Reshuffling scale-free networks: From random to assortative

R. Xulvi-Brunet and I. M. Sokolov

Institut für Physik, Humboldt Universität zu Berlin, Newtonstraße 15, D-12489 Berlin, Germany (Received 5 May 2004; revised manuscript received 13 July 2004; published 2 December 2004)

To analyze the role of assortativity in networks we introduce an algorithm which produces assortative mixing to a desired degree. This degree is governed by one parameter p. Changing this parameter one can construct networks ranging from fully random (p=0) to totally assortative (p=1). We apply the algorithm to a Barabási-Albert scale-free network and show that the degree of assortativity is an important parameter governing the geometrical and transport properties of networks. Thus, the average path length of the network increases dramatically with the degree of assortativity. Moreover, the concentration dependences of the size of the giant component in the node percolation problem for uncorrelated and assortative networks are strongly different. The behavior of the clustering coefficient is also discussed.

DOI: 10.1103/PhysRevE.70.066102 PACS number(s): 05.50.+q, 89.75.Hc

INTRODUCTION

Complex networks have recently attracted a burst of interest as an indispensable tool for a description of different complex systems. Thus, technological webs such as the Internet and World Wide Web, as well as other natural and social systems like intricate chemical reactions in the living cell, the networks of scientific and movie actor collaborations, and even human sexual contacts, have been successfully described through scale-free networks, networks with the degree distribution $P(k) \sim k^{-\gamma}$ [1,2]. The degree distribution P(k) is one of the essential measures used to capture the structure of a network and gives the probability that a node chosen at random is connected with exactly k other vertices of the network.

Recently, it was pointed out that the existence of degree correlations among nodes is an important property of real networks [3-15]. Thus, many social networks show that nodes having many connections tend to be connected with other highly connected nodes [4,6]. In the literature this characteristic is usually denoted as assortativity or assortative mixing. On the other hand, technological and biological networks often have the property that nodes with a high degree are preferably connected with ones with a low degree, a property referred to as dissortativity [3,7]. Such correlations have an important influence on the topology of networks, and therefore they are essential for the description of spreading phenomena, like spreading of information or infections, as well as for the robustness of networks against an intentional attack or random breakdown of their elements [16-21].

In order to assess the role of correlations, especially of assortative mixing, several authors have proposed procedures to build correlated networks [3,23–25]. The most general procedures are the ones proposed by Newman [3] and Boguñá and Pastor-Satorras [25], who suggest two different ways to construct general correlated networks with prescribed correlations. Following the same goal, we however adopt a different perspective in this paper. We propose a simple algorithm producing assortative mixing, in which, instead of putting in correlations by hand, we only try to im-

pose the intuitive condition that "nodes with similar degree connect preferably." We then investigate the correlations which come out of our simple model. Thus, we present an algorithm, governed by only the parameter p, capable of generating assortative correlations to a desired degree. In order to study the effect of assortative mixing, we apply our algorithm to a Barabási-Albert scale-free network [26], the one leading to the degree distribution $P(k) \sim k^{-3}$, and investigate the properties of the emerging networks in some detail. The idea behind this work is to check what are the effects of assortativity alone on the properties of the networks, which stay random in any other respect. We show, for example, that the assortative correlations change drastically their average path length and strongly influence their percolation properties.

ALGORITHM

In what follows we treat undirected networks. Starting from a given network, at each step two links of the network are chosen at random, so that the four nodes, in general with different degrees, connected through the links two by two are considered. The step of our algorithm looks as follows. The four nodes are ordered with respect to their degrees. Then, with probability p, the links are rewired in such a way that one link connects the two nodes with the smaller degrees and the other connects the two nodes with the larger degrees; otherwise, the links are randomly rewired (Maslov-Sneppen algorithm [11]). In the case when one or both of these new links already existed in the network, the step is discarded and a new pair of edges is selected. This restriction prevents the appearance of multiple edges connecting the same pair of nodes. A repeated application of the rewiring step leads to an assortative version of the original network. Note that the algorithm does not change the degree of nodes involved and thus the overall degree distribution in the network. Changing the parameter p, it is possible to construct networks with different degree of assortativity.

