\mathbf{k}) G_i(s|\mathbf{k}), \quad G_o(s) = \sum_{\mathbf{k}} P(\mathbf{k}) G_o(s|\mathbf{k}) \quad (7)$$

and

$$G(s, s') = \sum_{\mathbf{k}} P(\mathbf{k}) G(s, s'|\mathbf{k}), \quad (8)$$

where the functions  $G$  inside the summations have the same meaning as the original ones under the condition of starting from a vertex of degree  $\mathbf{k}$ . Thus, the complete solution of the problem goes through finding the conditional probabilities  $G_i(s|\mathbf{k})$ ,  $G_o(s|\mathbf{k})$ , and  $G(s, s'|\mathbf{k})$ . In the following subsections we will show how to compute them in the general correlated case.

### A. In- and out components

To proceed further, we first focus our attention on the out-component size distribution of a vertex of degree  $\mathbf{k}$ ,

$G_o(s|\mathbf{k})$ . Starting from a vertex of degree  $\mathbf{k}=(k_i, k_o, k_b)$ , we can leave it using the  $k_o$  outgoing edges and the  $k_b$  bidirectional ones. Then, the number of reachable vertices will be the sum of the reachable vertices of each of the  $k_o+k_b$  neighbors plus 1. In mathematical terms, this translates into

$$G_o(s|\mathbf{k}) = \sum_{s_1} \cdots \sum_{s_{k_o+k_b}} g_o(s_1|\mathbf{k}) \cdots g_o(s_{k_o}|\mathbf{k}) \times g_o^b(s_{k_o+1}|\mathbf{k}) \cdots g_o^b(s_{k_o+k_b}|\mathbf{k}) \delta_{s_1+\cdots+s_{k_o+k_b}+1, s}, \quad (9)$$

where  $g_o(s|\mathbf{k})$  [ $g_o^b(s|\mathbf{k})$ ] is the distribution of the number of reachable vertices from a vertex given that we have arrived to it from another source vertex of degree  $\mathbf{k}$  following one of its outgoing [bidirectional] edges. In writing Eq. (9), we have used the fact that sparse random networks—those with finite average degree—are locally tree like.<sup>1</sup> Equations of the type of Eq. (9) find in the discrete Laplace space their natural representation in terms of the generating function formalism. Using this formalism, these equations simplify enormously and can be manipulated very easily. Within this formalism, Eq. (9) simplifies as

$$\hat{G}_o(z|\mathbf{k}) = z[\hat{g}_o(z|\mathbf{k})]^{k_o}[\hat{g}_o^b(z|\mathbf{k})]^{k_b}, \quad (10)$$

where we have adopted the notation  $\hat{f}(z) \equiv \sum_s f(s)z^s$ . In what follows, we will work in the discrete Laplace space, using the generating function formalism.

The functions  $g_o$  and  $g_o^b$  satisfy the following set of coupled equations:

$$\hat{g}_o(z|\mathbf{k}) = z \sum_{\mathbf{k}'} P_o(\mathbf{k}'|\mathbf{k}) [\hat{g}_o(z|\mathbf{k}')]^{k'_o} [\hat{g}_o^b(z|\mathbf{k}')]^{k'_b},$$

$$\hat{g}_o^b(z|\mathbf{k}) = z \sum_{\mathbf{k}'} P_b(\mathbf{k}'|\mathbf{k}) [\hat{g}_o(z|\mathbf{k}')]^{k'_o} [\hat{g}_o^b(z|\mathbf{k}')]^{k'_b-1}. \quad (11)$$

The term  $k_b-1$  in the second equation of Eqs. (11) comes from the fact that one of the bidirectional edges has already been used to reach the vertex of degree  $\mathbf{k}'$  and, thus, cannot be used again to leave it. Notice that this restriction is not needed in the first equation of Eqs. (11) since, in this case, we have reached the vertex using an outgoing edge of the source vertex. The set of equations (11) is closed for the functions  $g$ . Its solution for  $z=1$  will allow us to compute the relative size of the GIN component,

$$g_{in} = 1 - \sum_{\mathbf{k}} P(\mathbf{k}) [\hat{g}_o(1|\mathbf{k})]^{k_o} [\hat{g}_o^b(1|\mathbf{k})]^{k_b}. \quad (12)$$

The trivial solution of Eqs. (11) is  $\hat{g}_o(1|\mathbf{k})=1$ ,  $\hat{g}_o^b(1|\mathbf{k})=1$ , corresponding to the only case without giant component.

