

## Class of correlated random networks with hidden variables

Marián Boguñá<sup>1</sup> and Romualdo Pastor-Satorras<sup>2</sup>

<sup>1</sup>*Departament de Física Fonamental, Universitat de Barcelona, Avinguda Diagonal 647, 08028 Barcelona, Spain*

<sup>2</sup>*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord, Mòdul B4, 08034 Barcelona, Spain*

(Received 3 June 2003; published 15 September 2003)

We study a class of models of correlated random networks in which vertices are characterized by *hidden variables* controlling the establishment of edges between pairs of vertices. We find analytical expressions for the main topological properties of these models as a function of the distribution of hidden variables and the probability of connecting vertices. The expressions obtained are checked by means of numerical simulations in a particular example. The general model is extended to describe a practical algorithm to generate random networks with an *a priori* specified correlation structure. We also present an extension of the class, to map nonequilibrium growing networks to networks with hidden variables that represent the time at which each vertex was introduced in the system.

DOI: 10.1103/PhysRevE.68.036112

PACS number(s): 89.75.-k, 87.23.Ge, 05.70.Ln

### I. INTRODUCTION

A large effort has been recently devoted to the study of a very large ensemble of interacting systems that can be described in terms of complex networks (or graphs), in which the vertices represent typical units and the edges represent the interactions between pairs of units [1–3]. Stimulated by this finding, a theory of complex networks, deeply rooted in the classical graph theory [4], has hence been developed, finding fruitful applications in fields as diverse as the Internet [5–8], the WorldWideWeb [9], social communities [10], food webs [11], or biological interacting networks [12–15].

The study of complex networks, boosted by the new availability of powerful computers capable of dealing with very large databases, was initially focused in the study of global properties such as the average shortest path length, the average clustering coefficient, or the degree distribution [1–3]. This work led to the discovery that most natural complex networks usually exhibit two typical properties: (i) the *small-world* property [16], which is defined by an average path length—average distance between any pair of vertices—increasing very slowly (usually logarithmically) with the network size  $N$  and (ii) a *scale-free* degree distribution. If we define the degree distribution  $P(k)$  as the probability that a vertex is connected to  $k$  other vertices, then scale-free networks are characterized by a power-law behavior  $P(k) \sim k^{-\gamma}$ , where  $\gamma$  is a characteristic degree exponent. These properties imply a large connectivity heterogeneity and a short average distance between vertices, which have considerable impact on the behavior of physical processes taking place on top of the network, such as the resilience to random damage [17–19] or the spreading of infective agents [20–23].

It was soon realized, however, that these properties do not provide a sufficient characterization of natural networks. In particular, these systems seem to exhibit also ubiquitous degree *correlations*, which translate in the fact that the degrees of the vertices at the end points of any given edge are not independent [7,8,24,25]. This observation has led to a classification of networks according to the nature of their degree correlations [24]: In the presence of positive correlations

(vertices with large degree tend to connect more preferably with vertices with large degree), the network is said to show *assortative mixing*. On the other hand, negative correlations (highly connected vertices are preferably connected to vertices with low degree) imply the presence of *dissortative mixing*. At the same time, it has been pointed out that the presence of correlations might have important consequences in dynamical processes taking place in the topology defined by the network [26–29]. Motivated by these observations, several works have been recently devoted to set up a general framework to study the origin of correlations in random networks [30,31]. At this respect, it is particularly interesting the models introduced by Caldarelli *et al.* [32] and Söderberg [33]. These models consider graphs in which each vertex has assigned a tag (*type* or *fitness*), randomly drawn from a fixed probability distribution. Edges are assigned to pairs of vertices with a given connection probability, depending on the values of the tags assigned at the edge end points. This construction generates random networks that exhibit peculiar correlation and percolation properties [32,33].

In this paper we present a generalization of the models described in Refs. [32,33], which can be encompassed in a general class of models with *hidden variables* tagging the vertices, and completely determining the topological structure of the ensuing network. We develop a detailed analysis of the correlations present in this class of network models, providing explicit analytical expressions for both two and three vertices degree correlations. We distinguish between sparse networks (with finite average degree  $\langle k \rangle$ ) and non-sparse networks (with diverging  $\langle k \rangle$  for a number of vertices  $N \rightarrow \infty$ ). Even though both cases are enclosed in this class of networks, analytical expressions are simpler in the former case. As an example of our formalism, we consider the intrinsic fitness model introduced in Ref. [32], which belongs to the subset of nonsparse networks, and which has attracted a great deal of attention as an alternative to generate scale-free networks without growth nor preferential attachment [34]. The solution of this model in the continuous degree approximation is compared with extensive numerical simulations, yielding a remarkable agreement for all the topological properties considered. As a particular case of the general

class of models with hidden variables, we propose a practical algorithm to generate correlated random networks with a given correlation structure. The algorithm leverages in the assignment of hidden variables with the structure of the degrees of a real network. Following this approach, it is possible to easily generate networks matching any desired correlation pattern, as we show by means of analytical calculations and numerical simulations. Finally, we present the extension of this class of models to nonequilibrium growing networks. By mapping the hidden variables to the time in which vertices are introduced in the network [33], and by means of an appropriately chosen connection probability, we define an algorithm that yields networks exhibiting all the properties (in particular aging) exhibited by traditional scale-free growing models.

The paper is organized as follows. In Sec. II we review some general results concerning the measure of correlations in complex networks, which will be useful through the rest of the paper. In Sec. III we introduce the general analytical formulation of the class of correlated networks with hidden variables. Section IV is devoted to the analytical and numerical study of the intrinsic fitness model introduced in Ref. [32]. In Sec. V we present an algorithm to generate correlated random networks with a given *a priori* correlation structure. Sec. VI deals with the mapping into this class of models of nonequilibrium growing networks. Finally, in Sec. VII we draw our conclusions and perspectives.

## II. MEASURING CORRELATIONS IN COMPLEX NETWORKS

### A. Two vertices correlations

Let us consider the class of unstructured undirected networks, in which all vertices with the same degree can be considered to be statistically equivalent. In this sense, the following results will not apply to structured networks, in which a distance ordering can be defined; for instance, when the small-world property is absent [27,35]. A network is said to be *uncorrelated* when the probability that an edge departing from a vertex of degree  $k$  arrives at a vertex of degree  $k'$  is independent of the degree of the initial vertex  $k$ . Most natural networks are not uncorrelated, in the sense that the degrees at the end points of any given edge are not independent. This kind of two vertices degree correlation can be measured in undirected networks by means of the conditional probability  $P(k'|k)$  that a vertex of degree  $k$  is connected to a vertex of degree  $k'$ . From the point of view of correlations, it is useful to consider the restricted subset of undirected *Markovian* random networks [26], which are completely defined by the degree distribution  $P(k)$  and the conditional probability  $P(k'|k)$ . The Markovian nature of this class of networks implies that all higher order correlations can be expressed as a function of  $P(k'|k)$ .

The functions  $P(k)$  and  $P(k'|k)$  are assumed to be normalized, i.e.,

$$\sum_k P(k) = \sum_{k'} P(k'|k) = 1, \quad (1)$$

and they are constrained by a degree detailed balance condition [26] stating the physical conservation of edges among vertices: The total number of edges pointing from vertices with degree  $k$  to vertices with degree  $k'$  must be equal to the number of edges that point from vertices  $k'$  to vertices  $k$ . There is an intuitive way to derive the degree detailed balance condition [36]. Let us denote by  $N_k$  the number of vertices of degree  $k$ . Since  $\sum_k N_k = N$ , where  $N$  is the size of the network, we can define the degree distribution as  $P(k) = N_k/N$  [37]. To completely define the network, we need to specify also how the different degree classes are connected. To this end, let us define the symmetric matrix  $E_{kk'}$  that gives the number of edges between vertices of degree  $k$  and  $k'$ , for  $k \neq k'$ , and two times the number of self-connections for  $k = k'$  (the number of connections between vertices in the same degree class). This matrix fulfills the identities

$$\sum_{k'} E_{kk'} = kN_k, \quad (2)$$

$$\sum_{k,k'} E_{kk'} = \langle k \rangle N = 2E, \quad (3)$$

where  $E$  is the total number of edges in the network. This last identity allows us to define the *joint distribution*

$$P(k,k') = \frac{E_{kk'}}{\langle k \rangle N}, \quad (4)$$

where the symmetric function  $(2 - \delta_{k,k'})P(k,k')$  is the probability that a randomly chosen edge connects two vertices of degrees  $k$  and  $k'$ . The conditional probability  $P(k'|k)$  defined as the probability that an edge from a  $k$  vertex points to a  $k'$  vertex can be easily written as

$$P(k'|k) = \frac{E_{k'k}}{kN_k} = \frac{\langle k \rangle P(k,k')}{kP(k)}. \quad (5)$$

From the symmetry of  $P(k,k')$  it follows immediately the degree detailed balance condition

$$kP(k'|k)P(k) = k'P(k|k')P(k') = \langle k \rangle P(k,k'). \quad (6)$$

The joint distribution  $P(k,k')$  conveys all the information needed to construct a Markovian random network. In fact, it is easy to see that

$$P(k) = \frac{\langle k \rangle}{k} \sum_{k'} P(k,k'). \quad (7)$$

This relation, together with Eq. (5), completely defines the network properties, i.e.,  $P(k)$  and  $P(k'|k)$ , as a function of the joint distribution  $P(k,k')$ . Notice that Eqs. (5) and (7) define the degree distribution and the conditional probability in the whole  $k$  range, except for  $k=0$ . This fact does not represent a problem, however, since vertices without edges are usually not considered in natural complex networks.