CORRELATIONS AND ASSORTATIVITY

Let \mathcal{E}_{ij} be the probability that a randomly selected edge of the network connects two nodes, one with degree i and an-

other with degree j. The probabilities \mathcal{E}_{ij} determine the correlations of the network. We say that a network is uncorrelated when

$$\mathcal{E}_{ij} = (2 - \delta_{ij}) \frac{iP(i)}{\langle i \rangle} \frac{jP(j)}{\langle j \rangle} := \mathcal{E}_{ij}^r, \tag{1}$$

i.e., when the probability that a link is connected to a node with a certain degree is independent from the degree of the attached node. Here $\langle i \rangle = \langle j \rangle$ denotes the first moment of the degree distribution.

Assortativity means that highly connected nodes tend to be connected to each other with a higher probability than in an uncorrelated network. Moreover, the nodes with similar degrees tend to be connected with larger probability than in the uncorrelated case, i.e., $\mathcal{E}_{ii} > \mathcal{E}_{ii}^r \ \forall \ i$. The degree of assortativity of a network can thus be characterized by the quantity [3]

$$\mathcal{A} = \frac{\sum_{i} \mathcal{E}_{ii} - \sum_{i} \mathcal{E}_{ii}^{r}}{1 - \sum_{i} \mathcal{E}_{ii}^{r}},\tag{2}$$

which takes the value 0 when the network is uncorrelated and the value 1 when the network is totally assortative. (Note that finite-size effects and the constraint that no vertices be connected by more than one edge bound \mathcal{A} from above by the values lower than 1 [22].)

Now, starting from the algorithm generator, we can obtain a theoretical expression for \mathcal{E}_{ij} as a function of p. Let E_{ij} be the number of links in the network connecting two nodes, one with degree i and another with degree j, so that $\mathcal{E}_{ij} = E_{ij}/L$, where L is the total number of links of the network. (Since undirected networks satisfy $E_{ij} = E_{ji}$, the restriction $i \leq j$ can be imposed without loss of generality.) We now define the variable

$$F_{ln} = \sum_{r=l}^{n} \sum_{s=r}^{n} E_{rs}, \ r \le s, \ l \le n.$$
 (3)

A careful analysis of the algorithm reveals that, every time the rewiring procedure is applied, F_{ln} either does not change or increases or decreases by unity. We can then calculate the probabilities that it changes—i.e., that $F_{ln} \rightarrow F_{ln} + 1$ or $F_{ln} \rightarrow F_{ln} - 1$. The effect of multiple edges can be disregarded since they are rare in the thermodynamical limit. Taking into account all corresponding possibilities, we obtain for the probabilities of changes the following expressions:

$$(X_{ln}-f_{ln})^2+p(X_{ln}-f_{1n}+f_{1,l-1})^2$$

for $F_{ln} \rightarrow F_{ln} + 1$ and

$$f_{ln}[(1-p)(1-2X_{ln})+p(X_{1,l-1}-f_{1,l-1}-f_{1n})+f_{ln}]$$

for $F_{ln} \rightarrow F_{ln} - 1$. Here $f_{ln} = F_{ln}/L$, and X_{ln} is given by

$$X_{ln} = \frac{1}{\langle k \rangle} \sum_{k=1}^{n} kP(k), \ l \le n.$$

(Note that X_{ln} and f_{ln} vanish when one of the indices is smaller than 1, the minimal tolerated degree.) Using this,

we can calculate the expected value for f_{ln} . The process of repeated applications of our algorithm corresponds to an ergodic Markov chain, and the stationary solution in the thermodynamical limit is given by the condition

$$(X_{ln} - f_{ln})^{2} + p(X_{ln} - f_{1n} + f_{1,l-1})^{2}$$

$$= f_{ln} [(1 - p)(1 - 2X_{ln}) + p(X_{1,l-1} - f_{1,l-1} - f_{1n}) + f_{ln}]$$
(4)

for all l > 1. For l = 1 this condition reduces to

$$(1+p)(X_{1n}-f_{1n})^2 = (1-p)f_{1n}[1-2X_{1n}+f_{1n}].$$
 (5)

