Scale-free network model of node and connection diversity

Xiang Cheng, Hongli Wang, and Qi Ouyang*
Department of Physics and Mesoscopic Physics Laboratory, Peking University, Beijing 100871, China
(Received 26 December 2001; published 21 June 2002)

A network model with node and connection diversity is proposed in this paper. Distinctly from to other models whose nodes and connections are represented by identical simple points and lines, we investigate inhomogeneous networks with two kinds of sites and link by growth of preferential attachment. Scale-free networks with varied centralizations and exponents (ranging from 2.0 to theoretically infinity) are obtained, and the influences of the relative ratio of the two kinds of sites p, the number of links connected from each site m, and initial attractiveness δ are studied. A mean-field theory that agrees well with our numerical results was proposed and analyzed. The theory gives the analytical scaling exponent of the form $\gamma = 2 + p + \delta/m - \delta(1-p)/(mp+m+\delta)$.

DOI: 10.1103/PhysRevE.65.066115 PACS number(s): 05.10.−a, 05.40.−a, 87.18.Sn

Important properties of small-world and scaling behavior observed in real-life networks [1–7] have led to the recognition of networks as a prototype for studying complex systems. We realize that the view angle of the network is fundamental to our understanding of the complexity in the real world, which is essentially a huge set of complex networks [8]. Recently, attention along this line of research has been focused on hierarchically organized networks, whose connectivity distributions follow a power law. As distinct from homogeneous networks, such as those described by the classical random graph model of Erdos and Renyi [9], where no node is particularly important, in this type of network a small fraction of vertices with high numbers of connections play key roles in the function of the whole system. The presence of the scaling behavior is observed in a wide range of real-world networks. Examples include the Internet [2], the World Wide Web [3,5], an electronic power grid [4], the collaboration graph of movie actors [3], the citation pattern of scientific publications [1], and metabolic and protein interaction networks [6]. In order to describe the empirical results, considerable interest in designing scale-free network models with various specific features, as well as in their topological and dynamical properties is currently being aroused. A variety of models with different emphases on the growth process are proposed [10–13]. The question of stability to random or intentional attacks [14,15], topological properties such as fractal dimensions, spectral dimensions [16,17], and eigenvalue spectra [18] are investigated. Dynamical processes such as the spreading of a computer virus on the Internet has also been discussed recently [19].

Most scale-free network models considered so far ignore the diversity of nodes and connections. Network vertices are viewed as identical simple sites, and links as simple nondirectional edges. The intricate real-world networks are, however, typically composed of distinct nodes and connections. There can be many kinds of nodes, and the links between them can have directions and weights. For instance, in the biochemical network that controls cell division in mammals [20], enzymes and substrates as vertices are wildly diverse, and their interactions served as links are unilateral: a protein contributes a peptide that binds to a receptor site or pocket on another protein, but usually the latter does not affect the former the same way. In the example of the World Wide Web (WWW), HTML pages as nodes are obviously distinct: a page that has a link to another page does not assure that it is referenced back by the latter. This discrepancy sometimes leads to a qualitative difference between model systems and networks found in the real world. For example, from real data [21] we calculate that the centralization index [for definition see Eq. (2)] of WWW is about 0.02; while a network built using the rule of Barabasi and Albert [3] gives a value of 0.0021, ten times smaller than that found in the real world. In order to incorporate this ubiquitous diversity of node and connection in real networks, models whose constituents are nonidentical nodes and links are desirable.

In this paper, we propose a simple inhomogeneous network model that consists of two kinds of nodes and connections. The nodes are linked with unidirectional or bidirectional edges according to the specific types of two nodes that connect. We show that networks with scaling structure can be built using these nonidentical elements by the mechanism of preferential attachment. By adjusting the relative ratio of the two distinct sites and other parameters, we obtain scale-free networks with different centrality. The scaling exponent ranges from 2.0 to infinity and the centralization index varies in a broad range. A mean-field theory that agrees well with numerical results is introduced and analyzed.

I. THE MODEL

We consider two classes of nodes denoted by A and B, representing two types of interacting elements with certain distinct functions, and two kinds of links for their interactions. One kind of link is bidirectional edge that connects between type A nodes; the other one is unidirectional connection that points from a type B site to a type A site. Links between B sites are prohibited in our model. In other words, a type B site is only permitted to be linked by a type A site, while a type A site can be connected either by an A or B site.

