Local versus global knowledge in the Barabási-Albert scale-free network model

Jesús Gómez-Gardeñes[1,2] and Yamir Moreno[2]
[1]*Departamento de Teoría y Simulación de Sistemas Complejos, Instituto de Ciencia de Materiales de Aragón, C.S.I.C.—Universidad de Zaragoza, Zaragoza 50009, Spain*
[2]*Instituto de Biocomputación y Física de Sistemas Complejos, Universidad de Zaragoza, Zaragoza 50009, Spain*
(Received 30 October 2003; published 31 March 2004)

The scale-free model of Barabási and Albert (BA) gave rise to a burst of activity in the field of complex networks. In this paper, we revisit one of the main assumptions of the model, the preferential attachment (PA) rule. We study a model in which the PA rule is applied to a neighborhood of newly created nodes and thus no global knowledge of the network is assumed. We numerically show that global properties of the BA model such as the connectivity distribution and the average shortest path length are quite robust when there is some degree of local knowledge. In contrast, other properties such as the clustering coefficient and degree-degree correlations differ and approach the values measured for real-world networks.

DOI: 10.1103/PhysRevE.69.037103 PACS number(s): 89.75.Fb, 05.70.Jk

During the last several years, many scientists have scrutinized the world around us to unravel the complex patterns of interconnections that characterize seemingly diverse social [1], biological [2,3], and technological systems [4,5]. These systems have been shown to exhibit common features that can be captured using the tools of graph theory or in more recent terms, network modeling. At the same time, network models of diverse kinds have been proposed with the aim of describing and explaining the properties of real webs [6,7]. It turns out that most real networks are better described by growing models in which the number of nodes (or elements) forming the net increases with time and that the probability that a given node has k connections to other nodes follows a power law $P_k \sim k^{-\gamma}$, with $\gamma \leqslant 3$. Additionally, the study of processes taking place on top of these networks has led us to reconsider classical results obtained for regular lattices or random graphs due to the radical changes of the system's dynamics when the heterogeneity of complex networks cannot be neglected [8–11].

The first scale-free (SF) network model, introduced by Barabási and Albert (BA), postulated that there are two fundamental ingredients of many real networks [12,13]: their growing character and the preferential attachment (PA) rule. The preferential attachment rule considers that the probability that an old node links to newly added nodes is proportional to its degree k. It summarizes the common belief that the richer you are, the more likely it is that your richness grows; that is why the term rich-gets-richer has been used to refer to the PA rule [13]. However, the BA model assumes that one knows the connectivity of all nodes when a new node links to the network. This is clearly an unrealistic assumption. This drawback of the model construction has not passed unnoticed and many models have been introduced to produce scale-free networks and to test whether or not the basic assumptions of the BA recipe are necessary conditions to build up these networks [6,7].

Growing models which produce scale-free graphs with arbitrary γ exponents, and nonrandom correlations can be found nowadays in the scientific literature. On the other hand, there are some models in which the PA rule is limited to a neighborhood due to geographic constraints [14], or where its linear character is investigated [15]. Recently, Caldarelli *et al.* [16] have shown that one can produce SF networks without assuming preferential attachment at all. As a by-product, other properties of the network fit well with those of real-world graphs. They introduced an intrinsic fitness model in which two nodes are connected with a probability that depends on their fitness. Note, additionally, that the way in which the fitness parameter was introduced is different from the model in Ref. [17].

In this paper, we adopt a different perspective. Our aim is to test to what extent the global character of the PA rule in the original BA model is important. We introduce a model in which the PA is applied only to a neighborhood of the newly added node depending on the value of a variable which measures the affinity between different nodes. By going down from the BA limit of the model to the the limit where all nodes are distinct, we test to what extent the global knowledge of each node's connectivity is fundamental to get a scale-free graph. Through numerical simulations we find that in a wide range of the model parameters, average quantities such as the connectivity distribution and the shortest path length are not affected by the use of local knowledge of the network, whereas other properties such as the clustering coefficient are more sensitive to local details.