Using Eqs. (4) and (5) we can calculate f_{ln} . The solution reads

$$f_{ln} = \frac{X_{ln}^2 + (B_n - B_{n-1})^2}{(1 - p)/2 + pX_{ln} + B_n + B_{n-1}}, \ l \le n,$$

with

$$B_n = \sqrt{\left\lceil pX_{1n} + \frac{1-p}{4} \right\rceil^2 - pX_{1n}^2 \left(\frac{1+p}{2}\right)}.$$

Applying the definition, Eq. (3), we obtain the correlations

$$\mathcal{E}_{ii} = f_{ii} - f_{i,i-1} - f_{i+1,i} + f_{i+1,i-1}. \tag{6}$$

Finally, note that Eq. (6) reduces to the corresponding uncorrelated case \mathcal{E}_{ij}^r when p=0, and reduces to

$$\mathcal{E}_{ij} = \delta_{ij} \frac{iP(i)}{\langle i \rangle} \tag{7}$$

for the case p=1.

SIMULATIONS RESULTS

Let us start this section with drawing a small network to show how our algorithm works. The initial network is a Barabási-Albert scale-free construction with only N=200 nodes and L=400 links; see Fig. 1(a). To obtain other networks with exactly the same degree distribution but different degree of assortativity we apply the algorithm discussed. Figure 1 shows the changes in the network with varying parameter p. In the figure we have placed the nodes in such a way that nodes of degree 2 are shown in the left part of each panel, all nodes of degree 3 lie to the right of any node of degree 2, all nodes of degree 4 lie to the right of any node of degree 3, etc. The nodes of the same degree are randomly spread within the corresponding area of the figure to better show the links.

The network of maximal accessible assortativity is shown in Fig. 1(d). In this network almost all nodes with the same degree are linked only between themselves. Figure 1(d) shows that all nodes with degree k=2 form separated clusters (a more careful analysis unveils that there are three "pearl necklace" clusters with N=23, N=30, and N=48 nodes). All nodes with k=3 are linked between themselves except for one, which is linked with a node of connectivity k=4. Note

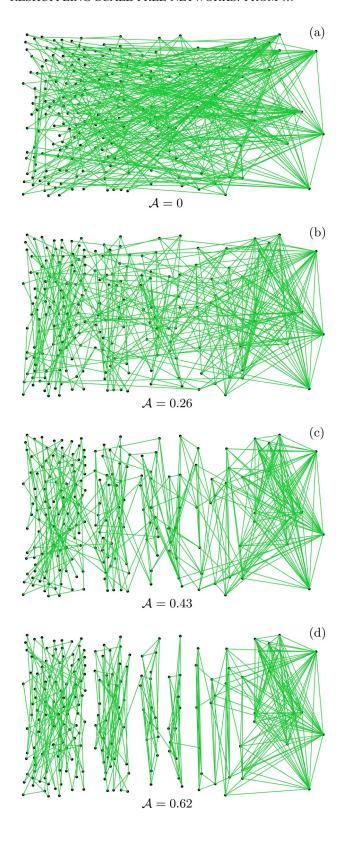


FIG. 1. Scale-free networks for different degrees of assortativity (see text for details). The nodes of the same degree are grouped together; the degree is nondecreasing from left to right. The panels show (a) $\mathcal{A}=0$ (uncorrelated network), (b) $\mathcal{A}=0.26$, (c) $\mathcal{A}=0.43$, and (d) $\mathcal{A}=0.62$ (maximal assortativity).