The empirical evaluation of  $P(k,k')$  [or  $P(k'|k)$ ] is, in most real networks, a quite difficult task, since the available

data, restricted to finite sizes, usually yield results extremely noisy and difficult to interpret. For this reason, it is more useful for practical purposes to analyze instead the average degree of the nearest neighbors (ANND) as a function of the vertex degree, defined by [7]

$$\bar{k}_{nn}(k) = \sum_{k'} k' P(k'|k). \quad (8)$$

For uncorrelated networks, in which  $P(k'|k)$  does not depend on  $k$ , application of the normalization condition (1) into Eq. (6) yields  $P_0(k'|k) = k' P(k')/\langle k \rangle$ . In this case, we obtain  $\bar{k}_{nn}^0(k) = \langle k^2 \rangle / \langle k \rangle$ , independent of  $k$ . Therefore, a function  $\bar{k}_{nn}(k)$  with an explicit dependence on  $k$  signals the presence of degree correlations in the network. Based on the ANND, it is possible to characterize the correlation properties of the network [24]: When  $\bar{k}_{nn}(k)$  is an increasing function of  $k$ , the network shows assortative mixing. Examples of assortative behavior can be found in several social networks [24]. On the other hand, when  $\bar{k}_{nn}(k)$  is a decreasing function of  $k$ , the network shows disassortative mixing, as found, for example, in technological systems such as the Internet [7].

### B. Three vertices correlations

Correlations among three vertices can be measured by means of the probability  $P(k', k''|k)$  that a vertex of degree  $k$  is simultaneously connected to two vertices with degrees  $k'$  and  $k''$ . In the particular case of Markovian networks, this function is related to the two vertices correlation through  $P(k', k''|k) = P(k'|k)P(k''|k)$ . For non-Markovian networks, however, the functions  $P(k', k''|k)$  and  $P(k'|k)$  are in principle not related.

Information about three vertices correlations can be obtained from the clustering coefficient. The concept of clustering in a graph refers to the tendency to form cliques (complete subgraphs [4]) in the neighborhood of any given vertex. In this sense, clustering implies that if vertex  $i$  is connected to vertex  $j$ , and at the same time  $j$  is connected to  $l$ , then, with high probability,  $i$  is also connected to  $l$ . The probability that two vertices with a common neighbor are also connected to each other is called the *clustering coefficient* of the common vertex [16]. Numerically, the clustering coefficient  $c_i$  of vertex  $i$  can be computed as the ratio between the number of edges existing between  $k_i$  neighbors of  $i$ ,  $e_i$ , and its maximum possible value  $k_i(k_i-1)/2$ , that is,

$$c_i = \frac{2e_i}{k_i(k_i-1)}. \quad (9)$$

On the other hand, the clustering coefficient of a vertex of degree  $k$ ,  $\bar{c}(k)$  [8], can be formally computed as the probability that it is connected to vertices  $k'$  and  $k''$ , and that those two vertices are, on their turn, joined by an edge, averaged over all the possible values of the degrees of the neighbor vertices. Therefore, we can write  $\bar{c}(k)$  as a function of the three vertices correlations as

$$\bar{c}(k) = \sum_{k', k''} P(k', k''|k) p_{k', k''}, \quad (10)$$

where the function  $p_{k', k''}$  is the probability that the vertices  $k'$  and  $k''$  are connected [38]. The quantity  $\bar{c}(k)$  has been recently used to study the level of hierarchy and modularity in real complex networks [39].

### III. HIDDEN VARIABLE MODELS OF CORRELATED NETWORKS

Recently, Caldarelli *et al.* [32] and Söderberg [33,40] (see also Refs. [25,41]) have proposed different models of inhomogeneous random graphs that represents a natural generalization of the classical Erdős-Rényi random graph model [42,43]. These models consider inhomogeneous graphs in which each vertex is characterized by a different *type* or *fitness*. Types can be either discrete or continuous variables and are assigned to vertices according to a certain probability distribution. Then, pairs of vertices are independently joined by an undirected edge with a probability depending on the type of the respective end points. This construction leads to an ensemble of undirected random networks, which inherits the simplicity of the Erdős-Rényi model while allowing freedom for general forms of the degree distribution and correlation structure. References [33,40] were mainly concerned with the component distribution and the onset of the giant component in these kinds of models, and Ref. [32] reported numerical simulations for different model parameters, and analytical arguments for the form of the degree distribution.

The models defined in Refs. [32,33] can be generalized as a class of models with *hidden variables*. The hidden variables play the role of tags assigned to the vertices, and they completely determine the topological properties of the network through their probability distribution and the probability to connect pairs of vertices.

We define the class of models with hidden variables as follows. Let us consider a set of  $N$  disconnected vertices and a general hidden variable  $h$ , which can be a natural or a real number. An undirected graph is generated by the following two rules.

- (1) Each vertex  $i$  is assigned a variable  $h_i$ , independently drawn from the probability distribution  $\rho(h)$ .
- (2) For each pair of vertices  $i$  and  $j$ , with respective hidden variables  $h_i$  and  $h_j$ , an undirected edge is created with probability  $r(h_i, h_j)$  (the *connection probability*), where  $r(h, h') \geq 0$  is a symmetric function of  $h$  and  $h'$ .

Given the independent assignment of hidden variables and edges among vertices, this procedure generates correlated random networks with neither loops nor multiple edges, which are Markovian at the hidden variable level [44] and whose degree distribution and correlation properties are encoded in the two functions  $\rho(h)$  and  $r(h, h')$ . Here we will focus in the case in which the distribution  $\rho(h)$  is *independent* of the network size  $N$ . The case in which  $\rho(h)$  is allowed to depend on  $N$  will be considered in Sec. VI.

In this section we will provide analytic expressions for the correlation function and clustering coefficient of the networks generated with this class of models as a function of

the distribution of hidden variables and the probability to connect pairs of vertices.

### A. Degree distribution

The degree distribution  $P(k)$  is defined as the probability that any given vertex has  $k$  edges attached to it. Therefore, in order to compute it, we need to know the conditional probability  $g(k|h)$  (propagator) that a vertex with initial hidden variable  $h$  ends up connected to other  $k$  vertices. The degree distribution can then be written as

$$P(k) = \sum_h g(k|h)\rho(h), \quad (11)$$

where the summation sign must be exchanged by an integral for continuous  $h$ . The propagator, which is obviously normalized,  $\sum_k g(k|h) = 1$ , provides full information about the dependence of the actual degree  $k$  on the hidden variable  $h$ . In particular, we can see that the average degree of the vertices with hidden variable  $h$ ,  $\bar{k}(h)$ , is given by

$$\bar{k}(h) = \sum_k k g(k|h), \quad (12)$$

and the average degree can be expressed as

$$\langle k \rangle = \sum_k k P(k) = \sum_h \bar{k}(h)\rho(h). \quad (13)$$

On the other hand, the probability that a vertex of actual degree  $k$  has associated a hidden variable  $h$ ,  $g^*(h|k)$ , can be computed as the inverse of the propagator by means of Bayes' formula [45],

$$P(k)g^*(h|k) = \rho(h)g(k|h). \quad (14)$$

In order to get an explicit expression for the propagator, we start by noticing that it can be written as

$$g(k|h) = \sum_{k_1, \dots, k_c} g_1^{(h)}(k_1|h_1)g_2^{(h)}(k_2|h_2) \cdots g_c^{(h)}(k_c|h_c) \times \delta_{k_1+k_2+\dots+k_c, k}, \quad (15)$$

where  $g_i^{(h)}(k_i|h_i)$  is the probability that a vertex with hidden variable  $h$  ends up with  $k_i$  connections with vertices of hidden variable  $h_i$ ,  $h_c$  being the maximum value of  $h$ . Since the connections between vertices with hidden variables  $h$  and  $h'$  are independently drawn with probability  $r(h, h')$ , the probability  $g_i^{(h)}(k_i|h_i)$  is simply given by a binomial distribution, i.e.,

$$g_i^{(h)}(k_i|h_i) = \binom{N_i}{k_i} r(h, h_i)^{k_i} [1 - r(h, h_i)]^{N_i - k_i}, \quad (16)$$

where  $N_i = N\rho(h_i)$  is the number of vertices with hidden variable  $h_i$ . Let us define now the generating function [46]

$$\hat{g}(z|h) = \sum_k z^k g(k|h). \quad (17)$$

Since the propagator is given by a convolution, Eq. (15), we can write its generating function as the product of the generating functions of the partial propagators  $g_i^{(h)}(k_i|h_i)$ , which on their turn, being binomial distributions, yield

$$\hat{g}_i^{(h)}(z|h_i) = [1 - (1-z)r(h, h_i)]^{N_i}. \quad (18)$$

Inserting this expression into the definition of  $\hat{g}(z|h)$  and taking logarithms on both sides we are led to the equation

$$\ln \hat{g}(z|h) = N \sum_{h'} \rho(h') \ln [1 - (1-z)r(h, h')]. \quad (19)$$

For general probabilities  $\rho(h)$  and  $r(h, h')$ , Eq. (19) must be solved and inverted in order to obtain the corresponding propagator. The degree distribution is then obtained applying Eq. (11). Even without solving the previous equation, however, it is already possible to obtain some information on the connectivity properties of the network. From the definition, Eq. (17), the first moment of  $g(k|h)$  is given by the first derivative of  $\hat{g}(z|h)$  evaluated at  $z=1$ . Therefore we have

$$\bar{k}(h) = N \sum_{h'} \rho(h') r(h, h'), \quad (20)$$

and

$$\langle k \rangle = N \sum_{h, h'} \rho(h) r(h, h') \rho(h'), \quad (21)$$

where we have used Eq. (19) in computing these expressions.