Our model is defined in two layers. The first discriminates among all the nodes by assigning to each node at the moment of its creation a parameter a_i which measures how close or distinct a given node is from the rest of the elements that compose the network. Then, we apply the preferential attachment rule in the neighborhood defined by nodes with common affinities. Specifically, the network is constructed by repeated iteration of the following rules.

(i) Start from a small core of nodes, m_o, linked together. Assign to each of these m_o nodes a random affinity a_i taken form a probability distribution. In what follows, we will use for simplicity a uniform distribution between $(0,1)$.

(ii) At each time step, a new node j with a random affinity a_j is introduced and linked to m nodes already present in the network according to the rules specified below.

(iii) Search through all nodes of the network verifying whether or not the condition $a_i - \mu \leqslant a_j \leqslant a_i + \mu$ is fulfilled,

where μ is a parameter that controls the affinity tolerance of the nodes. The nodes that satisfy the affinity condition are grouped in a set A as potential candidates to gain new links.

(iv) Apply the preferential attachment rule to the set A [18], i.e., when choosing the nodes to which the new vertex links, we impose that the probability that vertex i connects to the new node depends on its connectivity such that

$$\Pi(k_i) = \frac{k_i}{\sum_{s \in A} k_s}. \quad (1)$$

(v) Repeat steps (ii)–(iv) such that the final size of the network is $N = m_o + t$.

Thus, after t time steps a network made up of N nodes builds up. It is worth mentioning that the inclusion of the affinity parameter a is not a mere artifact. Indeed, most real systems are formed by nonidentical elements and thus it is natural to assume that although a given node could have a large connectivity, a newly created element will not link to that node because they have very little in common. This feature is clearly manifested in social networks such as the WWW—where individuals bookmark different web pages according to their ''affinity''—or the scientist citation network [19]. In this way, it is very unlikely to find a citation in a condensed matter paper referring to a paper written by a psychologist. Additionally, the same argument can be translated to biological networks such as predator-prey webs or protein-protein interaction networks.

Obviously, when μ is large enough as to dilute the first layer of the model, we recover the BA model. The problem then consists of determining to what extent the local preferential attachment will give the same results, or in other words, does the knowledge of the entire network substantially contribute to the properties observed in the BA networks?

We have performed extensive numerical simulations of the model described in the preceding section. In all cases, the numerical results have been obtained after averaging over at least 500 iterations varying the system size from 10^3 up to 1.2×10^4 nodes. We first generate the BA network by setting the parameter μ to its maximum value such that the preferential attachment applies to the entire set of nodes and then tune μ in order to systematically reduce its value and therefore the size of the set A to which the second choice Eq. (1) is applied.

Figure 1 shows the number of nodes with connectivity k for several values of μ. It turns out that irrespective of the range to which the preferential attachment is applied the stationary probability of having a node with connectivity k is the same as for the BA model, namely, $P_k \sim k^{-\gamma}$ with $\gamma \approx 3$. This result could be intuitively understood by noting that the rules for the network generation have been changed only at a local level, but seeing from a global perspective the average properties should not change radically. To realize this point, think of the network as being made up of different small components, as given by the affinity constraint, each of which is constructed following the BA algorithm. It is then

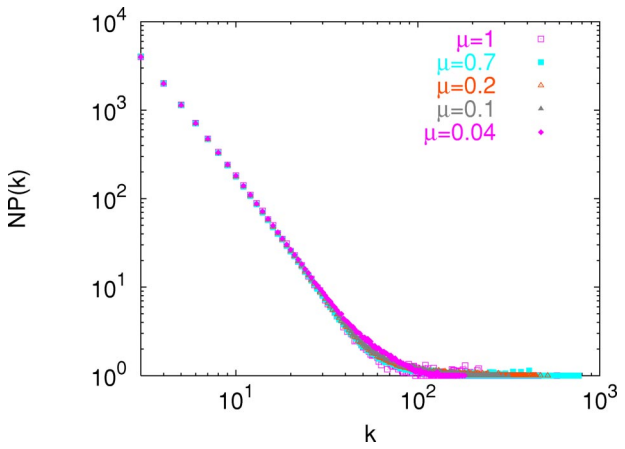

FIG. 1. (Color online) Number of nodes with connectivity k for different values of μ. The size of the network is $N = 10^4$ nodes and $m_o = m = 3$. The power-law distribution has an exponent equal to 3. Note that the BA limit corresponds to $\mu = 1$.

