Growing scale-free networks with small-world behavior

Konstantin Klemm* and Víctor M. Eguíluz†
Center for Chaos and Turbulence Studies,‡ Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark and Instituto Mediterráneo de Estudios Avanzados (IMEDEA) (CSIC-UIB), E07071 Palma de Mallorca, Spain
(Received 31 July 2001; published 8 May 2002)

In the context of growing networks, we introduce a simple dynamical model that unifies the generic features of real networks: scale-free distribution of degree and the small-world effect. While the average shortest path length increases logarithmically as in random networks, the clustering coefficient assumes a large value independent of system size. We derive analytical expressions for the clustering coefficient in two limiting cases: random [$C \sim (\ln N)^2/N$] and highly clustered ($C = 5/6$) scale-free networks.

DOI: 10.1103/PhysRevE.65.057102 PACS number(s): 89.75.Hc, 87.23.Ge, 89.65.−s

Many systems can be represented by networks, i.e., as a set of nodes joined together by links indicating interaction. Social networks, the internet, food webs, distribution networks, metabolic and protein networks, the networks of airline routes, scientific collaboration networks, and citation networks are just some examples of such systems. [1–11]. Most of these networks share three prominent features. (a) The average shortest path length L is small. In order to connect two nodes on the graph, typically only a few edges need to be passed. (b) The clustering coefficient C is large. Two nodes having a common neighbor are far more likely to be connected to each other than are two nodes picked at random. (c) The distribution of the degree is scale-free, i.e., it decays as a power law. The absence of a typical scale for the connectivity of nodes is often related to the organization of the network as a hierarchy.

In this Brief Report we present an attempt to explain the empirical observations by a model of network self-organization according to simple rules. To our best knowledge, all previous approaches at modeling complex networks have only partially taken into account the above properties (a), (b), and (c). Co-occurrence of high clustering and short distance between nodes was originally termed the ''small-world'' phenomenon. It can be obtained by departing from a regular lattice, randomly rewiring links with a probability $p \ll 1$ [4]. However, networks created in this way display a degree distribution sharply peaked around the mean value; a power-law decay is not observed. Barabási and Albert (BA) have given an explanation of the scale-free distribution by reformulating Simon's model [12,13] in the context of growing networks. New nodes join the network by attaching m links to other nodes, chosen according to linear preferential attachment. This means that a node obtains one of the new links with a probability proportional to the number of links it already has. The algorithm, henceforth called the BA model, generates networks with a degree distribution $P(k) = 2m^2k^{-3}$ with $k \geqslant m$. However, as the system size N grows, the clustering coefficient approaches zero as the network size increases. The value of the clustering coefficient predicted by the BA model is typically several orders of magnitude lower than found empirically.

Recently, an alternative algorithm has been suggested [14] to account for the high clustering found in scale-free networks. Here we generalize the model allowing for random connections as the network grows. We will show first that a small ratio of random connections is sufficient to obtain small path length, keeping the high clustering and scale-free degree distribution of the original model. Second we will show, numerically and theoretically, that the crossover can be explained by the different scaling of the path length and the clustering in the limiting cases.

The model. Each node of the network is assigned a binary state variable. A newly generated node is in the *active* state and keeps attaching links until eventually deactivated. Taking a completely connected network of m active nodes as an initial condition, each step of the time-discrete dynamics consists of the following three stages. (i) A new node joins the network by attaching a link to each of the m active nodes. For each of the m links of the new node it is decided randomly, with probability μ, whether the link connects to the active node (as in the original model) or connects to a random node. (ii) The new node becomes active. (iii) One of the active nodes is deactivated. The probability that node i is chosen for deactivation is $p_i = ak_i^{-1}$ with normalization $a^{-1} = \Sigma_j k_j^{-1}$. The random nodes are chosen according to linear preferential attachment, i.e., the probability that node j obtains a link is proportional to the node's degree k_j.