At this point we must consider the possibility of two different kinds of networks: Sparse networks, with a well-defined thermodynamic limit for the average degree  $\langle k \rangle$ , and nonsparse networks, in which the average degree diverges with the network size. In the case of sparse networks, the number of edges grows linearly with the system size and, therefore, the joint distribution is a well-defined quantity, independent of  $N$ . In the opposite case, nonsparse networks have a number of edges growing faster than linearly, which causes the breakdown of the thermodynamic limit and the emergence of the phenomenon of condensation of edges (see Sec. IV). In order to distinguish between sparse and nonsparse networks we must consider the value of the average degree, given by Eq. (21). If the density  $\rho(h)$  is independent of the size of the system, the only possibility to have a sparse network is that the connection probability scales as  $N^{-1}$ . This scaling behavior turns out to have a strong implication in the form of the propagator. Defining  $r(h, h') \equiv C(h, h')/N$  (as considered in Ref. [33]), where  $C(h, h')$  is a bounded symmetric function, independent of  $N$ , we can expand the right-hand side of Eq. (19) in the limit  $N \rightarrow \infty$  to obtain

$$\hat{g}(z|h) = \exp\left\{ (z-1) \sum_{h'} \rho(h') C(h, h') \right\}. \quad (22)$$

The generating function of the propagator is a pure exponential, which indicates that the propagator itself is a Poisson distribution,

$$g(k|h) = \frac{e^{-\bar{k}(h)} \bar{k}(h)^k}{k!}, \quad (23)$$

where in this case  $\bar{k}(h) = \sum_{h'} \rho(h') C(h, h')$ . Equation (23) is, indeed, a strong result, since it states the universality of the propagator for sparse networks regardless of the form of the connection probability.

It is worth mentioning that the degree distribution, Eq. (19), has been derived on the basis of a microcanonical ensemble, in the sense that the number of vertices of each class,  $N_h = N \rho(h)$ , is a fixed quantity. This is in contrast to the canonical ensemble in which the fixed quantity is the average number of vertices of each class. However, given the equivalence between ensembles, both approaches are equivalent in the thermodynamic limit [47].

### B. Degree correlations

Degree correlations are completely characterized by means of the conditional probability  $P(k'|k)$ , which gives the probability that an edge emanating from a vertex of degree  $k$  is connected to a vertex of degree  $k'$ . In order to construct the function  $P(k'|k)$  we consider a vertex of degree  $k$ , which with probability  $g^*(h|k)$  has associated a hidden variable  $h$ . Let us define  $p(h'|h)$  the conditional probability that a  $h$  vertex is connected to a  $h'$  vertex. Then, the conditional probability  $P(k'|k)$  can be written as

$$P(k'|k) = \sum_{h, h'} g(k'-1|h') p(h'|h) g^*(h|k), \quad (24)$$

where the propagator  $g(k'-1|h')$  gives the probability that the  $h'$  vertex ends up with degree  $k'$  (since one connection has already been used up for the conditional edge with  $h$ ). Using the form of  $g^*(h|k)$  given by Eq. (14) we have

$$P(k'|k) = \frac{1}{P(k)} \sum_{h, h'} g(k'-1|h') p(h'|h) \rho(h) g(k|h), \quad (25)$$

valid for  $k, k' = 1, 2, \dots$ . In order to close Eq. (25), we need, finally, to provide an expression for the conditional probability  $p(h'|h)$ . In order to do so, we consider that the probability of drawing an edge from  $h$  to  $h'$  is proportional to the probability of finding an  $h'$  vertex, times the probability of creating the actual edge. Taking into account normalization, we have that

$$p(h'|h) = \frac{\rho(h') r(h, h')}{\sum_{h''} \rho(h'') r(h, h'')} = \frac{N \rho(h') r(h, h')}{\bar{k}(h)}. \quad (26)$$

Finally, using Eqs. (25) and (26) we can compute the ANND as

$$\bar{k}_{nn}(k) = 1 + \frac{1}{P(k)} \sum_h g(k|h) \rho(h) \bar{k}_{nn}(h), \quad (27)$$

where we have defined the ANND of a  $h$  vertex [see Eq. (8)] as

$$\bar{k}_{nn}(h) \equiv \sum_{h'} \bar{k}(h') p(h'|h). \quad (28)$$

### C. Clustering coefficient

The clustering coefficient is defined as the probability that two vertices, adjacent to a third vertex, are also connected to each other. In the space of hidden variables, consider a  $h$  vertex, which is connected to two other vertices  $h'$  and  $h''$  with probability  $p(h', h''|h)$ . On the other hand,  $h'$  and  $h''$  are connected with probability  $r(h', h'')$ . Therefore, the clustering coefficient of a vertex  $h$  is given by  $c_h = \sum_{h', h''} p(h', h''|h) r(h', h'')$ . Note that this is the natural counterpart of Eq. (10) in the space of hidden variables. Now, since the network is Markovian at the hidden variable level, we have that  $p(h', h''|h) = p(h'|h) p(h''|h)$ . Thus we have that

$$c_h = \sum_{h', h''} p(h'|h) r(h', h'') p(h''|h). \quad (29)$$

The clustering coefficient of the vertices of degree  $k$ ,  $\bar{c}(k)$ , will be given by the probability that a vertex  $k$  has hidden variable  $h$ ,  $g^*(h|k)$ , times  $c_h$ , averaged over all the possible values of  $h$ . Thus

$$\bar{c}(k) = \frac{1}{P(k)} \sum_h \rho(h) g(k|h) c_h, \quad k = 2, 3, \dots, \quad (30)$$

where we have used the form of  $g^*(h|k)$  given by Eq. (14).

The results derived in this section represent the general solution of the class of networks with hidden variables. In the rest of the paper we will show how this formalism is able to deal with a wide variety of models, from sparse to non-sparse networks, and from equilibrium to nonequilibrium ones.

## IV. THE INTRINSIC FITNESS MODEL

As an example of the general class of models with hidden variables, Caldarelli *et al.* [32] considered the model defined by the probability distributions

$$\rho(h) = e^{-h} \quad \text{for } h \in [0, \infty[, \quad (31)$$

$$r(h, h') = \theta(h + h' - \zeta), \quad (32)$$

where  $\theta(x)$  is the Heaviside step function and  $\zeta$  is a constant. In this model, hereafter referred to as the intrinsic fitness (IF) model, vertices have assigned an exponentially distributed hidden variable (fitness), and are joined by an edge

whenever the sum of the fitness of the end points is larger than a given threshold  $\zeta$ . By means of numerical simulations and analytical arguments, Caldarelli *et al.* [32] showed that the degree distribution in this model is power-law distributed. This observation led to the very interesting conclusion that it is possible to generate scale-free networks without growth not preferential attachment [34].

### A. Analytic solution

Using the general formalism developed in the preceding section, we can provide analytic expressions for the main properties of the IF model. In order to do so, let us compute in the first place the propagator  $g(k|h)$ . Inserting Eqs. (31) and (32) into Eq. (19), we have, substituting the summation by an integral,

$$\begin{aligned} \ln \hat{g}(z|h) &= N \int_0^\infty dh' e^{-h'} \ln[1 - (1-z)\theta(h+h'-\zeta)] \\ &= N \ln z \begin{cases} e^{h-\zeta} & \text{if } 0 \leq h \leq \zeta \\ 1 & \text{if } h > \zeta, \end{cases} \end{aligned} \quad (33)$$

from where we obtain  $\hat{g}(z|h) = z^{Ne^{h-\zeta}}$  for  $0 \leq h \leq \zeta$ , and  $\hat{g}(z|h) = z^N$  for  $h > \zeta$ . In order to invert this generating function, we approximate  $k$  by a continuous variable. In this case, the propagator takes the simple form

$$g(k|h) = \delta(k - Ne^{h-\zeta})\theta_h(0, \zeta) + \delta(k - N)\theta(h - \zeta), \quad (34)$$

where  $\delta(x)$  is the Dirac delta function and we have introduced the window function

$$\theta_x(a, b) = \begin{cases} 1 & \text{for } a \leq x \leq b \\ 0 & \text{otherwise.} \end{cases} \quad (35)$$

This approximation is expected to perform poorly for small values of  $k$ , as we will see when comparing the analytical results with computer simulations of the IF model.