clear that for large system sizes, each graph will follow the power-law distribution $P_k \sim k^{-3}$ and so will be for the entire network.

The above argument applies only to average global properties, but there is nothing that guarantees *a priori* that the components of the network will link together in such a way that other properties will not be affected. This is the case of the average shortest path length L. The average shortest path length of a graph is defined as the minimum number of nodes one has to pass by to go from one node of the network to another randomly chosen node averaged over all possible pairs of nodes. Complex networks show the noticeable property, known as small-world property, that the average path length increases only with the logarithm of its size. We expect that for high values of μ the network is composed by a unique giant component and no fragmentation arises. When the range to which the affinity criterion is applied decreases, the network will gradually lose its compactness and will stretch approaching a one-dimensional structure with some small components. Further reduction of μ provokes the breakdown of the network in many isolated clusters.

Figures 2 and 3 substantiate this picture. Figure 2 represents the ratio between the average path length obtained for different values of μ and that of the BA network, for several system sizes. As μ restricts the PA range, the network undergoes a transition characterized by a growth of $L(\mu)$ and eventually becomes fragmented giving rise to an infinite shortest path length. We note here that although the results shown in the figure have been obtained for a uniform distribution of affinity values a_i, the qualitative behavior does not change for other probability distributions and only the value at which the transition is observed slightly shifts to the right. The shape of the network as the parameter μ is varied can be observed in Fig. 3, where we have represented how the network looks like for the limiting values of μ. It is clear that when the PA range reduces too much, the structure of the network radically changes while keeping the same degree distribution.

We now focus our attention on other properties with a

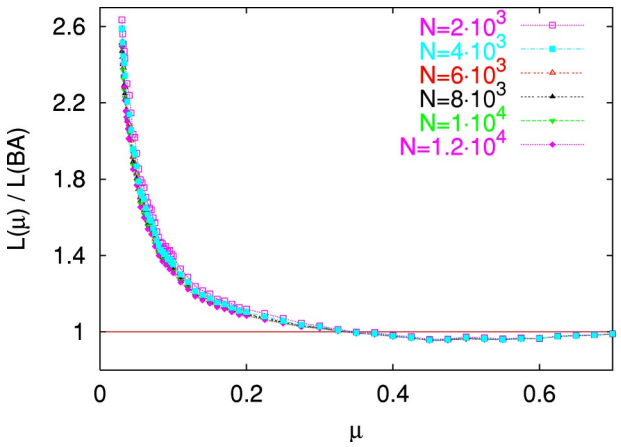

FIG. 2. (Color online) Ratio between the average shortest path length for different μ values, $L(\mu)$, and that of the BA network $[L(1)]$ for several system sizes. The horizontal line marks the BA limit. A transition from graphs fulfilling the small-world property to a regime in which networks break down in many small pieces raising the value of $L(\mu)$ is observed. See the text for further details.