For $\mu = 0$ we recover the high clustering model; the case $\mu = 1$ is the BA model. Both cases generate networks with degree distribution $P(k) = 2m^2k^{-3}$ ($k \geqslant m$) and average connectivity $\langle k \rangle = 2m$. Varying μ in the interval [0,1] allows us to study the crossover between the two models. We are especially interested in the behavior of the topological properties, namely, the average shortest path length and the clustering coefficient, as a function of the crossover parameter μ. Figure 1 shows the variation of the average shortest path length and the clustering coefficient with the parameter μ. When increasing μ from zero to small finite values, the average shortest path length L drops rapidly and approaches the low value of the BA model. The clustering coefficient C remains essentially constant in this same range $0 < \mu \ll 1$. We

*Email address: klemm@nbi.dk
†Email address: martinez@nbi.dk
‡URL: http://www.nbi.dk/CATS

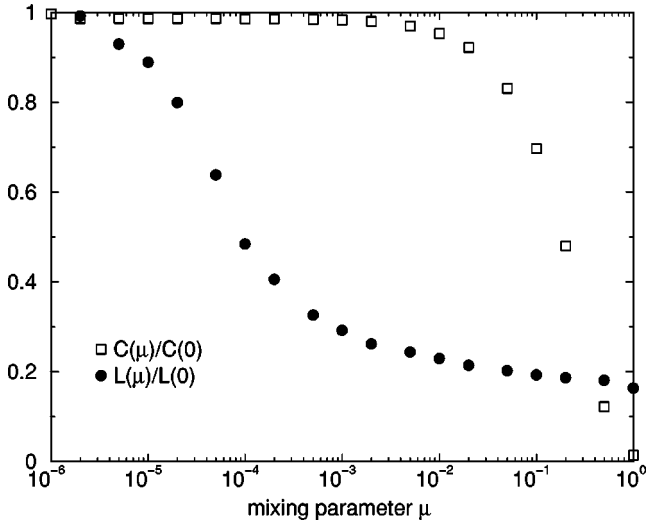

FIG. 1. Small-world effect in scale-free networks. Introducing the ratio $\mu \ll 1$ of random links into the highly clustered scale-free networks drastically reduces the typical distance L between nodes. However, the strongly interconnected neighborhoods of the original model ($\mu = 0$) are preserved, as the clustering coefficient remains at its large value. Only when μ reaches the order of 1 does the clustering coefficient drop significantly. All plotted values are averages over 100 independent realizations. The networks have $N = 10^4$ nodes with average degree $\langle k \rangle = 20$.

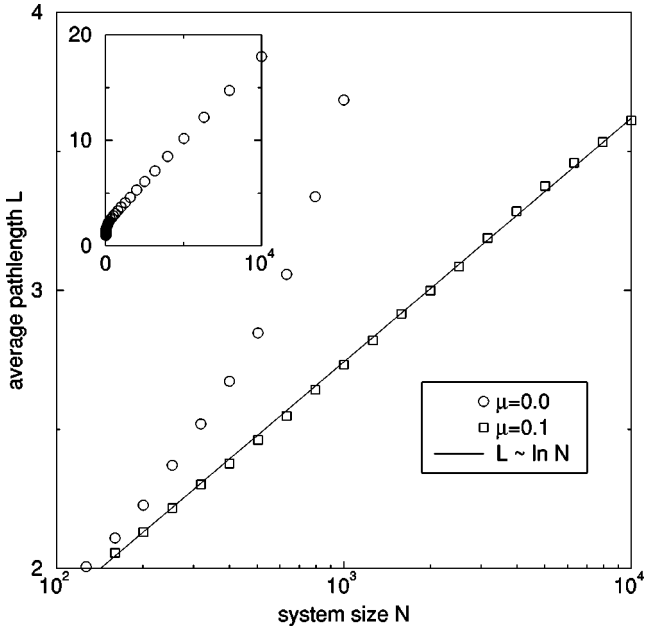

FIG. 2. Average shortest path length L as a function of system size N. In networks without long-range connections ($\mu = 0$) the relation between L and N is linear. This is seen best in the inset with linear scales on both axes. When attaching a fraction $\mu = 0.1$ of all links to random nodes instead of the currently active ones, L grows merely logarithmically with N. The values can be fitted well by a straight line in the plot with logarithmic N scale (main panel). All values plotted are averages over 100 independent realizations. The average degree is $\langle k \rangle = 20$.

have checked that the power-law distribution of the degree (not shown here) is still obtained in this range. Thus the model with $0 < \mu \ll 1$ reproduces the three generic properties (a), (b), and (c) of real-world networks. The model is robust against changes in the rule for the introduction of random links. The small-world transition shown in Fig. 1 does not change significantly when the attachment is not preferential, i.e., every node receives a random link with the same probability.