*Corresponding author. Email address: qi@phy.pku.edu.cn

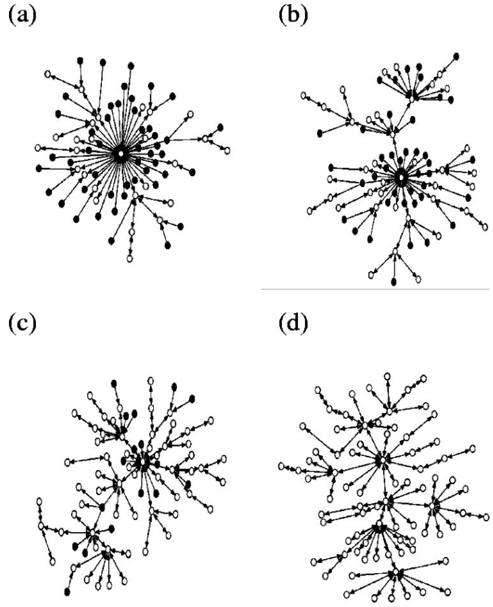

FIG. 1. Examples of evolving networks built with different relative ratios of A (open circle) and B (black dot) nodes, $p=0.3$, 0.5, 0.8, and 1.0 for (a), (b), (c), and (d), respectively. The total number of nodes is 80 with $m=1$, $\delta=0$, $m_0=1$.

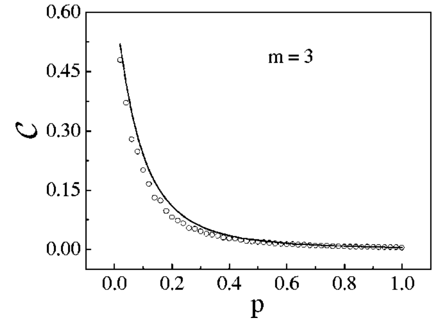

FIG. 2. The centralization as a function of p. Circles in the figure are numerical simulations; the solid line is theoretical predictions of Eq. (13). Parameters: $m=3$, $\delta=1$, $m_0=3$, with a total of 30 000 nodes.

We define the connectivity (or node degree) k_i of site i in the network as the total number of incoming links (connections pointing to it) from other sites, and simulate a growth of a network using the following rules: start with a small number of sites (m_0), among them there is at least one A-type site. At each time step, a new A or B site is created with probability p or $1-p$, respectively. It is added to the existing network by making m connections with old sites. When two type A sites are linked by a bidirectional edge, their connectivities are both increased by 1; if instead an A site is connected to a B site by a unidirectional edge, the former increases its connectivity by 1 while the degree of the latter remains unchanged. The connection of a new member to the existing network follows the rule of preferential attachment,

$$W(k_i)=\frac{k_i}{\sum_j k_j}, \tag{1}$$

where k_i is the connectivity of an old site i in the network. $W(k_i)$ is the probability that site i is to be connected by the new-coming site. It is determined by its node-degree contribution to the summation of connectivity in the system. In order to endow each B node with the chance to be connected by A sites, either A or B sites are assigned with an initial connectivity δ; therefore both types have a nonzero initial linking probability. Notice that a connection between two B sites is prohibited, so that the connectivity for a B site never increases. When the network grows, nodes with high values of connectivity establish preferential relations with the newly added site.

Figure 1 illustrates four examples of growing networks with different values of p, which takes respectively 0.3, 0.5, 0.8, and 1.0 for (a), (b), (c), and (d). All networks are hierarchically organized with a small fraction of highly connected sites. However, there is a qualitative difference among them: the network in Fig. 1(a) is highly centralized, it is very close to a star network where most sites are connected with a hub of the type A site; while the network in Fig. 1(d) is less centralized. We observe that network with a higher value of p bears a lower centrality and vice versa. As a characteristic quantity for hierarchical networks, the centralization $\mathcal{C}$ is used to measure the centrality quantitatively [22]. It is defined in our model as

$$\mathcal{C}=\frac{\sum_i (k_{\max}-k_i)}{m(G-1)(G-p-1)}, \tag{2}$$

where $k_{\max}$ is the highest connectivity of all sites in the network, and G is the total number of sites. The centrality is measured by the summation of the connectivity discrepancy of each site from $k_{\max}$. The denominator in Eq. (2) is a normalization factor, which is the maximal centralization that a network can be built with the given constraint (i.e., with the same G and p). In the simulation, we calculated the dependence of the centralization on the value of p. The result is shown in Fig. 2 (circles). One observes that the centralization decreases fast with the increase of p.

