Generation of uncorrelated random scale-free networks

Michele Catanzaro,[1] Marián Boguñá,[2] and Romualdo Pastor-Satorras[1]
[1]*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain*
[2]*Departament de Física Fonamental, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain*
(Received 5 August 2004; published 24 February 2005)

Uncorrelated random scale-free networks are useful null models to check the accuracy and the analytical solutions of dynamical processes defined on complex networks. We propose and analyze a model capable of generating random uncorrelated scale-free networks with no multiple and self-connections. The model is based on the classical configuration model, with an additional restriction on the maximum possible degree of the vertices. We check numerically that the proposed model indeed generates scale-free networks with no two- and three-vertex correlations, as measured by the average degree of the nearest neighbors and the clustering coefficient of the vertices of degree k, respectively.

DOI: 10.1103/PhysRevE.71.027103 PACS number(s): 89.75.−k, 87.23.Ge, 05.70.Ln

Complex networks constitute a general framework for the topological characterization of many natural and technological systems whose complexity prevents a more detailed microscopic description [1–3]. Within this framework, these systems are represented in terms of networks or graphs [4], in which vertices stand for the units composing the system, while edges among vertices represent the interactions or relations between pairs of units. The focus is thus shift to the topological characterization of the representative network, a task which is largely more feasible and yields, nevertheless, a noticeable amount of information on the structure and properties of the original system. The empirical analysis of many real complex networks has unveiled the presence of several typical properties, widely found in systems belonging to a large variety of realms. One of the most relevant is given by the scale-free nature of the degree distribution $P(k)$ [1,3,5], defined as the probability that a randomly chosen vertex has degree k (i.e., it is connected to other k vertices). In mathematical terms, the scale-free property translates into a power-law function of the form

$$P(k) \sim k^{-\gamma}, \tag{1}$$

where γ is a characteristic degree exponent. The presence of a scale-free degree distribution can have an important impact on the behavior of dynamical processes taking place on top of the network. Indeed, scale-free networks with exponent γ in the range $2<\gamma\leqslant 3$ show large fluctuations in their degrees, evident in the presence of a diverging second moment $\langle k^2\rangle$ in the infinite-network-size limit $N\to\infty$. This divergence, in turn, shows up in a remarkable weakness of the network in front of targeted attacks [6,7] or the propagation of infectious agents [8,9].

It has been recently realized that, besides their degree distribution, real networks are also characterized by the presence of degree correlations. This translates in the observation that the degrees at the end points of any given edge are not usually independent. This kind of degree-degree correlations can be theoretically expressed in terms of the conditional probability $P(k'|k)$ that a vertex of degree k is connected to a vertex of degree k'. From a numerical point of view, it is more convenient to characterize degree-degree correlations by means of the average degree of the nearest neighbors (NN) of the vertices of degree k, which is formally defined as [10]

$$\overline{k}_{nn}(k) = \sum_{k'} k' P(k'|k). \tag{2}$$

Degree-degree correlations have led to a first classification of complex networks according to this property [11]. Thus, when $\overline{k}_{nn}(k)$ is an increasing function of k, the corresponding network is said to exhibit *assortative mixing by degree*; i.e., highly connected vertices are preferentially connected to highly connected vertices and vice versa, while a decreasing $\overline{k}_{nn}(k)$ function is typical of *disassortative mixing*, highly connected vertices being more probably connected to poorly connected ones. For uncorrelated networks, the degrees at the end points of any edge are completely independent. Therefore, the conditional probability $P(k'|k)$ can be simply estimated as the probability that any edge points to a vertex of degree k', leading to $P_{\rm nc}(k'|k)=k'P(k')/\langle k\rangle$, independent of k. Inserting this equation into Eq. (2), the average nearest-neighbor degree reads

$$\overline{k}^{\rm nc}_{nn}(k) = \frac{\langle k^2\rangle}{\langle k\rangle}, \tag{3}$$

that is, independent of the degree k.