<sup>1</sup>The condition of being sparse is typically fulfilled by real random networks. This allows for a proper definition of statistical quantities, like the degree distribution, independently of the size of the network. On the other hand, the condition of being locally tree like is, in some cases, not totally satisfied and the clustering coefficient must be taken under consideration. Nevertheless, percolation theory for random networks represents an upper bound to the real percolation phenomena and, thus, is equally important.

Therefore, the network will percolate at the directed level when this trivial solution becomes unstable. To analyze the stability of this solution we use the approach adopted in [11] and find solutions of the form  $\hat{g}_o(1|\mathbf{k})=1-\epsilon x(\mathbf{k})$ ,  $\hat{g}_o^b(1|\mathbf{k})=1-\epsilon y(\mathbf{k})$  in the limit  $\epsilon \rightarrow 0$ . Replacing these expressions in Eqs. (11) and taking the limit  $\epsilon \rightarrow 0$  we obtain

$$\begin{pmatrix} x(\mathbf{k}) \\ y(\mathbf{k}) \end{pmatrix} = \sum_{\mathbf{k}'} \mathbf{C}_{\mathbf{k}\mathbf{k}'}^o \begin{pmatrix} x(\mathbf{k}') \\ y(\mathbf{k}') \end{pmatrix}, \quad (13)$$

where the matrix  $\mathbf{C}_{\mathbf{k}\mathbf{k}'}^o$  is defined as

$$\mathbf{C}_{\mathbf{k}\mathbf{k}'}^o \equiv \begin{pmatrix} k'_o P_o(\mathbf{k}'|\mathbf{k}) & k'_b P_o(\mathbf{k}'|\mathbf{k}) \\ k'_o P_b(\mathbf{k}'|\mathbf{k}) & (k'_b - 1) P_b(\mathbf{k}'|\mathbf{k}) \end{pmatrix}. \quad (14)$$

The stability of the solution  $\hat{g}_o(1|\mathbf{k})=1$ ,  $\hat{g}_o^b(1|\mathbf{k})=1$  is thus determined by the maximum eigenvalue of the matrix  $\mathbf{C}_{\mathbf{k}\mathbf{k}'}^o$ ,  $\Lambda_m$ . When  $\Lambda_m \leq 1$  this solution is stable and the GIN component does not exist. In contrast, when  $\Lambda_m > 1$  a nontrivial solution of the set of equations (11) exists and the GIN component emerges.

The analysis for the in-component of individual vertices is identical to the case of the out one if we replace in Eqs. (10) and (11)

$$\begin{aligned} k_o &\rightarrow k_i, \\ g_o(s|\mathbf{k}) &\rightarrow g_i(s|\mathbf{k}), \\ g_o^b(s|\mathbf{k}) &\rightarrow g_i^b(s|\mathbf{k}), \\ P_o(\mathbf{k}'|\mathbf{k}) &\rightarrow P_i(\mathbf{k}'|\mathbf{k}). \end{aligned} \quad (15)$$

In this case, the matrix controlling the onset of the GOUT component is

$$\mathbf{C}_{\mathbf{k}\mathbf{k}'}^i \equiv \begin{pmatrix} k'_i P_i(\mathbf{k}'|\mathbf{k}) & k'_b P_i(\mathbf{k}'|\mathbf{k}) \\ k'_i P_b(\mathbf{k}'|\mathbf{k}) & (k'_b - 1) P_b(\mathbf{k}'|\mathbf{k}) \end{pmatrix}. \quad (16)$$

As before, the condition for the appearance of the GOUT component is ruled by the maximum eigenvalue of the matrix  $\mathbf{C}_{\mathbf{k}\mathbf{k}'}^i$ . At first glance, one could be tempted to conclude that, since the matrices  $\mathbf{C}_{\mathbf{k}\mathbf{k}'}^i$  and  $\mathbf{C}_{\mathbf{k}\mathbf{k}'}^o$  are different, their eigenvalues are also different, leading to different phase transitions for the appearance of the GIN and GOUT components. However, it can be proved that the eigenvalues spectra of both matrices are identical and, then, both the GIN and GOUT components appear simultaneously.