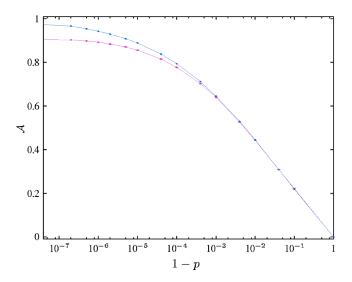


FIG. 2. The lower curve corresponds to the measured assortativity \mathcal{A} of our simulations, whereas the upper curve corresponds to the theory. We note that both curves coincide for $\mathcal{A} < 0.7$. Above this value the finite-size corrections get important, leading to the measured value of $\mathcal{A} < 1$ for $p \rightarrow 1$.

that since there are N_3 =41 nodes with k=3 in our network their links cannot be redistributed within the set. If this were possible, the overall number of links would be $41 \times 3/2$ =61.5, since each node bears 3 links and each of these links is counted twice in the set. All nodes with degree k=4 form a single cluster, with two outgoing links, one to the cluster of nodes with k=3 and one to a cluster of nodes of connectivity k=5. In fact, the network is not a set of isolated clusters of nodes with the same connectivity only because of restrictions imposed by the given degree distribution. These restrictions are also responsible for the fact that A<1 (for our network the maximal assortativity is A_{max} =0.62).

In this work we apply our algorithm only to the Barabási-Albert construction [26] and just like in our example, with double number of links than of nodes $L=2\dot{N}$. Thus, in the rest of simulations we use $N=10^5$ nodes and $L=2\times10^5$ links. We measure \mathcal{E}_{ij} as functions of p and use them to calculate the corresponding values of \mathcal{A} . All simulation results are averaged over ten independent realizations of the algorithm as applied to the original network.

Figure 2 shows the relation between the parameter p and the coefficient of assortativity \mathcal{A} . The lower curve corresponds to the measured assortativity and the upper to our theoretical prediction. Both curves coincide for $\mathcal{A} < 0.7$. However, whereas the theoretical curve reaches the value 1 for $p \rightarrow 1$, the measured assortativity increases until the maximal value smaller than one $(\mathcal{A} \rightarrow 0.913)$ is reached. This was expected [22] and is due to the finite-size effects mentioned above.

To assess the goodness of Eq. (6) we compare in Fig. 3 the theoretical values of \mathcal{E}_{kk} , given by Eq. (6), with the simulations. The points correspond to the simulations and the curves are the corresponding theoretical results obtained based on the actual degree distribution of a particular realization of the network discussed. We note that the agreement is really excellent.

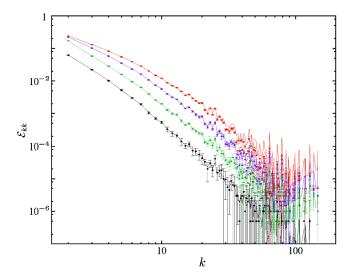


FIG. 3. \mathcal{E}_{kk} as a function of k for different values of \mathcal{A} . From bottom to top: $\mathcal{A}=0$, $\mathcal{A}=0.221$, $\mathcal{A}=0.443$, and $\mathcal{A}=0.640$. The points are the results of the simulations and the curves correspond to the theory.

Average path length. The average path length of a network is the average distance between every pair of vertices of the network, being defined as the number of edges along the shortest path connecting them. Uncorrelated scale-free networks show a very small path length, typically growing as the logarithm of the network's size (small-world behavior). For networks with $N \approx 10^5$ it is about $l \approx 6$. The results of the simulations show that the average path length grows rapidly when the assortativity of the network increases (Fig. 4), so that it becomes two orders of magnitude larger than for the uncorrelated network when the coefficient of assortativity tend to its maximal value. In the inset of Fig. 4 we plot the average path length as a function of $\mathcal{K}-\mathcal{A}$, where $\mathcal{K}=0.913$ corresponds to this maximal value of \mathcal{A} attainable in