Analogously, from a theoretical point of view, correlations concerning three vertices can be characterized by means of the conditional probability $P(k'',k'|k)$ that a vertex of degree k is simultaneously connected to two vertices of degrees k' and k''. We can estimate this kind of three-point correlations by means of the clustering coefficient of the vertices of degree k, $\overline{c}(k)$ [12,13], defined as the probability that two neighbors of a vertex of degree k are also neighbors themselves. This function can be formally written as

$$\overline{c}(k) = \sum_{k',k''} P(k'',k'|k)p_{k',k''}, \tag{4}$$

where $p_{k',k''}$ is the probability that vertices k' and k'' are connected given that they have a common neighbor [14,15]. An important class of random networks is composed of the so-called *Markovian networks* [16], for which all topological information is encoded into the degree distribution $P(k)$ and the conditional probability $P(k'|k)$. In this case, the three-vertex conditional probability can be factorized as $P(k'',k'|k)=P(k''|k)P(k'|k)$, for $k>1$. Furthermore, when the network is totally uncorrelated, the connection probability can also be computed as $p_{k',k''}=(k'-1)(k''-1)/\langle k\rangle N$, where the term -1 comes from the fact that one of the connections of each vertex has already been used [14,15,17]. From the above relations, the clustering coefficient for uncorrelated random networks becomes

$$\overline{c}_{\mathrm{nc}}(k) = \frac{(\langle k^2\rangle - \langle k\rangle)^2}{\langle k\rangle^3 N}. \tag{5}$$

This expression was first derived by Newman [17] (see also [14,15,18]). As in the previous case, for uncorrelated networks, the function $\overline{c}(k)$ is constant and independent of k. Therefore, any nontrivial dependence of the functions $\overline{k}_{nn}(k)$ and $\overline{c}(k)$ on the degree is a signature of the presence of two- and three-point correlations, respectively.

While most real networks show indeed the presence of correlations, uncorrelated random networks are nevertheless equally important from a practical point of view, especially as null network models in which to test the behavior of dynamical systems whose analytic solution is usually available only in the absence of correlations [6–8,19]. Therefore, it becomes an interesting issue the possibility to generate random networks which have a guaranteed lack of correlations. In the particular case of scale-free networks, however, finding such algorithms is far more difficult than one would expect *a priori*. In this paper, we observe that classical algorithms, which are supposed to generate uncorrelated networks, do, indeed, generate correlations when the desired degree distribution is scale free and no more than one edge is allowed between any two vertices [20,21]. To solve this drawback, we present and test an algorithm capable to generate uncorrelated scale-free networks.

The classical algorithm to construct random networks with any prescribed degree distribution $P(k)$ is the so-called configuration model (CM) [17,22–25]. To construct a network with the original definition of this algorithm, we start assigning to each vertex i, in a set of N vertices, a random number k_i of "stubs"—ends of edges emerging from the vertex—drawn from the probability distribution $P(k)$, with $m\leqslant k_i<N$ (no vertex can have a degree larger than $N-1$) and imposing the constraint that the sum $\Sigma_i k_i$ must be even. The network is completed by connecting pairs of these stubs chosen uniformly at random to make complete edges, respecting the preassigned sequence k_i. The result of this construction is a random network whose degrees are, by definition, distributed according to $P(k)$ and in which, in principle, there are no degree correlations, given the random nature of the edge assignment.