It is illustrative to recover from this formalism the results for the purely undirected and purely directed cases. In undirected networks, only bidirectional edges are present,  $k_i \equiv 0$  and  $k_o \equiv 0$ , and the matrices turn into

$$\begin{aligned} \mathbf{C}_{\mathbf{k}\mathbf{k}'}^o &\rightarrow C_{k_b k'_b} = (k'_b - 1) P(k'_b|k_b), \\ \mathbf{C}_{\mathbf{k}\mathbf{k}'}^i &\rightarrow C_{k_b k'_b} = (k'_b - 1) P(k'_b|k_b), \end{aligned} \quad (17)$$

recovering the results in [11]. In the case of purely directed networks,  $k_b \equiv 0$  and we obtain

$$\mathbf{C}_{\mathbf{k}\mathbf{k}'}^o \rightarrow C_{\mathbf{k}\mathbf{k}'}^o = k'_o P_o(\mathbf{k}'|\mathbf{k}),$$

$$\mathbf{C}_{\mathbf{k}\mathbf{k}'}^i \rightarrow C_{\mathbf{k}\mathbf{k}'}^i = k'_i P_i(\mathbf{k}'|\mathbf{k}). \quad (18)$$

This result generalizes the percolation theory for purely directed random networks developed in [5,6,20] to the case of arbitrary degree-degree correlations.

### B. Strongly connected component

The analysis of the GSCC requires a more careful development of the ideas introduced in the previous section. In this case, the joint distribution of the in- and out-components of a vertex of degree  $\mathbf{k}$  reads

$$\hat{G}(z, z'|\mathbf{k}) = z z' [\hat{g}_o(z|\mathbf{k})]^{k_o} [\hat{g}^b(z, z'|\mathbf{k})]^{k_b} [\hat{g}_i(z'|\mathbf{k})]^{k_i}, \quad (19)$$

where the function  $g^b(s, s'|\mathbf{k})$  is defined analogously to  $g_o^b(s|\mathbf{k})$  and  $g_i^b(s|\mathbf{k})$ . It is worth mentioning that, if the network contains bidirectional edges,  $G(s, s'|\mathbf{k}) \neq G_o(s|\mathbf{k})G_i(s'|\mathbf{k})$  because, in this case, such edges are common to the in- and out-components of the vertex. Using the same reasoning as in the previous section, we can write down a closed equation for the joint distribution  $g^b(s, s'|\mathbf{k})$ :

$$\begin{aligned} \hat{g}^b(z, z'|\mathbf{k}) &= z z' \sum_{\mathbf{k}'} P_b(\mathbf{k}'|\mathbf{k}) \\ &\times [\hat{g}_o(z|\mathbf{k}')]^{k'_o} [\hat{g}^b(z, z'|\mathbf{k}')]^{k'_b - 1} [\hat{g}_i(z'|\mathbf{k}')]^{k'_i}. \end{aligned} \quad (20)$$

This equation, together with Eq. (10) and Eqs. (11), is the complete solution of the problem. The solutions for arbitrary values of  $z$  and  $z'$  allow one to find the distribution of the sizes of the in- and out-components of single vertices whereas the nontrivial solution for  $z=1$  and  $z'=1$ , combined with Eqs. (5) and (6) and Eqs. (7) and (8), will provide us with a method to calculate the size of the different giant components of the network.

## IV. UNCORRELATED NETWORKS

As mentioned before, real networks are usually correlated in the sense that the degrees of pairs of connected vertices are correlated random quantities. Nevertheless, uncorrelated networks are equally useful as benchmarks, or null models, to test topological and dynamical properties and compare them to the results obtained in correlated networks.