Inserting the propagator, Eq. (34), into the general expression (11), and performing the integrals corresponding to the Dirac  $\delta$  functions, we obtain the degree distribution

$$P(k) = Ne^{-\zeta} \frac{1}{k^2} \theta_k(Ne^{-\zeta}, N) + e^{-\zeta} \delta(k - N). \quad (36)$$

That is, the networks generated by the IF model exhibit a scale-free degree distribution, with degree exponent  $\gamma=2$ , for degrees in the range  $Ne^{-\zeta} \leq k \leq N$ , plus an accumulation point at  $k=N$ , given by the  $\delta$  function, with weight  $e^{-\zeta}$ . This accumulation point signals the presence of a *condensation* of edges in the fraction  $e^{-\zeta}$  of the vertices of the network with  $h > \zeta$ , which establish connections to all the other vertices [48]. This condensation, reminiscent to that observed in models with nonlinear preferential attachment [49], is the result of the nonsparse nature of the network, which, from Eq. (21), has average degree  $\langle k \rangle = Ne^{-\zeta}(\zeta + 1)$ .

In order to characterize the correlations of the model, we compute the ANND, given by Eq. (27). In the continuous  $k$  approximation, the function  $\bar{k}(h)$  takes the form

$$\bar{k}(h) = Ne^{h-\zeta} \theta_h(0, \zeta) + N\theta(h - \zeta). \quad (37)$$

Inserting this expression into the formula for the ANND, we obtain

$$\begin{aligned} \bar{k}_{nn}(k) &= 1 + \frac{Ne^{-2\zeta}}{P(k)} \left[ (1 + \zeta) \delta(k - N) \right. \\ &\quad \left. + \frac{N^2}{k^3} \left\{ 1 + \zeta + \ln\left(\frac{k}{N}\right) \right\} \theta_k(Ne^{-\zeta}, N) \right]. \end{aligned} \quad (38)$$

The regular part of this expression (discarding the  $\delta$  function singularities, signaling again the effect of the condensation of edges in the correlation function) takes the form

$$\bar{k}_{nn}^r(k) = 1 + \frac{N^2 e^{-\zeta}}{k} \left[ 1 + \zeta + \ln\left(\frac{k}{N}\right) \right] \theta_k(Ne^{-\zeta}, N). \quad (39)$$

That is, the regular part of the ANND is proportional to  $k^{-1}$ , times a logarithmic correction term. We are therefore in the presence of disassortative mixing. Note that, in the limit  $N \rightarrow \infty$ , we have that  $\bar{k}_{nn}^r(k) \rightarrow \infty$ , in agreement with the theoretical prediction made in Ref. [29].

Finally, to estimate the clustering coefficient, we have to compute first the conditional probability at the level of hidden variables, given by Eq. (26). Using Eqs. (31) and (32), we obtain

$$p(h'|h) = e^{-h'} \theta(h' + h - \zeta) [e^{\zeta-h} \theta_h(0, \zeta) + \theta(h - \zeta)]. \quad (40)$$

From this expression we can obtain the clustering coefficient at the level of the hidden variables

$$\begin{aligned} c_h &= \theta_h(0, \zeta/2) + e^{\zeta-2h} (2h - \zeta + 1) \theta_h(\zeta/2, \zeta) \\ &\quad + e^{-\zeta} (\zeta + 1) \theta(h - \zeta), \end{aligned} \quad (41)$$

and the clustering coefficient as a function of the degree  $k$ ,

$$\begin{aligned} \bar{c}(k) &= \frac{Ne^{-\zeta}}{k^2 P(k)} \theta_k(Ne^{-\zeta}, Ne^{-\zeta/2}) + \frac{N^3 e^{-2\zeta}}{k^4 P(k)} \\ &\quad \times \left[ 2 \ln\left(\frac{k}{N}\right) + \zeta + 1 \right] \theta_k(Ne^{-\zeta/2}, N) + \frac{1}{P(k)} \\ &\quad \times e^{-2\zeta} (\zeta + 1) \delta(k - N). \end{aligned} \quad (42)$$

The regular part of this formula is finally

$$\begin{aligned} \bar{c}^r(k) &= \theta_k(Ne^{-\zeta}, Ne^{-\zeta/2}) + \frac{N^2 e^{-\zeta}}{k^2} \\ &\quad \times \left[ 2 \ln\left(\frac{k}{N}\right) + \zeta + 1 \right] \theta_k(Ne^{-\zeta/2}, N). \end{aligned} \quad (43)$$

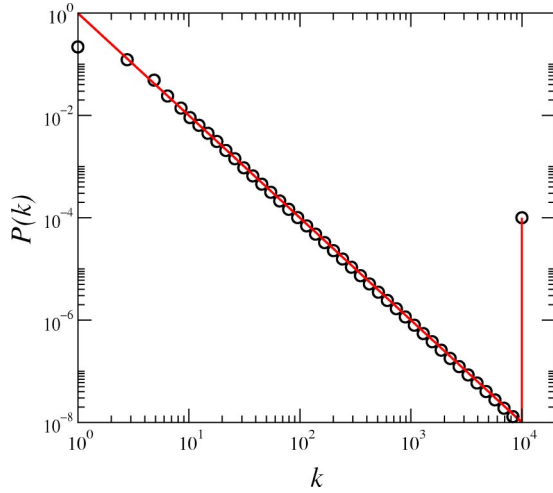


FIG. 1. Comparison between the theoretical prediction Eq. (44) for the degree distribution (solid line) and computer simulations (hollow circles) of the IF model. The isolated point at  $k=N$  corresponds to the analytical Dirac  $\delta$  function, with strength  $e^{-\zeta} = N^{-1}$ .

That is, for  $k \leq N e^{-\zeta/2}$ , the clustering coefficient is constant and equal to its maximum possible value 1. The presence of this flat region in the clustering coefficient is easy to understand. The degree range  $k \leq N e^{-\zeta/2}$  corresponds, from Eq. (37), to vertices with fitness  $h < \zeta/2$ . These vertices can only establish connections with vertices with  $h' > \zeta/2$ , which are on their turn fully interconnected among them. From here, it follows a maximum clustering coefficient equal to 1 for all vertices with  $h < \zeta/2$ . On the other hand, for  $N e^{-\zeta/2} \leq k \leq N$ , the clustering coefficient decreases as  $k^{-2}$ , modulated again by a logarithmic correction term.

### B. Numerical simulations

In order to check the validity of the proposed analytical expressions, we have performed numerical simulations of the IF model. To simplify the comparison, we have considered the particular case  $\zeta = \ln N$ , in which the relevant expressions take the form

$$P(k) = \frac{1}{k^2} \theta_k(1, N) + \frac{1}{N} \delta(k - N), \quad (44)$$

$$\bar{k}_{nn}^r(k) = \left[ 1 + N \frac{1 + \ln k}{k} \right] \theta_k(1, N), \quad (45)$$

$$\bar{c}^r(k) = \theta_k(1, N^{1/2}) + \frac{2N}{k^2} \left[ \ln \left( \frac{k}{N^{1/2}} \right) + \frac{1}{2} \right] \theta_k(N^{1/2}, N). \quad (46)$$

Simulations were performed for networks of size  $N = 10^4$  (corresponding to  $\zeta = \ln 10^4 \approx 9.2103$ ), averaging all statistical distributions over  $10^3$  to  $10^5$  network realizations.

In Fig. 1 we depict the results corresponding to the degree distribution. As we can see, the theoretical prediction (solid line) overestimates the value of the actual degree distribution

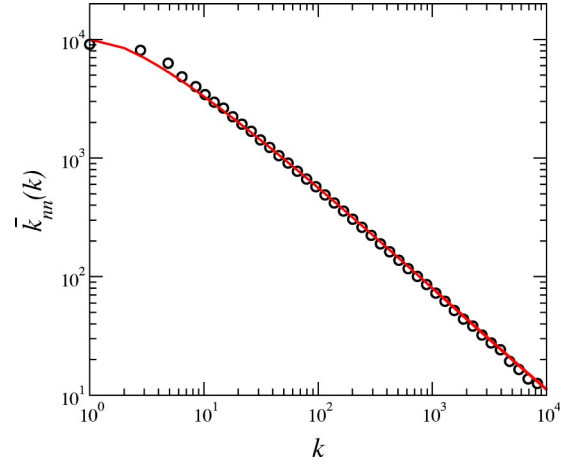


FIG. 2. Comparison between the theoretical prediction, Eq. (45), for the ANND (solid line) and computer simulations (hollow circles) of the IF model.

for small  $k$ . This is a natural effect of the continuous  $k$  approximation, which can be readily understood from the form of Eq. (44): The form  $P(k) \sim k^{-2}$  cannot be correct in a discrete approximation, since it does not fulfill the normalization condition. The condensation of edges at  $k=N$  is clearly visible in the presence of an isolated peak, of height approximately equal to  $N^{-1}$ . Figures 2 and 3, on the other hand, represent the ANND and the clustering coefficient as a function of the degree  $k$ , respectively. As we can see, the fit between the computer simulations and the analytical expressions is quite good.