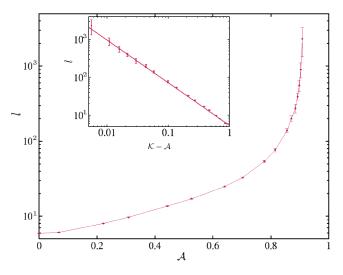


FIG. 4. Average path length l of the network versus coefficient of assortativity. We note that l grows rapidly when \mathcal{A} increases. In the inset the average path length is plotted on double-logarithmic scales as function of $\mathcal{K}-\mathcal{A}$, being $\mathcal{K}=0.913$. The slope of the straight line is -1.12.

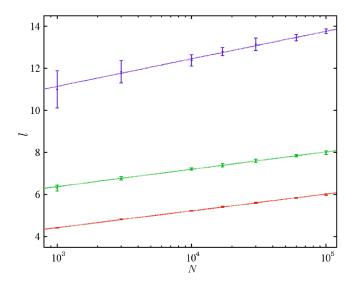


FIG. 5. The average path length l is plotted as function of N for three values of the coefficient of assortativity A. From bottom to top: A=0, A=0.221, and A=0.443. Note the logarithmic scale.

the network. For our particular Barabási-Albert network we thus have $l \propto (0.913 - A)^{-1.12}$.

Although assortative networks present large mean path lengths, they are still small worlds—i.e., show the logarithmic dependence of l on the network's size N. Figure 5 shows this behavior for three different values of \mathcal{A} . The error bars result from averaging over ten realizations of the algorithm. This small-world behavior is preserved for all tested values of $\mathcal{A} \leq 0.6$ (this maximal value of \mathcal{A} is still larger than the ones found in real assortative networks, where \mathcal{A} ranges between 0 and 0.4). Thus, assortative networks are the "large" small worlds.

Natural networks, like the different coauthorship networks (physics, biology, mathematics, etc.), the film actor collaboration network, etc. (all of them assortative networks), seem to show somewhat smaller average path lengths than the ones found here [1,3]. We attribute this finding to the fact that the mean degree of such networks is 2–4 times larger than in our case ($\langle k \rangle = 4$). Therefore one has to be cautious about comparing absolute numerical values.

Clustering coefficient. Clustering coefficients of a network are a measure of the number of loops (closed paths) of length 3. The notion has its roots in sociology, where it was important to analyze the groups of acquaintances in which every member knows every other one. To discuss the concept of clustering, let us focus first on a vertex, having k edges connected to k other nodes termed as nearest neighbors. If these nearest neighbors of the selected node were forming a fully connected cluster of vertices, there would be k(k)-1)/2 edges between them. The ratio between the number of edges that really exist between these k vertices and the maximal number k(k-1)/2 gives the value of the clustering coefficient of the selected node. The clustering coefficient of the whole network C is then defined as the average of the clustering coefficients of all vertices. One can also speak about the clustering coefficient of nodes with a given degree k, referring to the average of the clustering coefficients only

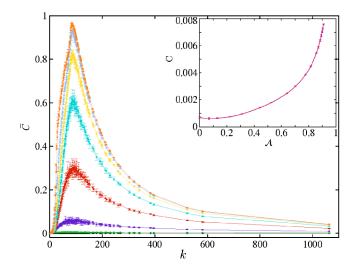


FIG. 6. $\bar{C}(k)$ as a function of the degree of nodes, k. Different curves correspond to different values of \mathcal{A} . From bottom to top: $\mathcal{A}=0$, $\mathcal{A}=0.069$, $\mathcal{A}=0.221$, $\mathcal{A}=0.443$, $\mathcal{A}=0.640$, $\mathcal{A}=0.777$, $\mathcal{A}=0.856$, and maximal assortativity $\mathcal{A}=0.913$. Inset: clustering coefficient C versus the degree of assortativity, \mathcal{A} .

over this type of nodes. We shall denote this degree-dependent clustering coefficient by $\bar{C}(k)$, to distinguish it from C.