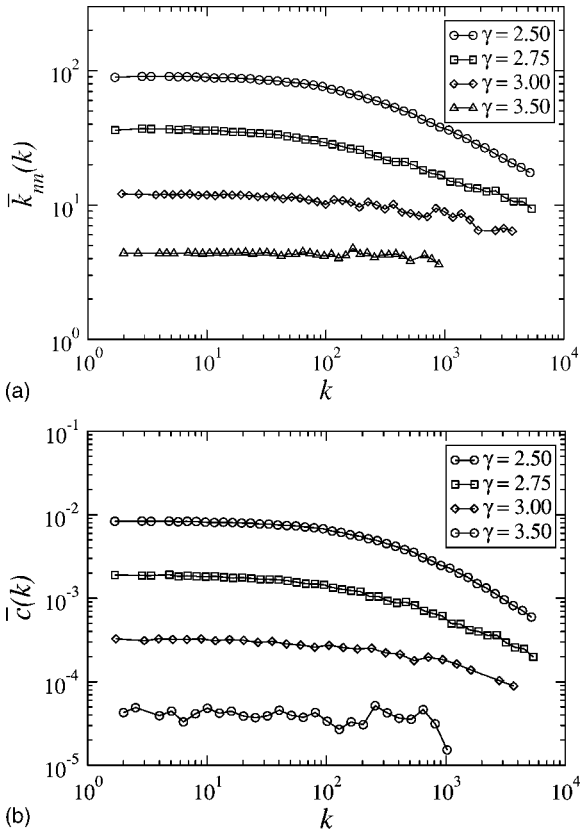

FIG. 1. Average nearest-neighbor degree of vertices of degree k, $\overline{k}_{nn}(k)$ (a), and average clustering coefficient $\overline{c}(k)$ (b) for the original CM algorithm with different degree exponents γ. Network size is $N=10^5$.

While this prescription works well for bounded degree distributions, in which $\langle k^2\rangle$ is finite, one has to be more careful when dealing with networks with a scale-free distribution, which, for $2<\gamma\leqslant 3$, yield diverging fluctuations, $\langle k^2\rangle \to\infty$, in the infinite-network-size limit. In fact, it is easy to see that, if the second moment of the degree distribution diverges, a completely random assignment of edges leads to the construction of an uncorrelated network, but in which a non-negligible fraction of self-connections (a vertex joined to itself) and multiple connections (two vertices connected by more than one edge) are present [28]. While multiple and self-connections are completely natural in mathematical graph theory [4], they are somewhat undesired for simulation purposes, since most real network do not display such structures, and also in order to avoid ambiguities in the definition of the network and any dynamics on top of it. This situation can be avoided by imposing the additional constraint of forbidding multiple and self-connections. This constraint, however, has the negative side effect of introducing correlations in the network [20,21]. As an example of this fact, in Fig. 1 we show the functions $\overline{k}_{nn}(k)$ and $\overline{c}(k)$ computed from numerical simulations of the CM algorithm with no multiple and self-connections for different γ exponents and fixed net-

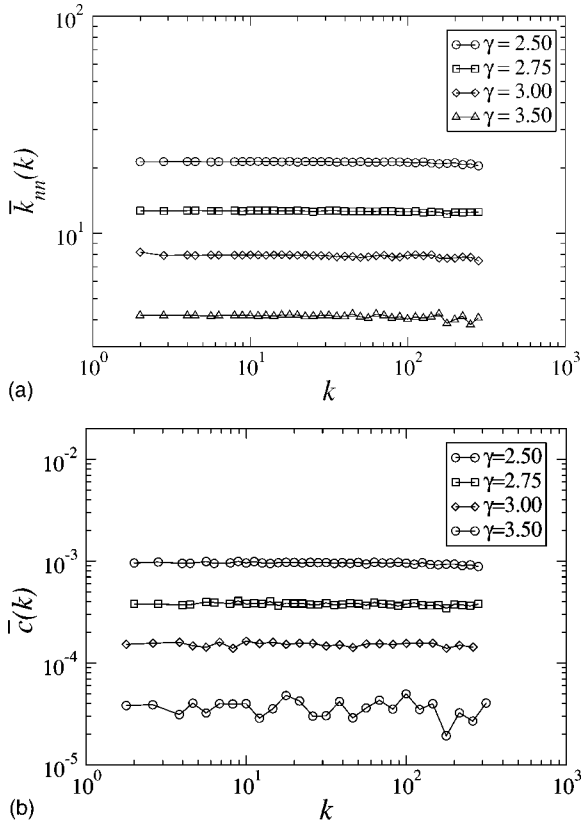

FIG. 2. Average nearest-neighbor degree of vertices of degree k, $\overline{k_{nn}}(k)$ (a), and average clustering coefficient $\overline{c}(k)$ (b) for the UCM algorithm with different degree exponents γ. Network size is $N=10^5$.

work size $N=10^5$. As we can observe, for $\gamma>3$, which corresponds to an effectively bounded degree distribution with finite $\langle k^2\rangle$, both functions are almost flat, signaling an evident lack of correlations. On the other hand, for values $\gamma\leqslant 3$ there is a clear presence of correlations. This correlations have a mixed disassortative nature: vertices with many connections tend to be connected to vertices with few connections, while low-degree vertices connect equally with vertices of any degree.