### V. A PRACTICAL ALGORITHM TO GENERATE CORRELATED RANDOM NETWORKS

The hidden variable class of models represents a natural extension of the Erdős-Rényi random graph model that allows to generate a broad class of correlated networks from which it is possible to compute the most relevant topological properties. From a practical point of view, however, it is still

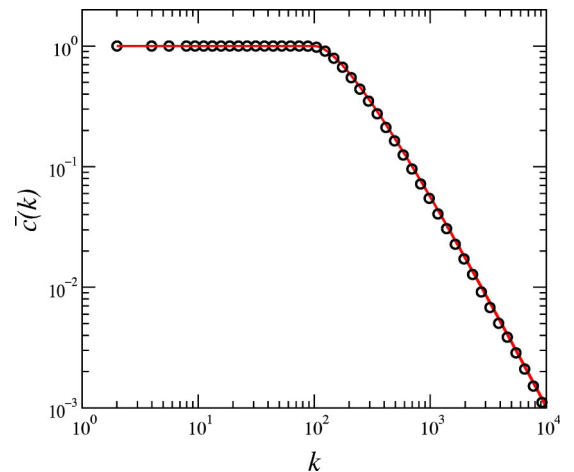


FIG. 3. Comparison between the theoretical prediction, Eq. (46), for the clustering coefficient as a function of degree  $k$  (solid line) and computer simulations (hollow circles) of the IF mode.

missing an important point. Indeed, there are many situations in which it is desirable to generate a network with a particular correlation structure given by a certain joint distribution  $P(k, k')$ . For a hidden variable model it is possible to compute this quantity as a function of the initial probabilities  $\rho(h)$  and  $r(h, h')$ . However, this relation is nontrivial, and it is generally not possible to invert it. Therefore, in order to implement an algorithm capable of generating networks with any *a priori* correlation structure, one must carefully choose the distribution of hidden variables and the connection probability.

One possible way to proceed is to define hidden variables  $h$  that have themselves the structure of the degrees of a real network (*hidden degrees*), with correlations given by a joint distribution  $\tilde{P}(h, h')$ . Those hidden degrees will then be natural numbers that are assigned to the vertices according to the probability distribution [see Eq. (7)]

$$\rho(h) = \frac{\langle h \rangle}{h} \sum_{h'} \tilde{P}(h, h'), \quad (47)$$

with  $\langle h \rangle = \sum_h h \rho(h)$ . In order to define the connection probability, we consider that, if the hidden degrees were the actual degrees characterizing the network, then the total number of edges between vertices  $h$  and  $h'$  would be  $E_{hh'} = \langle h \rangle \tilde{P}(h, h') N$ . Since the total number of  $h$  vertices is  $N_h = N \rho(h)$ , it is therefore natural to define

$$r(h, h') = \frac{\langle h \rangle}{N} \frac{\tilde{P}(h, h')}{\rho(h) \rho(h')}. \quad (48)$$

On the other hand, the conditional probability that a vertex  $h$  is connected to a vertex  $h'$  is given by [see Eq. (5)]

$$p(h'|h) = \frac{\langle h \rangle \tilde{P}(h, h')}{h \rho(h)}. \quad (49)$$

The quantities  $\rho(h)$  and  $p(h'|h)$  will be, in this case, related through the hidden degree detailed balance condition,

$$h p(h'|h) \rho(h) = h' p(h|h') \rho(h') = \langle h \rangle \tilde{P}(h, h'), \quad (50)$$

as can be checked by inserting into the definition of  $p(h'|h)$ , Eq. (26), the expression of Eq. (48). It is interesting to note that, if two-point correlations are absent at the level of the hidden variables, then we have that  $\tilde{P}_0(h, h') = hh' \rho(h) \rho(h') / \langle h \rangle^2$ . Thus, the connection probability reads

$$r_0(h, h') = \frac{hh'}{N \langle h \rangle}, \quad (51)$$

recovering the model recently introduced by Chung and Lu [50] (see also Ref. [51]).

Assuming that the connection probability is bounded and decreases for large network sizes as  $N^{-1}$ , we can compute the propagator  $g(k|h)$  applying Eq. (23) with  $\bar{k}(h)$

$= N \sum_{h'} \rho(h') r(h, h') = h$ , where we have used Eq. (48). Therefore, the propagator is a simple Poisson distribution, with average value  $h$ , i.e.,

$$g(k|h) = \frac{e^{-h} h^k}{k!}. \quad (52)$$

Note that the validity of this result leans on a quite strong assumption for the boundedness of the connection probability  $r(h, h')$ . The nature of this condition is more clearly seen in the case of uncorrelated networks, with  $r_0(h, h') = hh' / N \langle h \rangle$ . If  $r_0(h, h')$  has to decrease as  $N^{-1}$ , then the maximum value of the hidden degree must be smaller than  $h_c(N) = (N \langle h \rangle)^{1/2}$  [50], a condition that imposes restrictions on the maximum degree available for any vertex.

From the propagator, Eq. (52), the degree distribution as a function of  $\rho(h)$  follows immediately from Eq. (11):

$$P(k) = \sum_h \frac{e^{-h} h^k}{k!} \rho(h). \quad (53)$$

This relation between distributions implies a relation between the respective moments. Indeed, it is straightforward to prove that

$$\langle h^n \rangle = \langle k(k-1) \cdots (k-n+1) \rangle, \quad (54)$$

and, in particular, the first two moments read

$$\langle h \rangle = \langle k \rangle, \quad \langle h^2 \rangle = \langle k^2 \rangle - \langle k \rangle. \quad (55)$$

It is also instructive to see how we can recover the classical Erdős-Rényi random graph model from this formalism. The Erdős-Rényi model corresponds to joining pairs of vertices with a constant probability  $p$ . Such connection probability results from imposing uncorrelated hidden degrees with distribution  $\rho_{ER}(h) = \delta_{\langle k \rangle, h}$ , which yields a Poisson degree distribution with average degree  $\langle k \rangle$ .

In order to compute the ANND function from Eq. (27), we observe that in this case  $\bar{k}(h) = h$ . From here we obtain

$$\bar{k}_{nn}(k) = 1 + \frac{1}{P(k)} \sum_h \frac{e^{-h} h^k}{k!} \rho(h) \bar{h}_{nn}(h), \quad (56)$$

where, in this case, the average hidden degree of the nearest neighbors as a function of  $h$  [see Eq. (28)] is

$$\bar{h}_{nn}(h) = \sum_{h'} h' p(h'|h) = \frac{\langle h \rangle}{h \rho(h)} \sum_{h'} h' \tilde{P}(h, h'). \quad (57)$$

For uncorrelated networks at the hidden level, the ANND yields

$$\bar{k}_{nn}^0(k) = 1 + \frac{\langle h^2 \rangle}{\langle h \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}, \quad (58)$$

recovering the well-known result [52]. Finally, the clustering coefficient takes, from Eq. (30), the form



$$\bar{c}(k) = \frac{1}{P(k)} \sum_h \frac{e^{-h} h^k}{k!} \rho(h) c_h, \quad (59)$$

where the clustering coefficient in terms of the hidden degrees is given by

$$\begin{aligned} c_h &= \sum_{h', h''} p(h'|h) r(h', h'') p(h''|h) \\ &= \frac{\langle h \rangle^3}{h^2 \rho(h)^2 N} \sum_{h', h''} \frac{\tilde{P}(h, h') \tilde{P}(h', h'') \tilde{P}(h'', h)}{\rho(h') \rho(h'')}. \end{aligned} \quad (60)$$

When correlations are missing in the hidden degree distribution, we obtain

$$\bar{c}^0(k) = \frac{\langle h^2 \rangle^2}{N \langle h \rangle^3} = \frac{(\langle k^2 \rangle - \langle k \rangle)^2}{N \langle k \rangle^3}, \quad (61)$$

recovering the result previously derived by Newman [53].

The key point to notice in the above expressions is that the Poisson propagator, Eq. (52), is a sharply peaked function at  $k=h$ , which in the large  $k$  limit is analogous to a delta function  $\delta_{h,k}$ . Therefore, in the limit  $k \rightarrow \infty$ , we expect to observe the behavior

$$P(k) \sim \rho(k), \quad (62)$$

$$\bar{k}_{nn}(k) \sim 1 + \bar{h}_{nn}(k), \quad (63)$$

$$\bar{c}(k) \sim c_k. \quad (64)$$

That is, the main topological properties referred to the actual degree  $k$  tend to their analogs computed for the hidden degree  $h$ , with the sole exception of a constant of order unit added to the ANND function. We can take advantage of this observation to propose the following algorithm to generate a correlated random network with theoretical degree distribution  $P_i(k)$  and joint distribution  $P_i(k, k')$ .