The scale-free feature is reserved in our networks. The probability $P(k)$ that a site in the network has k incoming links follows a power law. Figure 3 shows the connectivity (or node degree) distribution for a network of 300 000 nodes with $p=0.5$. The double-logarithmic plot shows a bounded scale-free behavior with a connectivity cutoff. This cutoff comes out from the finite network size. For a graph of a much larger size, the linear regime will extend. Generally, for networks with a large enough amount of sites, the scale-free regime is well defined and the slope does not depend on the network size. Figure 3 has been obtained by calculating the average from the data of 20 runs of simulations. The result of a sample can differ a little from the average due to fluctuations (refer to the inset of Fig. 3 for a sample). The fluctuation is generally not significant when the network size is large enough. The exponent γ for $P(k)\propto k^{-\gamma}$ in the scale-free regime as a function of p is shown in Fig. 4 (circles). γ

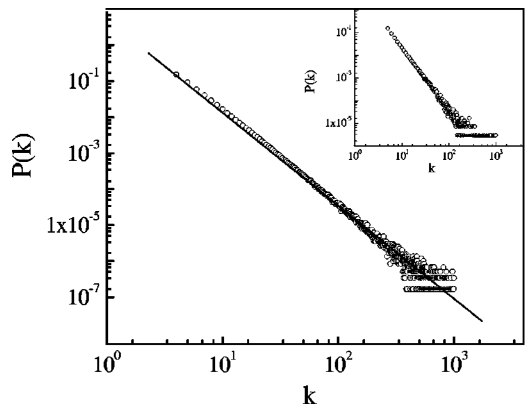

FIG. 3. Log-log plot of $P(k) \sim k$ for networks of 300 000 nodes with $p=0.5$. It was calculated from the average of 20 simulation runs. The solid line has a slope of 2.64. Other parameters: $m=m_0=3$, $\delta=1.0$.

decreases with the increase of p. Notice that index γ is not exactly linear as a function of p but with a little deviation. This property will be clear in the following theoretical analysis.

II. THE MEAN-FIELD THEORY

We next give a mean-field analysis for the networks described above. It allows us to determine the scaling exponents and the centralizations analytically. Since the growth probability of node degree for a site i is proportional to its connectivity contribution to the whole network, one readily writes the continuous growth dynamics of node degree k_i for site i as

$$\frac{dk_i(t)}{dt}=m\frac{k_i(t)}{\sum_{j=1}^{m_0+t}k_j(t)}. \tag{3}$$

The denominator in the above equation, which is the total connectivity of the system at time t, can be assumed to have the form

$$\sum_{j=1}^{m_0+t}k_j(t)=\delta(m_0+t)+pmt+\lambda mt. \tag{4}$$

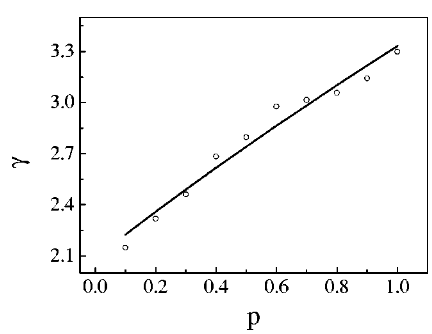

FIG. 4. Scaling exponent γ as a function of p from numerical simulations (circles) and theoretical prediction (solid line) of Eq. (11). Parameter values are as those for Fig. 2.