The observed drop in the average shortest path length L and clustering C is due to a qualitative change in the dependence of L and C on the system size. Thus, the change becomes more pronounced the larger is the system.

Shortest path length. In Fig. 2 we show L as a function of the system size N for $\mu = 0$ and $\mu = 0.1$. For $\mu = 0$, the average shortest path length grows linearly, $L \propto N$, the same behavior observed in one-dimensional regular lattices. This indicates that although there are highly connected nodes in the network there are no effective long-range connections that are able to reduce the path length. In clear contrast, a logarithmic growth of L is obtained for $\mu = 0.1$, $L \propto \ln N$. The logarithmic increase of L with system size is typical of the small-world effect [17].

Clustering. We derive analytical expressions for the clustering C in the two limiting cases $\mu = 0$ (the high clustering model) and $\mu = 1$ (the BA model).

Consider first the case $\mu = 0$. At any given time step the set of active nodes is completely interconnected, simply because a newly generated node always connects to all active nodes before being activated itself. It follows that a node l with degree $k_l = m$ has $C_l = 1$ because all the $m(m-1)/2$ possible links between neighbors of l actually exist. If l is deactivated in the time step of its generation its neighborhood does not change any more and it keeps $C_l = 1$. Otherwise a node $i \neq l$ is deactivated. In the next time step the node $l+1$ connects to l and all its neighbors apart from node i. Then $k_l(k_l-1)-1$ of the possible $k_l(k_l-1)$ links between neighbors of l exist, where now $k_l = m+1$. If node l keeps being active a node $j \neq l$ is deactivated. Node $l+2$ connects to all neighbors of l apart from i and j causing another two links to be missing in the neighborhood of l. See Fig. 3 for an illustration. By induction it follows that after n iterations $\sum_{\nu=1}^{n} \nu = n(n+1)/2$ links are missing in the neighborhood of l.

Thus the clustering C_l depends only on the degree k_l. The exact relation is

$$C(k) = 1 - \frac{(k-m+1)(k-m)}{k(k-1)}. \qquad (1)$$

This expression indicates that the local clustering scales as $C(k) \sim k^{-1}$ for large k. It is interesting to notice that a similar scaling has been found in the World Wide Web [15].

The clustering coefficient C can be obtained as the mean value of $C(k)$ with respect to the degree distribution $P(k) = 2m^2 k^{-3}$, $k \geqslant m$. The result is

$$C = \int_m^\infty \left(1 - \frac{(k-m+1)(k-m)}{k(k-1)}\right) 2m^2 k^{-3} dk \qquad (2)$$

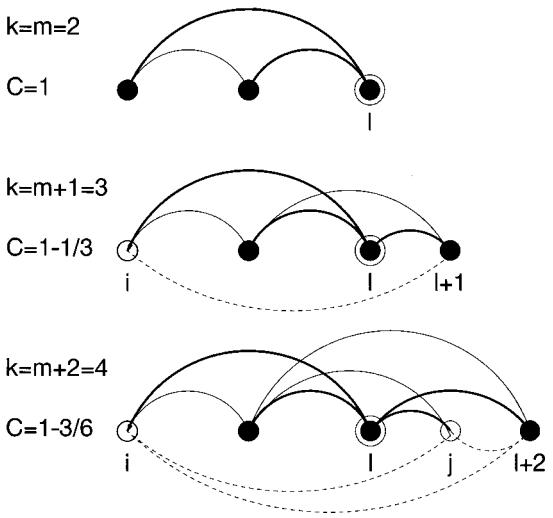

FIG. 3. Illustrating the calculation of the clustering coefficient of the highly clustered model ($\mu=0$, $m=2$). The encircled node is the node l under consideration. Links of this node are drawn as thick lines, links between its neighbors are thin lines. The dotted lines are links that are ''missing'' in the neighborhood of l. Active nodes are filled circles, inactive nodes are unfilled. For further explanation, see text.

