Scale-free networks with tunable degree-distribution exponents

H. Y. Lee, H. Y. Chan, and P. M. Hui
Department of Physics, The Chinese University of Hong Kong, New Territories, Shatin, Hong Kong, China
(Received 10 January 2004; published 2 June 2004)

We propose and study a model of scale-free growing networks that gives a degree distribution dominated by a power-law behavior with a model-dependent, hence tunable, exponent. The model represents a hybrid of the growing networks based on popularity-driven and fitness-driven preferential attachments. As the network grows, a newly added node establishes m new links to existing nodes with a probability p based on popularity of the existing nodes and a probability $1-p$ based on fitness of the existing nodes. An explicit form of the degree distribution $P(p,k)$ is derived within a mean field approach. For reasonably large k, $P(p,k) \sim k^{-\gamma(p)}\mathcal{F}(k,p)$, where the function $\mathcal{F}$ is dominated by the behavior of $1/\ln(k/m)$ for small values of p and becomes k independent as $p \rightarrow 1$, and $\gamma(p)$ is a model-dependent exponent. The degree distribution and the exponent $\gamma(p)$ are found to be in good agreement with results obtained by extensive numerical simulations.

DOI: 10.1103/PhysRevE.69.067102 PACS number(s): 89.75.Hc, 05.65.+b, 02.50.Cw

Many complex systems, including social, biological, physical, economic, and computer systems, can be studied by network models in which the nodes represent the constituents and links or edges represent the interactions between constituents [1–5]. Interesting findings in the statistics of real-world networks reveal that classical random networks [6,7] do not often represent the geometrical or topological structure of real-world networks [8–18] correctly. Most noticeable of the properties observed are so-called "six degree of separation" [19,20] from one node to any arbitrary node, and the highly clustering feature. In particular, many networks show a power-law degree distribution of the form $P(k) \sim k^{-\gamma}$, with the exponent γ taking on values between 2 to 3. This behavior has led to the construction of models of scale-free growing networks. Barabási and Albert (BA) [21] proposed a model in which a new node is added in each turn and m new links are established with existing nodes with the probability of establishing a link being proportional to the number of existing links of the nodes. This preferential attachment is thus driven entirely by the popularity of existing nodes. Detailed numerical simulations and analytic analysis showed that $\gamma=3$ for the BA model. The idea of incorporating preferential attachment in a growing network has led to proposals of a considerable number of models of scale-free networks [22–25]. Alternatively, models with preferential attachment driven entirely by fitness have also been proposed [26,27]. In these models, each node carries a randomly assigned fitness value that gives a collective character of the node other than its popularity, and the probability of establishing a new link to an existing node is proportional to the product of fitness and the number of existing links. It was found that the degree distribution $P(k) \sim k^{-\gamma}/\ln(k/m)$, with $\gamma=2.255$ for the fitness-driven model [26].

Many real-world networks such as paper citations in scientific journals, the World-Wide Web, the Internet, and the collaborative networks of actors and actresses, exhibit a degree distribution with a network-dependent exponent that takes on values close to but below 3. It is therefore interesting to construct and analyze models with a tunable degree-distribution exponent. In the present work, we propose and study a model representing a hybrid of the growing network models based on popularity-driven and fitness-driven preferential attachments of new links. As the network grows, a newly added node has a probability p of being popularity driven and a probability $1-p$ of being fitness driven in establishing new links. Thus the resulting network consists of a mixture of two types of nodes, with a fraction p establishing new links based entirely on popularity consideration. The model reflects the fact that not every one, taking the nodes as agents in a population, prefers to follow the popular persons, but instead may prefer to establish relationships with others based on characters other than the popularity of the agents. Our model thus incorporates the inhomogeneous nature of many real-world networks in which not all the nodes are identical. We report results of extensive numerical simulations on the degree distribution for networks of size 10^7 nodes, and compare numerical results with an analytic expression derived via a mean field approach. For reasonably large k, the degree distribution $P(p,k)$ follows the form $k^{-\gamma(p)}\mathcal{F}(k,p)$, where $\mathcal{F}(k,p)$ is dominated by the behavior of $1/\ln(k/m)$ for small p and becomes k independent for $p \rightarrow 1$. The exponent $\gamma(p)$ can be extracted numerically and results are found to be in good agreement with that of the mean field theory.

Our model is defined as follows. Initially a fully connected network of m_0 nodes is constructed, with m_0 typically of order unity. In this work, we use $m_0=5$. The network grows with one new node being added to the existing network at a time. Each newly added node establishes a number of m new links to existing nodes. With probability p, the new node establishes links by preferential attachments based on popularity of the existing nodes [21], i.e., the probability that an existing node i is connected is proportional to the degree or the number of links $k_i(t)$ that node i carries. With probability $1-p$, the new node establishes links by preferential attachments based on fitness of the existing nodes [26], i.e., the probability that an existing node i is connected is proportional to the product $k_i(t)\eta_i$, where the fitness η of a node is a randomly assigned value in the interval $0<\eta<1$ associated with the node when it is introduced into the network.