(1) Assign to each vertex  $i$  an integer random variable  $\tilde{k}_i$ ,  $i=1, \dots, N$ , drawn from the probability distribution  $P_i(k)$ .

(2) For each pair of vertices  $i$  and  $j$ , draw an undirected edge with probability  $r(\tilde{k}_i, \tilde{k}_j) = \langle k \rangle P_i(\tilde{k}_i, \tilde{k}_j) / N P_i(\tilde{k}_i) P_i(\tilde{k}_j)$ .

The outcome of this process will be a random network whose actual degree structure, in the large  $k$  limit, will be distributed according to the probability  $P_i(k)$ , with correlations given by  $P_i(k, k')$ .

In order to check the accuracy of the previous algorithm, we have tested it with the joint probability distribution

$$P_i(k, k') = A q^{kk'}, \quad k, k' = 1, 2, \dots, \quad (65)$$

where  $A$  is a normalization constant and  $q < 1$  is a constant parameter. With this choice, the degree distribution takes the form

$$P_i(k) = \frac{\langle k \rangle A q^k}{k(1-q^k)}, \quad (66)$$

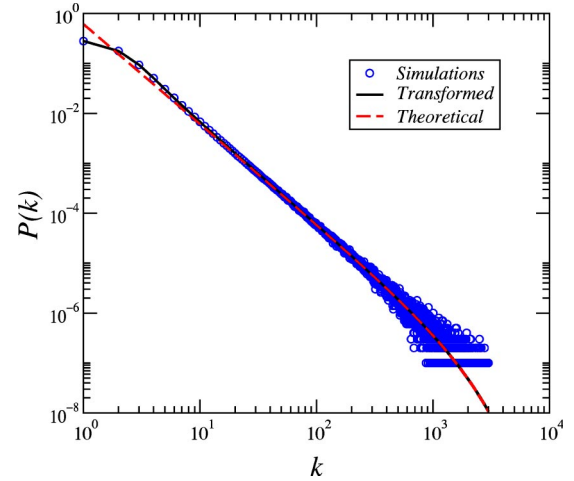


FIG. 4. Degree distribution obtained from numerical simulations of the proposed algorithm, applied to the joint distribution, Eq. (65), compared with the theoretical and transformed values.

which, for  $q \rightarrow 1$ , approaches a power-law distribution with exponent  $\gamma=2$ . In Figs. 4–6 we present the results for the degree distribution, the ANND function, and the clustering coefficient, respectively, from computer simulations of the proposed algorithm, using the joint probability distribution given by Eq. (65). The plots have been obtained for networks of size  $N=10^4$  and a parameter  $q=0.999$ , averaging over  $10^3$  realizations. In the same graphs we also represent the theoretical values corresponding to a network with a correlation structure given by  $P_i(k, k')$ , plus the transformed functions given by Eqs. (53), (56), and (59), respectively, that correspond to the actual topological properties of the network. As discussed previously, and to ease the comparison of the plots, a factor 1 has been subtracted to the ANND function obtained from computer simulations and the transformation, Eq. (56). We can see that for all three quantities, the matching between the computer simulations and the theoretical results is very good for values of  $k$  larger than 10. Being the discrepancy limited to such small degree values,

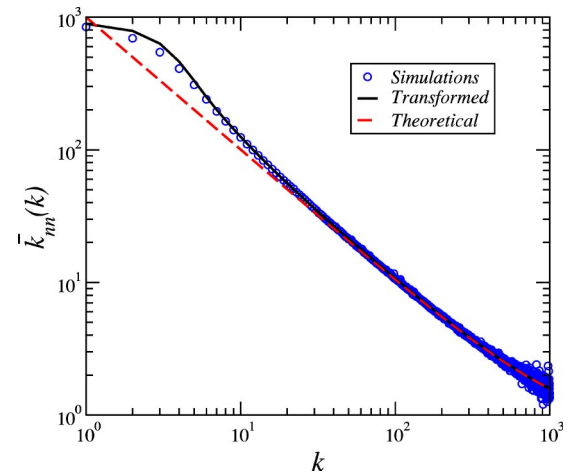


FIG. 5. ANND function obtained from numerical simulations of the proposed algorithm, applied to the joint distribution, Eq. (65), compared with the theoretical and transformed values.

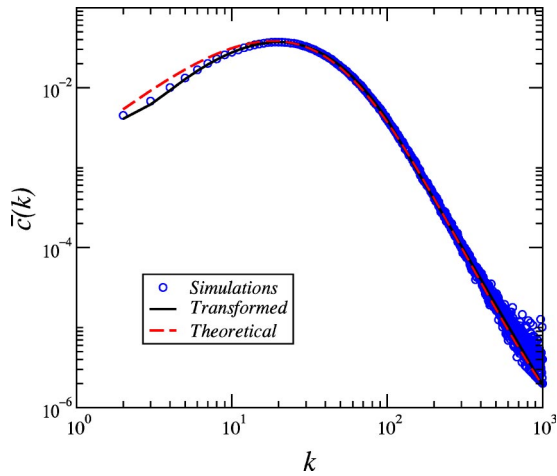


FIG. 6. Clustering coefficient obtained from numerical simulations of the proposed algorithm, applied to the joint distribution, Eq. (65), compared with the theoretical and transformed values.

we conclude that the proposed algorithm reproduces the desired correlation structure with an accuracy that is more than satisfactory for any purpose dealing with the large scale properties of the network.

## VI. NONEQUILIBRIUM CORRELATED RANDOM NETWORKS

The class of models we have introduced so far are static models, in which, starting from a fixed number  $N$  of vertices, edges are assigned with a given probability. As we have seen, this construction is extremely useful because it gives us control over the final network structure and, at the same time, the possibility to calculate important structural properties. Many real networks, however, are far from being static. Instead, many of them are the result of an evolving process [3], in which vertices are added to the network following some growing process (linear, exponential, etc.), establishing connections to other existing vertices with a given attachment rule (preferential attachment [34], deactivation of vertices [35], etc). From this growing mechanism the network reaches a nonequilibrium steady state where the statistical properties are time independent.

In the following we will see how it is possible to map nonequilibrium growing networks into a particular kind of model with hidden variables. The key point is to realize that, after the growth of the network, all vertices that joined the network at the same time are statistically equivalent and, thus, the hidden variable of a vertex must correspond to its injection time  $t$  [33]. We consider a time window  $t \in [t_0, T]$ , with  $T \gg t_0$ , in which the initial time can be taken as  $t_0 = 1$ , without lack of generality. If  $\lambda$  is the rate of creation of new vertices per unit time [54] the network size is given by  $N = \lambda(T - 1)$ . In this case, the density of hidden variables,  $\rho(t)$ , is a uniform distribution defined in the range  $[1, T]$ , that is,  $\rho(t) = 1/(T - 1)$ , reflecting the linear growth of the network. The main difference between this class of nonequilibrium networks and the classes discussed in the previous sections lies in the fact that the distribution  $\rho(t)$  has

now an explicit dependence on the system size  $N$  (or the final time  $T$ ).

The next step is to define the connection probability of two vertices that joined the network at times  $t$  and  $t'$ ,  $r(t, t')$ . The choice of this function is equivalent to the connection probability at the time that the vertices were added and its specific form will determine the final properties of the network. For instance, an homogeneous form for  $r(t, t')$ , that is,  $r(t, t') = f(|t - t'|)$ , will produce, in the large  $T$  limit, networks in which all the vertices will have the same statistical properties, independent of the injection time. In the opposite case of inhomogeneous networks, vertices introduced at early times will have different topological properties than those added later, giving rise to *aging* (that is, an explicit dependence on the injection time  $t$  of all the vertex properties evaluated at time  $T > t$ ).

In order to provide a particular example, we focus on the class of growing scale-free networks, whose most characteristic element is the Barabási-Albert model [34]. From general scaling arguments [2,3], it is possible to see that growing scale-free networks are described by a power-law degree distribution  $P(k) \sim k^{-\gamma}$ , while the average degree at time  $t$  of a vertex introduced at time  $t'$  is given by

$$k_{t'}(t) \sim \left(\frac{t}{t'}\right)^\beta, \quad 0 < \beta < 1, \quad t > t' \gg 1, \quad (67)$$

where the exponents  $\beta$  and  $\gamma$  fulfill the scaling relation

$$\gamma = 1 + \frac{1}{\beta}. \quad (68)$$

In the Barabási-Albert model, corresponding to  $\beta = 1/2$  and  $\gamma = 3$ , new edges are joined to old vertices following a preferential attachment prescription, that is, with probability proportional to the degree of the existing vertices. We can generalize this prescription and consider a preferential attachment as a function of time  $t$ , mapping the degree to the time by means of Eq. (67). In this case, however, when a new vertex is added at time  $t$ , it can establish connections with nodes introduced between 1 and  $t$  and, consequently, the connection probability is to be rescaled by a factor  $\int' k_{t'}(t) dt'$ , which is proportional to  $t$ . Therefore, the probability that a new vertex, created at time  $t$ , will be joined to a vertex injected at time  $t' < t$  is proportional to  $(t/t')^\beta/t$ .