$$=\frac{5}{6}-\frac{7}{30m}+\mathcal{O}(m^{-2}). \tag{3}$$

In the limit of large m the clustering coefficient is 5/6. It is worth noting that this value is higher than for regular lattices. The value $5/6\approx 0.83$ is similar to the one obtained in the film actor network (0.79), the coauthorship network in neuroscience (0.76), and networks of word synonyms (0.7) [16].

Let us now consider the BA model ($\mu=1$). When adding node j to the network, the probability for one link of node j to connect with node i is the ratio of the degree of the node i, k_i, and the sum of all nodes' degrees in the network, $2mj$. Thus the probability for the existence of a link from j to i is given by

$$\mathrm{Prob}\{(ij)\}=m\frac{k_i(j)}{2mj}, \tag{4}$$

where the prefactor m takes into account that m links per node are added to the network. By $k_i(j)$ we denote the degree of node i at the time that node j is added. Neglecting small fluctuations, the degree of the ith node is $k_i(j)=m(j/i)^{0.5}$ according to Ref. [11]. Inserting into Eq. (4) gives

$$\mathrm{Prob}\{(ij)\}=\frac{m}{2}(ij)^{-0.5}. \tag{5}$$

The local clustering $C_l(N)$ of the node l in a network of size N is defined as the number of links between neighbors of l, divided by the total number of pairs of neighbors l has. Only taking into account expectation values and treating the nodes as a continuum, we find

$$C_l(N)=\frac{\int_1^N di\int_1^N dj\,\mathrm{Prob}\{(li)\}\mathrm{Prob}\{(lj)\}\mathrm{Prob}\{(ij)\}}{k_l^2(N)}, \tag{6}$$

where we have approximated the total number of neighbors by $k_l^2/2$. Evaluating the probabilities according to Eq. (4) and using $k_l^2(N)=m^2N/l$ yields

$$C_l(N)=\frac{m^3}{8k_l^2(N)}\int_1^N di\int_1^N dj(li)^{-0.5}(lj)^{-0.5}(ij)^{-0.5} \tag{7}$$