When two-point correlations are absent, the transition probabilities become independent of the degree of the source vertex. In this situation, after some elemental algebra, we obtain

$$P_o(\mathbf{k}'|\mathbf{k}) = \frac{k'_i P(\mathbf{k}')}{\langle k_i \rangle}, \quad P_i(\mathbf{k}'|\mathbf{k}) = \frac{k'_o P(\mathbf{k}')}{\langle k_o \rangle} \quad (21)$$

and

$$P_b(\mathbf{k}'|\mathbf{k}) = \frac{k'_b P(\mathbf{k}')}{\langle k_b \rangle}, \quad (22)$$

where we have made use of the fact that  $\langle k_o \rangle = \langle k_i \rangle$ . Using these expressions, the set of equations (11) reduces to the following set of transcendent equations:



$$\begin{aligned}
y &= \langle k_i \rangle^{-1} \partial_x \hat{P}(1, y, z), & z &= \langle k_b \rangle^{-1} \partial_z \hat{P}(1, y, z), \\
x &= \langle k_i \rangle^{-1} \partial_y \hat{P}(x, 1, z'), & z' &= \langle k_b \rangle^{-1} \partial_z \hat{P}(x, 1, z'), \\
z'' &= \langle k_b \rangle^{-1} \partial_{z''} \hat{P}(x, y, z''), & &
\end{aligned} \tag{23}$$

where  $\hat{P}(x, y, z)$  is the generalized generating function of the degree distribution—that is,

$$\hat{P}(x, y, z) \equiv \sum_{k_i, k_o, k_b} x^{k_i} y^{k_o} z^{k_b} P(k_i, k_o, k_b). \tag{24}$$

The condition for the existence of a nontrivial solution of the set of equations (23) is easily obtained using the formalism developed in the previous sections. For the uncorrelated case, the matrices  $\mathbf{C}_{\mathbf{k}\mathbf{k}}^i$  and  $\mathbf{C}_{\mathbf{k}\mathbf{k}}^o$  become independent of  $\mathbf{k}$  and its maximum eigenvalue reads

$$\begin{aligned}
\Lambda_m &= \frac{1}{2} \left\{ \frac{\langle k_b(k_b - 1) \rangle}{\langle k_b \rangle} + \frac{\langle k_i k_o \rangle}{\langle k_i \rangle} \right. \\
&\quad \left. + \sqrt{\left( \frac{\langle k_b(k_b - 1) \rangle}{\langle k_b \rangle} - \frac{\langle k_i k_o \rangle}{\langle k_i \rangle} \right)^2 + \frac{4 \langle k_i k_b \rangle \langle k_o k_b \rangle}{\langle k_i \rangle \langle k_b \rangle}} \right\}, & (25)
\end{aligned}$$

so that, whenever the condition  $\Lambda_m \geq 1$  is fulfilled, the network is in the percolated phase. As can be seen from Eq. (25), the presence of bidirectional edges alters the point at which the giant component arises in a nontrivial way. The term  $\Gamma = \langle k_i k_b \rangle \langle k_o k_b \rangle / \langle k_i \rangle \langle k_b \rangle$  measures the strength of the coupling between directed and bidirectional edges. When this coupling is weak,  $\Gamma \sim 0$  and the maximum eigenvalue takes the simple form

$$\Lambda_m \sim \max \left\{ \frac{\langle k_b(k_b - 1) \rangle}{\langle k_b \rangle}, \frac{\langle k_i k_o \rangle}{\langle k_i \rangle} \right\}. \tag{26}$$

This result is easy to understand since, when  $\Gamma \sim 0$ , vertices cannot have directed and bidirectional edges simultaneously and, as a consequence, the network is composed of two isolated networks, one of them purely directed and the other one containing bidirectional edges only.<sup>2</sup>

In the purely undirected case,  $k_i \equiv 0$  and  $k_o \equiv 0$  and we recover the well-known condition for percolation in undirected networks with given degree distribution, Eq. (1):

$$\Lambda_m = \frac{\langle k_b(k_b - 1) \rangle}{\langle k_b \rangle} > 1. \tag{27}$$

In the case of purely directed networks,  $k_b \equiv 0$  and the maximum eigenvalue reads

$$\Lambda_m = \frac{\langle k_i k_o \rangle}{\langle k_i \rangle} > 1, \tag{28}$$

recovering Eq. (2).