The origin of this phenomenon can be traced back to the effects of the cutoff (or maximum expected degree) $k_c(N)$ in a network of size N. In fact, it is possible to show that in order to have no correlations in the absence of multiple and self-connections, a scale-free network with degree distribution $P(k)\sim k^{-\gamma}$ and size N must have a cutoff scaling at most as $k_s(N)\sim N^{1/2}$ (the so-called structural cutoff) [26–28]. For a power-law network generated using the CM algorithm defined above (i.e., generating random degrees in the range $m\leqslant k_i<N$), simple extreme value theory arguments show in fact that

$$k_c(N)\sim N^{1/(\gamma-1)}. \tag{6}$$

For $\gamma<3$, we have that $k_c(N)>N^{1/2}$ and therefore it is impossible to avoid the presence of correlations. Only for the particular case $\gamma\geqslant 3$ do we recover $k_c(N)\leqslant N^{1/2}$, which explains the lack of correlations observed in Fig. 1 for $\gamma=3.5$.

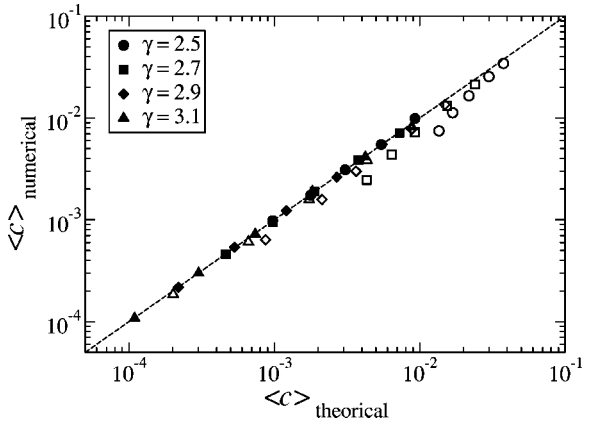

FIG. 3. Numerical average clustering coefficient $\langle c\rangle$ as a function of the corresponding theoretical value, given by Eq. (5), for the CM (open symbols) and the UCM (solid symbols) algorithms. The different points for each value of γ represent different network sizes $N=10^3$, 3×10^3, 10^4, 3×10^4, and 10^5.

Since it is the maximum possible value of the degrees in the network that rules the presence or absence of correlations in a random network with no multiple or self-connections, we propose the following *uncorrelated configuration model* (UCM) in order to generate random uncorrelated scale-free networks.

(i) Assign to each vertex i, in a set of N initially disconnected vertices, a number k_i of stubs, where k_i is drawn from the probability distribution $P(k)\sim k^{-\gamma}$ and subject to the constraints $m\leqslant k_i\leqslant N^{1/2}$ and $\Sigma_i k_i$ even.

(ii) Construct the network by randomly choosing stubs and connecting them to form edges, respecting the preassigned degrees and avoiding multiple and self-connections.

This algorithm can be implemented in practice as follows [29]: Once the degree k_i is assigned, a list of $\Sigma_i k_i$ elements is created, containing k_i copies of the ith vertex. A pair of elements in this list is randomly chosen to create an edge. If the elements are equal or correspond to an already existing edge, they are discarded and a new pair is drawn. Otherwise, the edge is accepted and the list is updated, eliminating the elements corresponding to the newly created edge. This procedure is iterated until all elements in the list are exhausted. The constraint on the maximum possible degree of the vertices ensures that $k_c(N)\sim N^{1/2}$, allowing for the possibility to construct uncorrelated networks. As an additional numerical optimization of this algorithm, we also impose the minimum degree $m=2$ to generate connected networks with probability one [25,30].

In Fig. 2 we check for the presence of correlations in the UCM for scale-free networks. As we can observe, both correlation functions show an almost flat behavior for all values of the degree exponent γ, compatible with the lack of correlations at the two and three vertex levels.