The first term on the right-hand side is the sum of all initial connectivities at time t. It is determined by the specific value of δ, which is a small number and represents the initial attractiveness of a new site added to the system, and the total number of sites (m_0+t) in the network. As vertices are added at equal time intervals to the system, a fraction (p) of them is A type sites. The second term comes from this class of sites. When an A site enters the system, its node degree increases by m no matter what kind of node it bumps into. The third term in Eq. (4) is due to the situation when a newly coming site happens to be connected with an A site in the system: whatever type of site is the newly added node, it increases the degree of the old A site it connected to by one. λ in Eq. (4) is the probability that a connection from the new coming site is linked with a type A site. It depends on t, and can be approximately written as

$$\lambda(t)=\frac{\delta(m_0+t)p+mpt+mt}{\delta(m_0+t)+mpt+mt}, \tag{5}$$

where the numerator is the total connectivity of all A sites; the denominator is roughly the sum of node degree for all sites in the system. λ grows quickly to a saturated platform as the time proceeds. The limit value of $\lambda(t)$ when time goes to infinity is

$$\lambda_\infty=1-\frac{(1-p)\delta}{mp+m+\delta}. \tag{6}$$

Taking into account Eq. (4) and Eq. (6), and with the initial condition that site i was added into the system at time t_i with connectivity $k_i(t_i)=m+\delta$, Eq. (3) can be solved as

$$k_i(t)=(m+\delta)\left[\frac{\Theta t+m_0\delta}{\Theta t_i+m_0\delta}\right]^{m/\Theta}, \tag{7}$$

where $\Theta \equiv \delta+mp+m\lambda_\infty$. From the above result, we can calculate the connectivity distribution. The probability that a vertex i has a connectivity $k_i(t)$ smaller than k can be written as

$$P(k_i(t)<k)=P\left(t_i>\frac{[\Theta t+m_0\delta]\left(\dfrac{m+\delta}{k}\right)^{\Theta/m}-m_0\delta}{\Theta}\right). \tag{8}$$

As the nodes are added to the system at equal time intervals, the probability density of t_i is $1/(m_0+t)$. $P(k_i(t)<k)$ then has the form

$$P(k_i(t)<k)=\left(1-\frac{[\Theta t+m_0\delta]\left(\dfrac{m+\delta}{k}\right)^{\Theta/m}-m_0\delta}{\Theta(m_0+t)}\right). \tag{9}$$

The connectivity distribution $P(k)$ is finally

$$P(k)=\frac{[\Theta t+m_0\delta](m+\delta)^{\Theta/m}}{m(m_0+t)}k^{-(1+\Theta/m)}. \tag{10}$$

From the above scaling law, the exponent γ is

$$\gamma = 2 + p + \frac{\delta}{m} - \frac{(1-p)\delta}{mp+m+\delta}. \quad (11)$$

The solid curve in Fig. 4 is the theoretical result calculated from Eq. (11). The numerical simulation (circles) in the figure fluctuates around the theoretical prediction. The agreement is not perfect because of limited total sites we simulated. A simulation of sufficient large network will eliminate the discrepancy. From Eq. (11), one sees that the exponent does not linearly depend on p. Instead, it is determined by p, m, and δ together in a complex manner. Its value is always larger than 2.0, and has an infinite upper bound due to the δ terms.

The centralization $\mathcal{C}$ can be calculated from Eq. (7). The vertex that has the largest connectivity in the system is the one created at $t_i = 0$. The connectivity $k_{\max}$ is therefore

$$k_{\max} = (m+\delta)\left[\frac{\Theta t + m_0\delta}{m_0\delta}\right]^{m/\Theta}. \quad (12)$$

The analytical centralization according to definition Eq. (2) has the form

$$\mathcal{C} = \frac{k_{\max}(m_0+t) - [\delta(m_0+t) + pmt + \lambda_\infty mt]}{m(t+m_0-1)(t+m_0-p-1)}. \quad (13)$$

The theoretical prediction of Eq. (13) agrees well with the numerical results. The solid line in Fig. 2 depicts the dependence of the centralization on p. It is in consistency with the numerical simulations.