<sup>2</sup>Notice that, in our context, uncoupled is not a synonym of uncorrelated. Indeed,  $\Gamma = 0$  corresponds to the strongest anticorrelation between  $k_i$  and  $k_b$  because when one of them is different from zero the other one is always zero.

When  $\Gamma \gg 1$ ,  $\Lambda_m$  must be computed using Eq. (25) and, in general, will depend on the density of edges, as well as on the type of correlations between directed and bidirectional edges. In particular, a positive correlation between  $k_b$  and  $k_i$  or  $k_o$  can strongly favor the emergence of the giant component even if the density of bidirectional edges is very small. We will illustrate this point in the example of the next section.

To finish the uncorrelated analysis, let us compute the relative sizes of the giant components. Let  $(x_c, y_c, z_c, z'_c, z''_c)$  be the nontrivial solution of the set of equations (23); then, using Eqs. (5) and (6), the relative sizes of the different giant components of the network read

$$g_{in} = 1 - \hat{P}(1, y_c, z_c), \quad g_{out} = 1 - \hat{P}(x_c, 1, z'_c), \tag{29}$$

$$g_{scc} = 1 - \hat{P}(1, y_c, z_c) - \hat{P}(x_c, 1, z'_c) + \hat{P}(x_c, y_c, z''_c). \tag{30}$$

This set of equations reduces to the results obtained in [6,20] if bidirectional edges are not present in the network. In the next section, we present a practical application of this formalism.

## V. BIDIRECTIONAL EDGES AS PERCOLATION CATALYSTS

Suppose we have a purely directed network with degree distribution  $P(k_i, k_o)$  and no two-point degree correlations. Suppose also that the network is in a regime in which the GWCC exists but not the GSCC—that is,

$$\frac{\langle k_i k_o \rangle}{\langle k_i \rangle} < 1, \quad \frac{\langle (k_i + k_o)(k_i + k_o - 1) \rangle}{\langle (k_i + k_o) \rangle} > 1. \tag{31}$$

Now we transform the original network by converting each directed edge into a bidirectional one with probability  $p$ . After this transformation, we end up with a network with  $pE$  bidirectional edges and  $(1-p)E$  directed ones, where  $E$  is the original number of directed edges on the network. The degree distribution of the transformed network can be written, in the discrete Laplace space, as

$$\hat{P}_p(z_i, z_o, z_b) = \hat{P}(pz_b + (1-p)z_i, pz_b + (1-p)z_o). \tag{32}$$

This transformation undergoes a phase transition as we increase the value of  $p$ . When  $p=0$ , the network is purely directed and, by construction, it has no GSCC. When  $p=1$ , all edges become bidirectional and, thus, the GSCC is identical to the GWCC of the original network. Therefore, at some intermediate value  $p_c$ , the network percolates and a GSCC emerges. The value of  $p_c$  can be easily obtained using the expression for the maximum eigenvalue, Eq. (25), and the final degree distribution, Eq. (32).

The most interesting case corresponds to networks with marginal degree distributions following power laws of the form  $P_i(k_i) \sim k_i^{-\gamma_i}$  and  $P_o(k_o) \sim k_o^{-\gamma_o}$  with  $\gamma_i, \gamma_o \leq 3$ . When the transformation described above is performed in this type of networks, some of the terms in Eq. (25) are proportional to  $\langle k_i^2 \rangle$  and  $\langle k_o^2 \rangle$  and, consequently,  $\Lambda_m \rightarrow \infty$  in the thermodynamic limit. This, in turn, implies that  $p_c = 0$ ; that is, even an