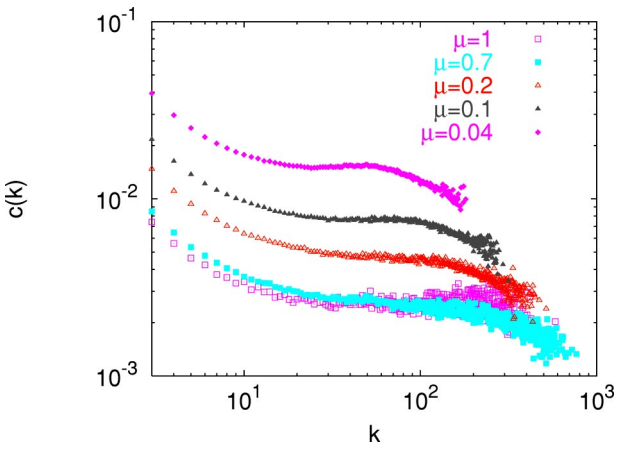

FIG. 4. (Color online) Average clustering coefficient c_k of nodes with degree k for five different values of the parameter μ. Note that as μ decreases, the clustering coefficient departs from the BA limit ($\mu=1$). The parameters used for the generation of the networks are as in Fig. 1.

local character. This is the case of the clustering coefficient c_i. The clustering coefficient of a node i is defined as the ratio between the number of edges e_i among the k_i neighbors of i and its maximum possible value, $k_i(k_i-1)/2$, i.e., $c_i = 2e_i/k_i(k_i-1)$. In this way, the average clustering coefficient c is given by the average of c_i over all nodes of the network. The clustering coefficient is of local character as it gives the probability that two nodes with a common neighbor are also linked together. Thus, it is expected that this magnitude, in our model, depends on the affinity of each node and the range of preferential attachment given by μ. Figure 4 shows the average clustering coefficient of nodes with a given connectivity k, for different values of the parameter μ. The BA limit exhibits almost no correlations with the degree k of the vertices and the smallest value for the clustering coefficient. As μ is reduced, the first selection of nodes by their affinity values plays a more dominant role contributing to the rising of c_i for small and large connectivities. Near the transition, $\mu\sim0.04$, the average coefficient is about one order of magnitude greater than that of the BA network.

Recently, a lot of attention has been given to network motifs [20,21], which can be defined as graph components that are observed in a given network more frequently than in a completely random graph with identical P_k. Triangles and rectangular loops are among these graph components, also known as cycles. They are important because they express the degree of redundancy and multiplicity of paths among nodes in the topology of the network. The results obtained for c_k indicate that as the region where the PA applies is reduced, the number of cycles increases and nonrandom correlations arise. This is illustrated in Fig. 5, where the average nearest neighbor degree, $k_{NN}(k)$ of a node with connectivity k is depicted. While the BA model exhibits no correlations, it manifests the tendency that networks generated with small values of μ display disassortative mixing at both ends of the connectivity range.

In this paper, we have studied a version of the Barabási and Albert scale-free model that allows to tune the range to which the preferential attachment is applied. The model con-

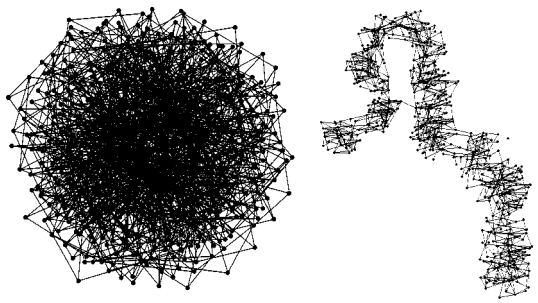

FIG. 3. Graph representations of two networks produced with different values of μ. From left to right, $\mu=1$, and $\mu=0.04$. Each network is made up of $N=500$ nodes.

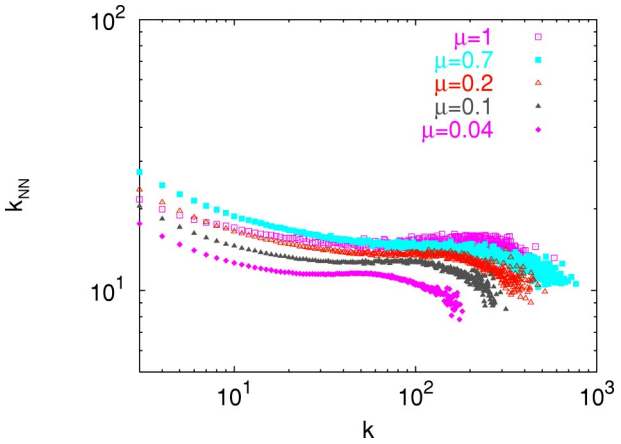

FIG. 5. (Color online) Average nearest neighbor connectivity k_{NN} against k for several values of μ. Results are averaged over 100 network realizations for each μ value. Other parameters are as in Fig. 1.