Figure 6 shows the variation of both clustering coefficients with the assortativity of the network. The clustering coefficient C increases with the assortativity (inset of the figure). However, typical values of the clustering coefficients found in our simulations are still much smaller than the ones observed in real networks ($C \ge 0.1$) [1]. The last ones might, however, have a much more intricate structure, partly governed by the metrics of the underlying space, as in the models discussed in [27]. Thus, care must be exercised when applying our results to natural networks.

The variation of $\bar{C}(k)$ shows more interesting features. The simulations show a peak around k=90 (probably a finite-size effect) whose height increases with the assortativity of the network. In the uncorrelated case $\bar{C}(k)$ does not depend on k [13], but a strong tendency to clustering (for relatively large k) emerges when A grows. We also observe in our simulations that $\bar{C}(k=2)=0$ when $A \approx 1$ (k=2 corresponds to the minimal degree of the vertices). This is not surprising since in a strongly assortative case almost all nodes with degree k=2 are connected between themselves, forming one or several large loops of length larger than 3. This means that all nodes having this minimal degree (in our simulations half of the total number of vertices) do not tend to contribute to the clustering coefficient C.

In the present contribution we concentrate on an investigation of the properties of the proposed algorithm. However, we suggest, in relation to real networks, a simple modification of the algorithm that perhaps could be useful. Thus, in order to generate assortativity only among highly connected vertices, one can apply the algorithm above only when at least one of the four nodes selected at the corresponding step has a degree larger than some chosen *k*. Provided all four

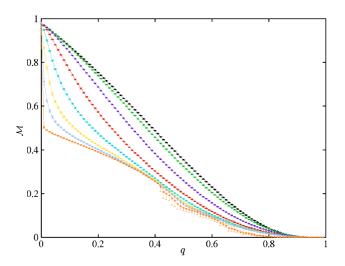


FIG. 7. Fraction of nodes, \mathcal{M} , in the giant component depending on the fraction of nodes removed from the network. The graph compares the results for different degrees of assortativity. From top to bottom: $\mathcal{A}=0$, $\mathcal{A}=0.069$, $\mathcal{A}=0.221$, $\mathcal{A}=0.443$, $\mathcal{A}=0.640$, $\mathcal{A}=0.777$, $\mathcal{A}=0.856$, and $\mathcal{A}=0.913$ (maximal assortativity).

nodes have a smaller degree, only the Maslov-Sneppen step is used. This procedure could lead to a larger value for the clustering coefficient.

Node percolation. Node percolation corresponds to the removal of a certain fraction of vertices from the network and is relevant when discussing their vulnerability to a random attack. Let q be the fraction of nodes removed. At a critical fraction q_c , the giant component (largest connected cluster) breaks into isolated clusters. Figure 7 shows the fraction of nodes, \mathcal{M} , in the giant component as a function of q for different degrees of assortativity of the network. The four upper curves correspond to the values of assortativity found in natural networks. We note that the behavior of $\mathcal{M}(q)$ changes gradually with A from the uncorrelated case (upper curve) to a quite different behavior when $A \rightarrow 1$ (lower curve), which indicates a very different topology in the network when it is strongly assortative. However, although the particular form of the \mathcal{M} dependence is different for different degrees of assortativity, the absence of the transition at finite concentrations $(q_c=1)$ and the overall type of critical behavior for correlated networks with the same P(k) seems to be the same as for uncorrelated networks-namely, the one discussed in Refs. [28,29]. We thus see that this generic behavior in node percolation is only quantitavely affected by resfuffling, lowering \mathcal{M} at a given fraction of removed nodes. This quantitative behavior, however, might depend on the network's precise nature which fact has to borne in mind when comparing our results with the ones for natural networks. We also point out that in the case $A \simeq 1$, a finite network is no longer fully connected: part of the nodes does not belong to the giant component even for q=0. The results suggest that, in the thermodynamical limit, the giant cluster at $q \rightarrow 0$ contains around a half of all nodes and that its density then decays smoothly with q.