Following this reasoning we propose a connection probability as a function of times  $t$  and  $t'$  given by

$$r(t, t') = \alpha \left[ \frac{1}{t} \left(\frac{t}{t'}\right)^\beta \theta(t - t') + \frac{1}{t'} \left(\frac{t'}{t}\right)^\beta \theta(t' - t) \right], \quad (69)$$

where  $\alpha$  is a parameter that controls the final average degree of the network. The connection probability has been symmetrized for  $t$  and  $t'$  to comply with the general conditions of this function. Its symmetric property, however, does not imply that the average properties of the vertices are independent of  $t$ . For example, computing the average degree of a  $t$

vertex (introduced at time  $t$ ), evaluated at the final time  $T$ , using Eq. (20), we obtain [55]

$$\bar{k}(t) = \lambda \int_1^T r(t, t') dt' = \frac{\alpha\lambda}{1-\beta} \left( 1 - \frac{1}{t^{1-\beta}} \right) + \frac{\alpha\lambda}{\beta} \left[ \left( \frac{T}{t} \right)^\beta - 1 \right]. \quad (70)$$

For large  $T$ , we have  $\bar{k}(t) \sim (T/t)^\beta$ , recovering the behavior obtained in growing network models, Eq. (67), if we consider the time  $T$  as the observation time. On the other hand,  $\bar{k}(t)$  is a decreasing function of  $t$  between the limits  $t=1$ , yielding  $\bar{k}_{max} \sim \alpha\lambda T^\beta/\beta$ , and  $t=T$ , where it converges to the constant  $\bar{k}_{min} \sim \alpha\lambda/(1-\beta)$ . This functional form implies that the oldest vertices (with smaller  $t$ ) have a larger average degree, which is the signature of aging in the network [34].

From Eq. (21), the average degree of the network can be computed as

$$\langle k \rangle = \frac{1}{T-1} \int_1^T \bar{k}(t) dt = \frac{2\alpha\lambda}{1-\beta} \left( 1 - \frac{1}{\beta} \frac{T^\beta - 1}{T-1} \right), \quad (71)$$

which, in the limit of large  $T$  and  $\beta < 1$ , tends to  $\langle k \rangle = 2\alpha\lambda/(1-\beta)$ , from where we identify the normalization parameter  $\alpha$  as a function of the average degree. For  $\beta = 1$ , on the other hand, the average degree diverges as  $\langle k \rangle \sim 2\alpha\lambda \ln T$ . Note that this choice of the connection probability, independent of the network size, yields for  $\beta < 1$  a sparse network (with finite average degree in the thermodynamic limit), in opposition with the case discussed in Sec. III. This fact is due to the explicit dependence on  $N$  of the distribution of times  $\rho(t)$ , and signals the crucial difference of nonequilibrium networks with hidden variables.

In order to obtain the form of the degree distribution, we observe that, even though the network is sparse, since  $r(t, t')$  is not proportional to  $N^{-1}$  we cannot rigorously apply Eq. (23). However, we note that the maximum value of  $r(t, t')$  takes place at  $t=t'$ , namely,

$$r(t, t) = \frac{\alpha}{t} = \frac{\langle k \rangle (1-\beta)}{2\lambda t}. \quad (72)$$

For not very large values of  $\langle k \rangle$  and large  $t$ , the connection probability is bounded by a small value, and therefore we can still approximate its propagator by means of Eq. (23). Working for simplicity with the generating function of the degree distribution,  $\hat{P}(z) = \sum_k z^k P(k)$ , we therefore write, from Eq. (22),

$$\hat{P}(z) = \frac{1}{T-1} \int_1^T e^{(z-1)\bar{k}(t)} dt. \quad (73)$$

Performing the change of variables  $\tau \equiv t/T$ , and considering the limit  $T \rightarrow \infty$ , we obtain

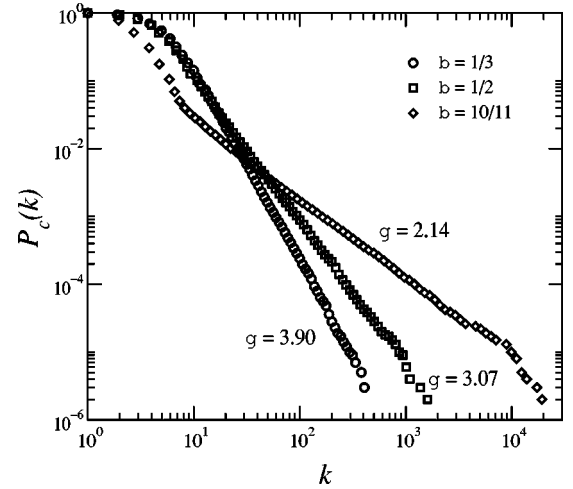


FIG. 7. Cumulative distribution for the mapping of the growing model for different values of the parameter  $\beta$  ( $\beta = 1/3, 1/2$ , and  $10/11$ ). The average degree is in all cases set to  $\langle k \rangle = 6$ . The size of the network is  $N = 10^6$ .

$$\begin{aligned} \hat{P}(z) &= \exp \left\{ \frac{\alpha\lambda(2\beta-1)(z-1)}{\beta(1-\beta)} \right\} \int_0^1 e^{-\alpha\lambda(1-z)\tau^{-\beta}/\beta} d\tau \\ &= \left[ \frac{\alpha\lambda}{\beta} (1-z) \right]^{1/\beta} \frac{\Gamma(-1/\beta, \alpha\lambda(1-z)/\beta)}{\beta} \\ &\quad \times \exp \left\{ \frac{\alpha\lambda(2\beta-1)(z-1)}{\beta(1-\beta)} \right\}, \end{aligned} \quad (74)$$

where  $\Gamma(x, y)$  is the incomplete Gamma function. Expanding  $\hat{P}(z)$  around  $z=1$ , we obtain the leading terms

$$\hat{P}(z) \approx 1 - \langle k \rangle (1-z) + \left( \frac{\alpha\lambda}{\beta} \right)^{1/\beta} \frac{\Gamma(-1/\beta)}{\beta} (1-z)^{1/\beta}. \quad (75)$$

Applying Tauberian theorems [56], it follows from the singularity  $\hat{P}(z) \sim (1-z)^{1/\beta}$  in the vicinity of  $z=1$ , the large  $k$  behavior of the degree distribution, namely,  $P(k) \sim k^{-1-1/\beta}$ ; that is, the model recovers a scale-free network,  $P(k) \sim k^{-\gamma}$ , with a degree exponent  $\gamma = 1 + 1/\beta$ , and a cutoff given by the maximum degree  $\bar{k}_{max} \sim T^\beta \sim N^{1/(\gamma-1)}$ , in agreement with the results obtained for growing network models [3].

In order to check this result, we have numerically generated networks of size  $N = 10^6$ , with the connection probability, Eq. (69). For a rate of addition of new vertices  $\lambda = 1$ , the numerical prefactor in  $r(t, t')$  is  $\alpha = \langle k \rangle (1-\beta)/2$ ; we impose  $\langle k \rangle = 6$ . Figure 7 shows the numerical cumulated degree distributions obtained for values  $\beta = 1/3, 1/2$ , and  $10/11$ . The plots show a clear power-law behavior in all three cases, with a degree exponent, estimated from a linear fit in the scaling region, given by  $\gamma = 3.90, 3.07$ , and  $2.14$ ,

respectively. These values compare very well with the theoretical prediction  $\gamma = 1 + 1/\beta$ , which provides the expected exponents 4, 3, and 2.1.