siders that all nodes are different such that they are, in principle, unable to link to very distinct nodes. By introducing an affinity selection before applying the preferential attachment rule, we tested whether or not the knowledge of the entire network is an essential requisite to get scale-free networks. Our results seem to support the idea that having at least some degree of preferential attachment is enough to get an SF growing network. We found that the connectivity distribution is not affected by the affinity constraints while the network is unable to link together if the tolerance range is reduced too much. On the other hand, local properties such as the clustering coefficient do change and reach values higher than those expected for random networks with the same degree distribution. However, the growth of the clustering coefficient due to the differentiation of nodes produces at the same time a rising in the value of the average shortest path length. Eventually the network breaks down in small pieces and loses its small-world character.

Finally, we point out that although the values found for several magnitudes cannot be directly associated with real data, there are some regions of the parameter space μ where nontrivial properties arise. In this sense, it would be interesting to perform the same analysis in more realistic growing network models looking for more similarities with real-world networks. For example, the exponent of the connectivity distribution can be tuned to small values by incorporating the first level of selection of the present model in the generalized BA model [6], which is known to give arbitrary γ values in the interval (2,3).

The authors thank F. Falo, J. L. García-Palacios, L. M. Floría, and A. F. Pacheco for helpful comments and discussions. J. G.-G. acknowledges financial support of the CSIC through Grant No. I3P-BPD2002-1 Y. M. was supported by the Secretaría de Estado de Educación y Universidades (Spain, SB2000-0357). This work was partially supported by the Spanish DGICYT project BFM2002-01798.

[1] M.E.J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
[2] H. Jeong, S.P. Mason, A.-L. Barabási, and Z.N. Oltvai, Nature (London) **411**, 41 (2001).
[3] R.V. Solé and J.M. Montoya, Proc. R. Soc. London, Ser. B **268**, 2039 (2001).
[4] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).
[5] G. Caldarelli, R. Marchetti, and L. Pietronero, Europhys. Lett. **52**, 386 (2000).
[6] S.N. Dorogovtsev and J.F.F. Mendes, Adv. Phys. **51**, 1079 (2002).
[7] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[8] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks. From Biological Nets to the Internet and the WWW* (Oxford University Press, Oxford, UK, 2003).
[9] *Handbook of Graphs and Networks*, edited by S. Bornholdt and H. G. Schuster (Wiley-VCH, Germany, 2003).
[10] A. Vázquez and Y. Moreno, Phys. Rev. E **67**, 015101(R) (2003).
[11] Y. Moreno, J.B. Gómez, and A.F. Pacheco, Phys. Rev. E **68**, 035103(R) (2003).
[12] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[13] A.-L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).
[14] S. Mossa, M. Barthelemy, H.E. Stanley, and L.A.N. Amaral, Phys. Rev. Lett. **88**, 138701 (2002).
[15] P.L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
[16] G. Caldarelli, A. Capocci, P. De Los Rios, and M.A. Muñoz, Phys. Rev. Lett. **89**, 258702 (2002).
[17] G. Bianconi and A.-L. Barabási, Europhys. Lett. **54**, 436 (2001).
[18] In the case that the number of elements in the set *A* is smaller than *m* we just add a link to all nodes in *A* without applying the PA rule.
[19] S. Redner, Eur. Phys. J. B **4**, 131 (1998).
[20] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[21] G. Bianconi and A. Capocci, Phys. Rev. Lett. **90**, 078701 (2003).