CONCLUSIONS

In summary, we present an algorithm to generate assortatively correlated networks. In the thermodynamical limit we

obtain a theoretical expression for the generated correlations, which only depend on the degree distribution of the network and on the tunable parameter p of the algorithm. Finally, we show that assortative correlations have a drastic influence on the statistical properties of networks, changing strikingly their average path length and percolation properties, as well as leading to an increase in their clustering coefficient.

We also indicate that with a minor change in our algorithm one can produce dissortative mixing too. The only change would be the following: after ordering the nodes with respect to their degree, one rewires, with probability p, the

edges so that one link connects the highest connected node with the node with the lowest degree and the other link connects the two remaining vertices; with probability 1-p one rewires the links randomly.

ACKNOWLEDGMENTS

Useful discussions with Professor S. Havlin are gratefully acknowledged. I.M.S. would like to thank the Fonds der Chemischen Industrie for partial financial support.

- [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] M. E. J. Newman, Phys. Rev. E 67, 026126 (2003).
- [4] M. E. J. Newman, Phys. Rev. Lett. 89, 208701 (2002).
- [5] A. Vázquez, M. Boguñá, Y. Moreno, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E 67, 046111 (2003).
- [6] A. Capocci, G. Caldarelli, and P. De Los Rios, Phys. Rev. E 68, 047101 (2003).
- [7] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. 87, 258701 (2001).
- [8] M. E. J. Newman and J. Park, Phys. Rev. E 68, 036112 (2003).
- [9] J. Berg and M. Lässig, Phys. Rev. Lett. 89, 228701 (2002).
- [10] K.-I. Goh, E. Oh, B. Kahng, and D. Kim, Phys. Rev. E 67, 017101 (2003).
- [11] S. Maslov and K. Sneppen, Science 296 910 (2002).
- [12] P. L. Krapivsky and S. Redner, Phys. Rev. E 63 066123 (2001).
- [13] S. N. Dorogovtsev, Phys. Rev. E 69, 027104 (2004).
- [14] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman, and S. H. Strogatz, Phys. Rev. E 64, 041902 (2001).
- [15] J. Park and M. E. J. Newman, Phys. Rev. E 68, 026112 (2003).
- [16] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **90**, 028701 (2003).

- [17] V. M. Eguíluz and K. Klemm, Phys. Rev. Lett. 89, 108701 (2002).
- [18] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E 66, 047104 (2002).
- [19] N. Schwartz, R. Cohen, D. ben-Avraham, A.-L. Barabási, and S. Havlin, Phys. Rev. E 66, 015104 (2002).
- [20] A. Vázquez and Y. Moreno, Phys. Rev. E 67, 015101 (2003).
- [21] Y. Moreno, J. B. Gómez, and A. F. Pacheco, Phys. Rev. E 68, 035103 (2003).
- [22] S. Maslov, K. Sneppen, and A. Zaliznyak, Physica A 333, 529 (2004).
- [23] A. Ramezanpour, V. Karimipour, and A. Mashaghi, Phys. Rev. E 67, 046107 (2003).
- [24] R. Xulvi-Brunet, W. Pietsch, and I. M. Sokolov, Phys. Rev. E 68, 036119 (2003).
- [25] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E 68, 036112 (2003).
- [26] A.-L. Barabási and R. Albert, Science 286, 509 (1999).
- [27] L. M. Sander, C. P. Warren, and I. M. Sokolov, Phys. Rev. E 66, 056105 (2002).
- [28] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. 85, 4626 (2000).
- [29] R. Cohen, D. ben-Avraham, and S. Havlin, Phys. Rev. E **66**, 036113 (2002).