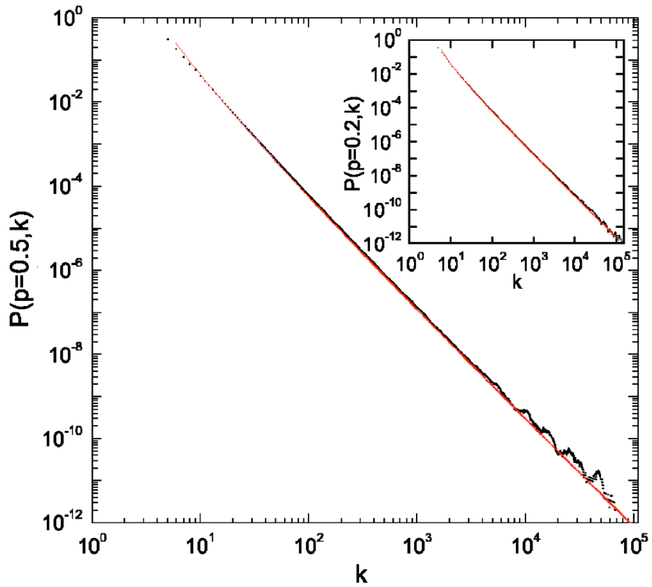

FIG. 1. The degree-distribution function $P(p,k)$ as a function of k on a log-log scale for $p=0.5$ and $p=0.2$ (inset). The symbols give numerical results averaging over 10 different realizations of networks of size of 10^7 nodes. The lines give the analytic results within the mean field theory.

For $p=1$ ($p=0$), the present model reduces to the popularity-driven BA [21] (fitness-driven [26]) model.

Detailed numerical simulations have been carried out for our model with networks of size up to 10^7 nodes. Each newly added node establishes $m=5$ new links. Figure 1 shows a typical degree-distribution function on a log-log scale for the case of $p=0.5$, i.e., a new node randomly chooses to follow the popularity-driven or fitness-driven rules in establishing new links, and for $p=0.2$ (see inset). The large size of the networks used in this study makes the comparison with analytic results and the extraction of the exponent in the degree distribution easier. The data shown in Fig. 1 represent an average over 10 different realizations of networks of the same size. To explore the functional form of the degree distribution and to extract possible exponent, we need some guidance from analytic treatment.

The model can be treated analytically by a mean field approach [1,2,26]. For sufficiently long time, the connectivity $k_i(t)$ of the ith node with fitness η_i evolves according to the following continuous time evolution equation

$$\frac{\partial k_i}{\partial t} = mp\frac{k_i(t)}{\sum_j k_j} + m(1-p)\frac{\eta_i k_i(t)}{\sum_j \eta_j k_j}, \tag{1}$$

where the first and second terms describe the increment in k_i due to popularity and fitness, respectively. Note that $\Sigma_j k_j = 2mt$, with the factor of 2 coming from the undirected nature of the links. We assume that k_i takes on the form

$$k_i(t,t_0) = m\left(\frac{t}{t_0}\right)^{\beta(\eta_i,p)}, \tag{2}$$

where t_0 is the time at which the ith node was introduced into the network. Since k_i can at most be increased by one at each time step, it cannot grow faster than t, thus implying $0<\beta(\eta_i,p)<1$.

The fitness η_i is, in general, chosen from a distribution $\rho(\eta)$. The average of the sum $\Sigma_j\ \eta_j k_j$ over $\rho(\eta)$ can be evaluated by

$$\left\langle \sum_j \eta_j k_j \right\rangle = \int \eta_j \rho(\eta_j) \int_1^t k_j(t,t_0)dt_0 d\eta_j$$
$$= \int \eta_j\rho(\eta_j)\frac{m(t-t^{\beta(\eta_j,p)})}{1-\beta(\eta_j,p)}d\eta_j. \tag{3}$$

For large t, since $t^\beta/t \to 0$, the contribution from the term $t^{\beta(\eta_j,p)}$ becomes negligible and we have

$$\left\langle \sum_j \eta_j k_j \right\rangle = C(p)mt, \tag{4}$$

where

$$C(p) = \int \rho(\eta_j)\frac{\eta_j}{1-\beta(\eta_j,p)}d\eta_j. \tag{5}$$

Substituting Eqs. (2) and (4) into Eq. (1), we obtain

$$\beta(\eta_i,p) = \frac{p}{2} + \frac{1-p}{C(p)}\eta_i. \tag{6}$$

From now on, we drop the subscripts i and j for brevity. Equations (6) and (5) can be solved self-consistently for $C(p)$. For $\rho(\eta)$ being a uniform distribution between 0 and 1, i.e., $\rho(\eta)=1$ for $\eta \in [0,1]$, $C(p)$ satisfies