We have additionally explored the validity of the expression for the average clustering coefficient [31] $\langle c\rangle$, defined as

$$\langle c\rangle=\sum_k P(k)\overline{c}(k), \tag{7}$$

which, for random uncorrelated networks, takes the form given by Eq. (5). For scale-free networks with a general cut-

off $k_c(N)$, we have that, in the large-N limit, $\langle k^2\rangle \sim k_c(N)^{3-\gamma}$. Therefore, for random networks generated with the classical CM model, in which $k_c(N) \sim N^{1/(\gamma-1)}$, we have that $\langle c\rangle_{\mathrm{CM}} \sim N^{(7-3\gamma)/(\gamma-1)}$. This expression is clearly anomalous for $\gamma < 7/3$, since it leads to a diverging clustering coefficient for large N, while, by definition, this magnitude, being a probability, must be smaller than 1. This anomaly vanishes in the UCM prescription. In this case, we have that $k_c(N) \sim N^{1/2}$ for any value of γ, leading to $\langle c\rangle_{\mathrm{UCM}} \sim N^{2-\gamma}$, which is a decreasing function of the network size for any $\gamma > 2$.

In Fig. 3 we plot the average clustering coefficient obtained from numerical simulations of the CM and UCM algorithms as a function of the theoretical value, Eq. (5), for different values of γ and different network sizes N. We can observe that, while the results for the uncorrelated UCM nicely collapses onto the diagonal line in the plot, meaning that the numerical values are almost equal to their theoretical counterparts, noticeable departures are observed for the implicitly correlated CM algorithm.

To sum up, in this Brief Report we have presented a model to generate uncorrelated random networks with no multiple and self-connections and arbitrary degree distribution. The lack of correlations is especially relevant for the case of scale-free networks. In this case, our algorithm is capable to generate networks with flat correlation functions and an average clustering coefficient in good agreement with theoretical predictions. Our algorithm is potentially interesting in order to check the accuracy of the many analytical solutions of dynamical processes taking place on top of complex networks, which are usually found in the uncorrelated limit and, which, up to now, lacked a proper benchmark to check the results for degree exponents $\gamma < 3$.

This work has been partially supported by EC-FET Open Project No. IST-2001-33555. R.P.-S. acknowledges financial support from the Ministerio de Ciencia y Tecnología (Spain) and from the Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya (Spain). M.B. acknowledges financial support from the Ministerio de Ciencia y Tecnología through the Ramón y Cajal program. M.C. acknowledges financial support from Universitat Politècnica de Catalunya.

[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 University Press, Oxford, 2003).
[4] B. Bollobás, *Modern Graph Theory* (Springer-Verlag, New York, 1998).
[5] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[6] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **86**, 3682 (2001).
[7] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).
[8] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[9] R. M. May and A. L. Lloyd, Phys. Rev. E **64**, 066112 (2001).
[10] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[11] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[12] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E **65**, 066130 (2002).
[13] E. Ravasz and A.-L. Barabási, Phys. Rev. E **67**, 026112 (2003).
[14] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **68**, 036112 (2003).
[15] S. N. Dorogovtsev, Phys. Rev. E **69**, 027104 (2004).
[16] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **66**, 047104 (2002).
[17] 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.
[18] Z. Burda, J. Jurkiewicz, and A. Krzywicki, Phys. Rev. E **70**, 026106 (2004).
[19] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, Eur. Phys. J. B **26**, 521 (2002).
[20] S. Maslov, K. Sneppen, and A. Zaliznyak, Physica A **333**, 529 (2004).
[21] J. Park and M. E. J. Newman, Phys. Rev. E **68**, 026112 (2003).
[22] A. Bekessy, P. Bekessy, and J. Komlos, Stud. Sci. Math. Hung. **7**, 343 (1972).
[23] E. A. Bender and E. R. Canfield, J. Comb. Theory, Ser. A **24**, 296 (1978).
[24] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).
[25] M. Molloy and B. Reed, Combinatorics, Probab. Comput. **7**, 295 (1998).
[26] F. Chung and L. Lu, Ann. Combinatorics **6**, 125 (2002).
[27] Z. Burda and A. Krzywicki, Phys. Rev. E **67**, 046118 (2003).
[28] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Eur. Phys. J. B **38**, 205 (2004).
[29] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman, and S. H. Strogatz, Phys. Rev. E **64**, 041902 (2001).
[30] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).
[31] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).