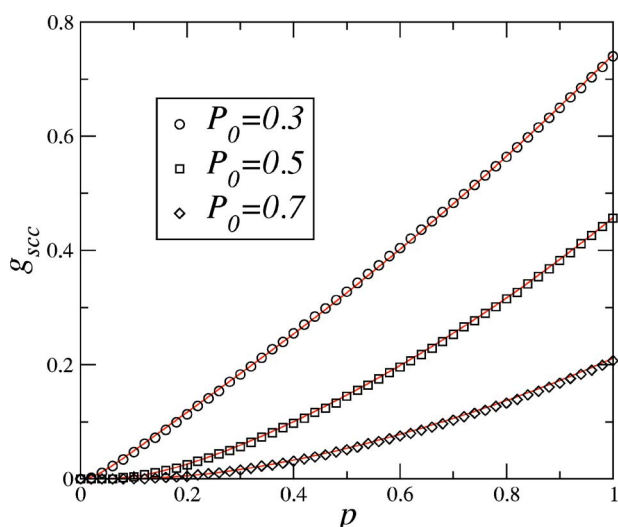


FIG. 2. (Color online) Relative size of the giant strongly connected component as a function of the conversion probability  $p$ . Simulation results are for a single network with  $N=10^6$  vertices. Solid lines are the numerical solution of the set of equations (23).

infinitely small fraction of bidirectional edges suffices to percolate the network.

#### A. Numerical simulation

To check the accuracy of our theory in the case of power-law marginal degree distributions, we have performed extensive numerical simulations. We first generate purely directed random networks with degree distribution of the form  $P(k_i, k_o) = P_i(k_i)P_o(k_o)$  and no two-point correlations. The in- and out-degree distributions are taken to be identical and to follow a scale-free form of the type

$$P_i(k) = P_o(k) = \begin{cases} P_0, & k = 0, \\ \frac{(1 - P_0)}{\zeta(\gamma)k^\gamma}, & k \geq 1, \end{cases} \quad (33)$$

where  $\zeta(\gamma)$  is the zeta Riemann function. To generate a purely directed random network, we use a natural extension of the configuration model [7,8,24,25]—an algorithm intended to generate uncorrelated random networks with a given degree distribution. The algorithm starts by first assigning to a set of  $N$  vertices a pair of “stubs,” one of them incoming and the other one outgoing,  $k_i$  and  $k_o$ , randomly drawn from the distribution  $P(k_i, k_o)$ . The only requirement is that  $\sum_i k_i = \sum_o k_o$ , whenever one wishes to close the network. The network is constructed by selecting pairs of in- and out-stubs chosen uniformly at random to create directed edges, avoiding multiple, bidirected, and self-connections among vertices. Once the network has been assembled, each directed edge is transformed into a bidirectional one with probability  $p$ , and the relative sizes of the giant components are measured.

Figures 2 and 3 show simulation results for a scale-free network following Eq. (33) with exponent  $\gamma=3$  and size  $N = 10^6$  as compared to the numerical solution of Eqs. (23). As can be seen, the agreement between simulation results and

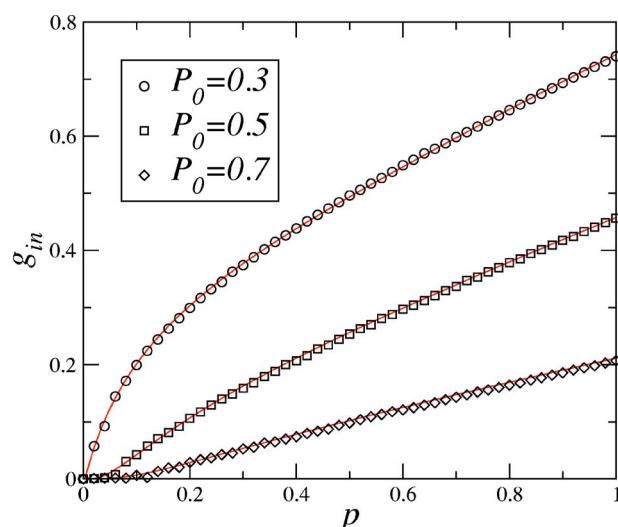


FIG. 3. (Color online) Relative size of the giant in component as a function of the conversion probability  $p$ . Simulation results are for a single network with  $N=10^6$  vertices. Solid lines are the numerical solution of the set of equations (23).

the theoretical prediction is excellent. The relative sizes of the GSCC and GIN are shown as a function of the conversion probability  $p$  for different values of the distribution parameter  $P_0$ , the probability of nodes having null in- or out-degree. Even for very small values of  $p$ , the GSCC and GIN are evident. As expected, small values of  $P_0$  favor the growth of bigger giant components. Then, bidirectional edges act as a percolation catalysts, favoring the appearance of a fine structure in the giant connected component. The scale-free property of many real networks is, once more, indicative of interesting features, since, in this case, the presence of an infinitesimal fraction of bidirectional edges is enough to ensure percolation at the level of the directed components.