The ANND function  $\bar{k}_{nn}(k)$  can also be analyzed using our formalism. By means of Eq. (27), assuming a Poisson form for the propagator, we can define the generating function  $\hat{\Psi}(z) = \sum_k z^k k P(k) [\bar{k}_{nn}(k) - 1]$ , which takes the form

$$\hat{\Psi}(z) = \frac{\lambda z}{T-1} \int_1^T dt e^{\bar{k}(t)(z-1)} \int_1^T dt' r(t, t') \bar{k}(t'), \quad (76)$$

from where it is possible to derive the large  $k$  limit of the ANND. We are primarily interested in the case  $\beta \geq 1/2$ , that is, the scale-free range of the model. Focusing in  $\beta > 1/2$ , the limit  $T \rightarrow \infty$  of Eq. (76) is

$$\hat{\Psi}(z) = \frac{\alpha^2 \lambda^2 z}{\beta(2\beta-1)} T^{2\beta-1} \int_0^1 \frac{1}{\tau^{1-\beta}} e^{(z-1)\bar{k}(\tau)} d\tau, \quad (77)$$

where we have used the same change of variables as before. An expansion of this equation around  $z = 1$  leads to

$$\hat{\Psi}(z) \approx \frac{\alpha^2 \lambda^2 T^{2\beta-1}}{\beta^2(2\beta-1)} \left( 1 + \frac{\alpha \lambda (1-z)}{\beta} \ln \frac{\alpha \lambda (1-z)}{\beta} \right). \quad (78)$$

Applying again Tauberian theorems, we can write that, in the large  $k$  limit,

$$k P(k) [\bar{k}_{nn}(k) - 1] \sim T^{2\beta-1} \frac{1}{k^2}, \quad (79)$$

from where it is straightforward to derive the large  $k$  limit of the ANND:

$$\bar{k}_{nn}(k) \sim N^{(3-\gamma)/(\gamma-1)} k^{-(3-\gamma)}. \quad (80)$$

The case  $\beta = 1/2$  is very similar to the previous one, except for the type of divergence with the system size appearing as a prefactor. In the large  $T$  limit, we can write

$$\hat{\Psi}(z) \sim 2 \alpha^2 \lambda^2 z \ln T \int_0^1 \frac{1}{\tau^{1/2}} e^{-2\alpha\lambda(1-z)\tau^{-1/2}} d\tau. \quad (81)$$

Using the same arguments we conclude that the right-hand side of this equation scales as  $k^{-2}$  and, therefore, the ANND converges to a constant value proportional to  $\ln N$ .

## VII. CONCLUSIONS

In this paper we have analyzed in detail a general class of complex network models that are based on the existence of a hidden space, in which vertices are located, and a connection probability that depends on the hidden variable of each vertex. The Markovian character at the hidden level allows to calculate analytical expressions for the most important structural properties, such as degree distribution, the ANND function—quantifying two vertices correlations—and clustering coefficient, as a measure of three vertices correlation. Our formalism is valid for both sparse and nonsparse networks, extending the applicability of our results to a wide range of complex networks. At this respect, one of the applications of our formalism is to provide the analytical solution of a recently introduced model with intrinsic fitness, which has recently attracted a great deal of interest as a way to obtain scale-free networks without preferential attachment. Our solution has been successfully contrasted with numerical simulations, thus validating the accuracy of our formalism.

Another interesting result of our analysis is to provide a new algorithm for generating correlated random networks with an *a priori* specified correlation structure. In this case, we also calculate exact formulas for the relevant quantities. We have tested the algorithm using a probe joint distribution  $P(k, k')$ , with very encouraging results.

Perhaps the most striking result concerns the application of the formalism to growing networks. Even though, in this case, the network is out of equilibrium, it is possible to map it to a specific kind of hidden variable model by the identification of the injection time as the hidden variable. In order to check this point we have applied the method to a general class of growing networks, which, as a particular case, contains the Barabási-Albert model. Using our formalism, we have recovered all the known results for this models, both for the degree distribution and the correlation structure. It is remarkable that, from a static approach, our formalism is able to derive correct results for nonequilibrium evolving networks. Therefore, this approach opens an appealing way to study such systems.

## ACKNOWLEDGMENTS

We thank G. Caldarelli, M. A. Muñoz, and A. Vespignani for helpful comments and discussions. This work was partially supported by the European Commission - Fet Open Project No. COSIN IST-2001-33555. R.P.-S. acknowledges financial support from the Ministerio de Ciencia y Tecnología (Spain).

- 
- [1] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).  
 [2] S.N. Dorogovtsev and J.F.F. Mendes, *Adv. Phys.* **51**, 1079 (2002).  
 [3] S.N. Dorogovtsev and J.F.F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford Uni-

versity Press, Oxford, 2003).

- [4] B. Bollobás, *Modern Graph Theory* (Springer-Verlag, New York, 1998).  
 [5] M. Faloutsos, P. Faloutsos, and C. Faloutsos, *Comput. Commun. Rev.* **29**, 251 (1999).

- [6] G. Caldarelli, R. Marchetti, and L. Pietronero, *Europhys. Lett.* **52**, 386 (2000).
- [7] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001).
- [8] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, *Phys. Rev. E* **65**, 066130 (2002).
- [9] R. Albert, H. Jeong, and A.-L. Barabási, *Nature (London)* **401**, 130 (1999).
- [10] S.H. Strogatz, *Nature (London)* **410**, 268 (2001).
- [11] J.M. Montoya and R.V. Solé, *J. Theor. Biol.* **214**, 405 (2002).
- [12] A. Wagner, *Mol. Biol. Evol.* **18**, 1283 (2001).
- [13] H. Jeong, S.S. Mason, A.L. Barabási, and Z.N. Oltvai, *Nature (London)* **411**, 41 (2001).
- [14] R.V. Solé, R. Pastor-Satorras, E. Smith, and T. Kepler, *Adv. Complex Syst.* **5**, 43 (2002).
- [15] A. Vázquez, A. Flammini, A. Maritan, and A. Vespignani, *ComPlexUs* **1**, 38 (2003).
- [16] D.J. Watts and S.H. Strogatz, *Nature (London)* **393**, 440 (1998).
- [17] R.A. Albert, H. Jeong, and A.-L. Barabási, *Nature (London)* **406**, 378 (2000).
- [18] D.S. Callaway, M.E.J. Newman, S.H. Strogatz, and D.J. Watts, *Phys. Rev. Lett.* **85**, 5468 (2000).
- [19] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, *Phys. Rev. Lett.* **86**, 3682 (2001).
- [20] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).
- [21] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. E* **63**, 066117 (2001).
- [22] R.M. May and A.L. Lloyd, *Phys. Rev. E* **64**, 066112 (2001).
- [23] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, *Eur. Phys. J. B* **26**, 521 (2002).
- [24] M.E.J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
- [25] M.E.J. Newman, *Phys. Rev. E* **67**, 026126 (2003).
- [26] M. Boguñá and R. Pastor-Satorras, *Phys. Rev. E* **66**, 047104 (2002).
- [27] Y. Moreno and A. Vázquez, *Eur. Phys. J. B* **31**, 265 (2003).
- [28] A. Vázquez and Y. Moreno, *Phys. Rev. E* **67**, 015101 (2003).
- [29] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, *Phys. Rev. Lett.* **90**, 028701 (2003).
- [30] J. Berg and M. Lässig, *Phys. Rev. Lett.* **89**, 228701 (2002).
- [31] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, e-print cond-mat/0206467.
- [32] G. Caldarelli, A. Capocci, P.D.L. Rios, and M.A. Muñoz, *Phys. Rev. Lett.* **89**, 258702 (2002).
- [33] B. Söderberg, *Phys. Rev. E* **66**, 066121 (2002).
- [34] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [35] K. Klemm and V.M. Eguíluz, *Phys. Rev. E* **65**, 036123 (2002).
- [36] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, e-print cond-mat/0301149.
- [37] Being more precise, we should write  $P(k) \equiv \lim_{N \rightarrow \infty} N_k / N$ , although, for the sake of simplicity, we will obviate the limit here and in the following.
- [38] Note that the probability  $p_{k',k''}$  will in general depend on the particular network considered, and can also be a function of the degree  $k$  of the common vertex.
- [39] E. Ravasz and A.-L. Barabási, *Phys. Rev. E* **67**, 026112 (2003).
- [40] B. Söderberg, e-print cond-mat/0305192.
- [41] K.-I. Goh, B. Kahng, and D. Kim, *Phys. Rev. Lett.* **87**, 278701 (2001).
- [42] P. Erdős and P. Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17 (1960).
- [43] E.N. Gilbert, *Ann. Math. Stat.* **30**, 1141 (1959).
- [44] But not generally at the actual degree level.
- [45] B.V. Gnedenko, *The Theory of Probability* (Chelsea, New York, 1962).
- [46] H.S. Wilf, *Generatingfunctionology*, 2nd ed. (Academic Press, San Diego, 1994).
- [47] The difference between ensembles can be of importance in cases where the fluctuations of the number of vertices does not vanish in the thermodynamic limit.
- [48] G. Bianconi and A.-L. Barabási, *Europhys. Lett.* **54**, 436 (2001).
- [49] P.L. Krapivsky, S. Redner, and F. Leyvraz, *Phys. Rev. Lett.* **85**, 4629 (2000).
- [50] F. Chung and L. Lu, *Ann. Combinatorics* **6**, 125 (2002).
- [51] J. Park and M.E.J. Newman, e-print cond-mat/0303327.
- [52] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, e-print cond-mat/0206084.
- [53] M.E.J. Newman, in *Handbook of Graphs and Networks: From the Genome to the Internet*, edited by S. Bornholdt and H.G. Schuster (Wiley-VCH, Berlin, 2003), pp. 35–68.
- [54] We consider here a linear growth of new vertices.
- [55] The average degree of a vertex injected at time  $t$ ,  $\bar{k}(t)$ , should not be confused with  $k_i(t')$ , which measures the average degree of the same vertex at an intermediate time  $t' > t$ .
- [56] R.A. Handelsman and J.S. Lew, *SIAM J. Math. Anal.* **5**, 425 (1974).