III. DISCUSSION

In the limit when p approaches 1, A-type nodes dominate the whole system. The network is still scale free, but loses its diversity in node and connection. At this limit, the exponent $3 < \gamma < \infty$. At the other end limit of $p = 0$, all sites are of B type, which deny any connection with each other. This leads to a complete disassembly of the system. The initial attractiveness δ plays a nontrivial role in the growth of the network by introducing the effect of m on γ. It can also drive the system out of scale-free structure and push it into a random network as δ goes to infinity. At zero initial attractiveness $(\delta = 0), \gamma = 2 + p$, which is independent of m. In this case, when $p = 1$, the network comes back to the original scale-free model of Barabasi and Albert (BA) with $\gamma = 3$ [3].

The model considered here brings out networks with scaling exponents and centralizations varying in a wide range, in which real random networks happen to fall. As reported previously [3], the actor collaboration network, WWW, paper citation pattern, and electrical power grid have a γ of value 2.3, 2.1, 3.0, and 4.0, respectively. We estimated the centralization value of WWW from real data containing 325 729 nodes [21]. It is about 0.02 with $m \approx 2.73$. Using this number and supposing δ to be about 1, we get $p = 0.3$ and the scaling index $\gamma = 2.5$. Compared with the BA model which gives the much smaller value of 0.0021 with the same m and $\gamma = 3$, our model is closer to the real situation.

From the results presented here, we see that the relative ratio of distinct nodes and connections has important effects on the behavior of the system. By adjusting the ratio, the scaling exponents vary accordingly. The network presented here is the simplest model that incorporates node and connection diversity. The ubiquitous diversity in node and connection of real-world networks is surely much more complicated than the case considered here. The elements can wildly vary in their properties, and the interactions among them are complex [23]. Profound topological properties and dynamical behaviors are expected in these situations so that more realistic network models that take into account various empirical aspects of real networks are desirable in order to understand the intricate behaviors of complex systems.

ACKNOWLEDGMENTS

The work was supported by the Chinese Natural Science Foundation and Special Funds for Major State Basic Research Projects; X.C. thanks the Taizhao Foundation for support and fruitful discussions with Y. Hu and F. Li.

[1] S. Redner, Eur. Phys. J. B **4**, 131 (1998).
[2] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).
[3] A.L. Barabasi and R. Albert, Science **286**, 509 (1999).
[4] D.J. Watts and S.H. Strogatz, Nature (London) **393**, 440 (1998).
[5] B.A. Huberman and L.A. Adamic, Nature (London) **401**, 131 (1999).
[6] H. Jeong, B. Tombor, Z.N. Oltvai, and A.-L. Barabasi, Nature (London) **407**, 651 (2000); H. Jeong, S.P. Mason, A.-L. Barabasi, and Z.N. Oltvai, *ibid.* **411**, 41 (2001).
[7] F. Liljeros *et al.*, Nature (London) **411**, 907 (2001).
[8] For a recent review, please refer to R.A.-L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002).
[9] P. Erdos and A. Renyi, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960); B. Bollabas, *Random Graphs* (Academic, London, 1985).
[10] S.N. Dorogovtsev and J.F.F. Mendes, Phys. Rev. E **62**, 1842 (2000); **63**, 025101(R) (2001); **63**, 056125 (2001); S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000); Phys. Rev. E **63**, 062101 (2001).
[11] P.L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[12] S.H. Yook, H. Jeong, A.L. Barabasi, and Y. Tu, Phys. Rev. Lett. **86**, 5835 (2001).
[13] O.V. Kirillova, Phys. Rev. Lett. **87**, 068701 (2001).
[14] R. Albert, H. Jeong, and A. Barabasi, Nature (London) **406**, 378 (2000).
[15] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **86**, 3682 (2001).

[16] S. Bilke and C. Peterson, Phys. Rev. E **64**, 036106 (2001).
[17] Z. Burda, J.D. Correia, and A. Krzywicki, Phys. Rev. E **64**, 046118 (2001).
[18] I.J. Farkas, I. Derenyi, A.L. Barabasi, and T. Vicsek, Phys. Rev. E **64**, 026704 (2001).
[19] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001); Phys. Rev. E **63**, 066117 (2001).
[20] K.W. Kohn, Mol. Biol. Cell **10**, 2703 (1999).
[21] We used the data from http://www.nd.edu/~networks.
[22] S. Wasserman and K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, 1994).
[23] S.H. Strogatz, Nature (London) **410**, 268 (2001).