## VI. CONCLUSIONS

We have derived a very general formulation of the theory of percolation in directed random networks with bidirectional edges and arbitrary two point degree correlations. Our formalism accounts for all the previously known results for percolation in purely directed and purely undirected random networks, which stand as limiting cases of our theory. The percolation threshold for the most general situation is derived as a function of the maximum eigenvalue of the connectivity matrices. In particular, for networks with no two-point correlations, explicit expressions are provided depending on the first and second moments of the degree distribution  $P(\mathbf{k})$ . In this case, we have also shown that bidirectional edges act as a catalyst for percolation, favoring the emergence of the GSCC, and for scale-free networks, only an infinitesimal fraction of bidirectional edges is needed.

After the completion of this work, we have become aware of a recent preprint [26] where the classical susceptible-infected-recovered (SIR) model of epidemiology is analyzed in uncorrelated directed random networks with bidirectional edges. Since there exists a mapping between the SIR model and percolation theory, some of the results derived in that reference overlap our results for the uncorrelated case.

#### ACKNOWLEDGMENTS

We would like to thank R. Pastor-Satorras for valuable comments. This work has been partially supported by DGES of the Spanish government, Grant No. FIS2004-05923-CO2-02, and EC-FET Open Project No. COSIN IST-2001-33555. M.B. acknowledges financial support from the MCyT (Spain).

- 
- [1] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From biological nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
  - [2] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
  - [3] S. N. Dorogovtsev and J. F. F. Mendes, *Adv. Phys.* **51**, 1079 (2002).
  - [4] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, England, 2004).
  - [5] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. Lett.* **85**, 5468 (2000).
  - [6] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001).
  - [7] M. Molloy and B. Reed, *Random Struct. Algorithms* **6**, 161 (1995).
  - [8] M. Molloy and B. Reed, *Combinatorics, Probab. Comput.* **7**, 295 (1998).
  - [9] M. E. J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
  - [10] M. E. J. Newman, *Phys. Rev. E* **67**, 026126 (2003).
  - [11] A. Vázquez and Y. Moreno, *Phys. Rev. E* **67**, 015101(R) (2003).
  - [12] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, *Phys. Rev. E* **64**, 066110 (2001).
  - [13] P. L. Krapivsky and B. Derrida, *Physica A* **340**, 714 (2004).
  - [14] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, *Comput. Netw.* **33**, 309 (2000).
  - [15] D. J. de S. Price, *Science* **149**, 510 (1965).
  - [16] S. Redner, *Eur. Phys. J. B* **4**, 131 (1998).
  - [17] M. E. J. Newman, S. Forrest, and J. Balthrop, *Phys. Rev. E* **66**, 035101 (2002).
  - [18] H. De Jong, *J. Comput. Biol.* **9**, 67 (2002).
  - [19] D. A. Fell and A. Wagner, *Nat. Biotechnol.* **18**, 1121 (2000).
  - [20] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, *Phys. Rev. E* **64**, 025101(R) (2001).
  - [21] N. Schwartz, R. Cohen, D. ben-Avraham, A.-L. Barabási, and S. Havlin, *Phys. Rev. E* **66**, 015104 (2002).
  - [22] D. Garlaschelli and M. I. Loffredo, *Phys. Rev. Lett.* **93**, 268701 (2004).
  - [23] M. Boguñá and R. Pastor-Satorras, *Phys. Rev. E* **66**, 047104 (2002).
  - [24] A. Bekessy, P. Bekessy, and J. Komlos, *Stud. Sci. Math. Hung.* **7**, 343 (1972).
  - [25] E. A. Bender and E. R. Candfield, *J. Comb. Theory, Ser. A* **24**, 296 (1978).
  - [26] L. A. Meyers, M. E. J. Newman, and B. Pourbohloul, e-print sfi/0412037.