$$=\frac{m^3}{8lk_l^2(N)}(\ln N)^2 \tag{8}$$

$$=\frac{m}{8}\frac{(\ln N)^2}{N}. \tag{9}$$

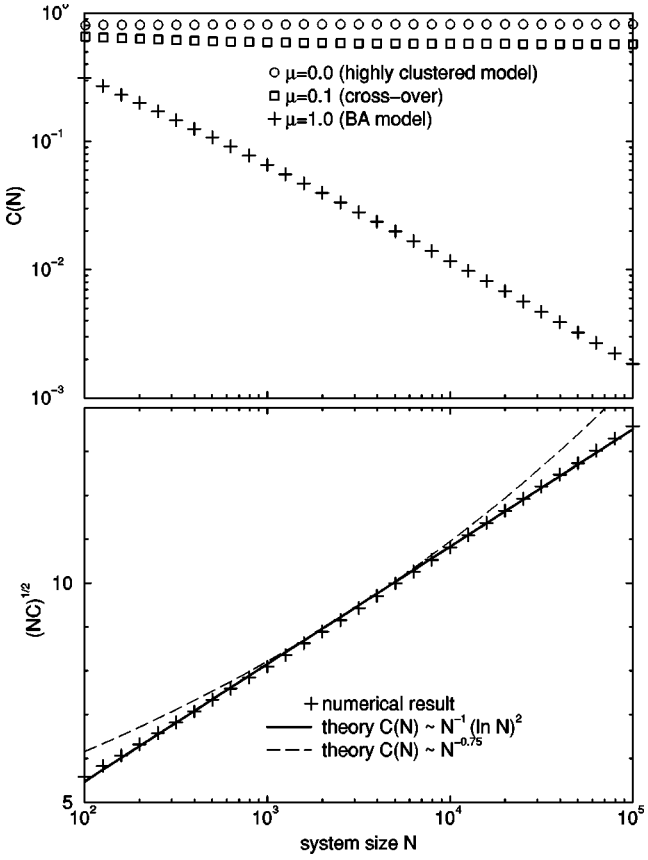

FIG. 4. Upper panel: The clustering coefficient C as a function of network size. Networks generated with $\mu=0.0$ quickly reach the large value predicted by the analytical calculations ($C\approx 0.83$). With 10% long-range connections ($\mu=0.1$) the clustering is lower but still approaches an asymptotic value clearly above zero. In the BA-model ($\mu=1.0$) the clustering coefficient decreases drastically with growing system size. Each of the three data sets is an average over 100 independent simulation runs. Lower panel: For the BA model, the function $[C(N)N]^{0.5}$ grows as $\ln N$, giving a straight line on a logarithmic-linear plot. This indicates very good agreement with the analytical result $C(N)\propto N^{-1}(\ln N)^2$. For comparison, the theoretical curve $C(N)\propto N^{-0.75}$ is shown, as suggested in Ref. [16].

The average value of the local clustering C_l does not depend on the node l under consideration. The networks generated by the BA model show homogeneous clustering, despite the inhomogeneous scale-free connectivity. With increasing network size N, the clustering coefficient decreases as N^{-1} in leading order. The difference with respect to a random graph, having a Poisson distribution of degree, is seen only in the logarithmic correction $(\ln N)^2$.

Figure 4, upper panel, shows the clustering coefficient obtained from numerical simulations. For $\mu=0$ we find an asymptotic value of approximately 0.83 as predicted analytically. Also for $\mu=0.1$ convergence to a finite value is observed. The BA model ($\mu=1.0$) displays a rapid decay of C as the network size N grows. The behavior of $C(N)$ in the BA model is analyzed in the lower panel of Fig. 4, clearly supporting the expression in Eq. (9). $C(N)$ is found to be inversely proportional to the system size, with logarithmic corrections. A pure power law with exponent -0.75 as proposed in Ref. [16] describes the numerical data less accurately.

Conclusions. In summary, we have defined and analyzed a simple model of self-organizing networks with high clustering, small path length, and a scale-free distribution of degree. The networks with these generic properties are obtained as a crossover between highly clustered scale-free networks [14] and scale-free random graphs [11]. Although we have considered growing networks, the dependence of the topology on the crossover parameter is very similar to the small-world transition observed in static networks when rewiring links into a regular grid [4]. Therefore our studies make a connection between small-world graphs and scale-free networks, essentially unifying both concepts in one model. We have also studied numerically the size dependence of the shortest path length and derived analytical expressions for the clustering.

[1] S.H. Strogatz, Nature (London) **410**, 268 (2001).
[2] L.A.N. Amaral, A. Scala, M. Barthélémy, and H.E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11 149 (2000).
[3] S. Wasserman and K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, England, 1994).
[4] D.J. Watts and S.H. Strogatz, Nature (London) **393**, 440 (1998).
[5] R. Albert, H. Jeong, and A.-L. Barábasi, Nature (London) **401**, 130 (1999).
[6] R.J. Williams and N.D. Martinez, Nature (London) **404**, 180 (2000).
[7] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabási, Nature (London) **407**, 651 (2000).
[8] H. Jeong, S.P. Mason, A.L. Barabási, and Z.N. Oltvai, Nature (London) **411**, 41 (2001).
[9] S. Redner, Euro. Phys. J. B **4**, 131 (1998).
[10] M.E.J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
[11] A.-L. Barabási and R. Albert, Science **286**, 509 (1999); A.-L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).
[12] H.A. Simon, Biometrika **42**, 425 (1955).
[13] S. Bornholdt and H. Ebel, Phys. Rev. E **64**, 035104 (2001).
[14] K. Klemm and V.M. Eguíluz, Phys. Rev. E **65**, 036123 (2002).
[15] J.-P. Eckmann and E. Moses, e-print cond-mat/0110338.
[16] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[17] M.E.J. Newman, J. Stat. Phys. **101**, 819 (2000).