$$C(p) = \int_0^1 \frac{\eta}{\left(1-\frac{p}{2}\right) - \frac{(1-p)\eta}{C(p)}}d\eta, \tag{7}$$

for which the integral can be performed to obtain the self-consistent equation

$$\frac{1}{C(p)} = \frac{1-p/2}{1-p}(1-e^{-2(1-p)/C(p)}). \tag{8}$$

Equation (8) can be solved numerically for $C(p)$ for given value of p. It is found that $C(p) \in [1,1.255]$ for $p \in [0,1]$. Note that for $p=1$, $\beta=1/2$ [see Eq. (6)] independent of η as in the BA popularity-driven model [21]. For $p=0$, $C(0)=1.255$, and $\beta(\eta,0)=\eta/1.255$, as in the purely fitness-driven model [26].

To proceed, the cumulative probability distribution function (CDF) $\mathcal{P}_\eta(p,k_i>k)$ for a particular fitness η and given p can be found by noting that a degree higher than some value k for a node corresponds to a cutoff in time before which the node must have been introduced into the network, i.e.,

$$\mathcal{P}_\eta(p,k_i>k) = \mathcal{P}_\eta\left[t_0 < t\left(\frac{m}{k}\right)^{1/\beta(\eta,p)}\right] = \left(\frac{m}{k}\right)^{1/\beta(\eta,p)}. \tag{9}$$

Note that a prefactor of $t/(m_0+t)$, which approaches unity for sufficiently long time, has been ignored in Eq. (9). To obtain the CDF for the whole network $\mathcal{P}(p,k)$, an average is taken over a uniform distribution of η, thus

$$\mathcal{P}(p,k) = \int_0^1 \left(\frac{m}{k}\right)^{1/\beta(\eta,p)} d\eta, \qquad (10)$$

with $\beta(\eta,p)$ given by Eq. (6). The probability distribution function (PDF) of degrees in the network $P(p,k)$ is related to $\mathcal{P}(p,k)$ through

$$P(p,k) = -\frac{\partial}{\partial k}\mathcal{P}(p,k). \qquad (11)$$

Using Eq. (10) and by making the substitution $x = \ln(k/m)/[p/2+(1-p)\eta/C(p)]$, Eq. (11) gives

$$P(p,k) = \frac{C(p)}{k(1-p)}\int_{x_0}^{x_1}\frac{e^{-x}}{x}dx, \qquad (12)$$

where the lower limit of the integral $x_0=\alpha_0\ln(k/m)$ and the upper limit $x_1=\alpha_1\ln(k/m)$, with

$$\alpha_0 = \frac{1}{p/2+(1-p)/C(p)} \qquad (13)$$

and

$$\alpha_1 = \frac{2}{p}. \qquad (14)$$

We aim at getting the functional form of $P(p,k)$ and in particular the dependence on k at given p. The integral in Eq. (12) can be carried out by parts, and only the "surface" term survives in the limit of $\ln(k/m) \gg 1$. It follows from the fact that $1/x$ takes on the maximum value of x_0^{-1} for x in the interval $[x_0,x_1]$, and thus the integral $\int_{x_0}^{x_1}(e^{-x}/x^2)dx \leqslant x_0^{-1}\int_{x_0}^{x_1}(e^{-x}/x)dx$, with $x_0^{-1} \ll 1$ for large k. Equation (12) thus gives

$$P(p,k) = \frac{C}{m(1-p)}\cdot\frac{\left(\frac{k}{m}\right)^{-(1+\alpha_0)}}{\alpha_0\ln(k/m)}\left[1-\frac{\alpha_0}{\alpha_1}\left(\frac{k}{m}\right)^{-\alpha_1+\alpha_0}\right]. \qquad (15)$$

Equation (15) is the main result of the mean field treatment. It gives the explicit k dependence of the degree-distribution function. It is worth noting that Eq. (15) gives the correct results in both limits of $p\to 0$ and $p\to 1$. For purely fitness-driven model, $P(0,k)=(k/m)^{-2.255}/[m\ln(k/m)]$ [26]. For purely popularity-driven model, the $\ln(k/m)$ term in the denominator can be shown to be canceled by the terms in the parentheses, giving $P(1,k)=2m^2k^{-3}$ [21].

Results obtained from the mean field theory can be compared with numerical results. The solid lines in Fig. 1 show the degree distribution for $p=0.5$ and $p=0.2$ (inset) using Eq. (12). Excellent agreements are found between mean field and numerical results. Equation (15) also suggests a functional form for $P(p,k)$. For reasonably large k, e.g., $k\sim 10^2$ or above, the second term in the parentheses is small compared to unity, especially for small values of p. Hence, Eq. (15) suggests that $P(p,k)\sim k^{-\gamma(p)}/\ln(k/m)$, where

$$\gamma(p) = 1+\alpha_0 = 1+\frac{1}{[p/2+(1-p)/C(p)]}, \qquad (16)$$

with $C(p)$ given by Eq. (8) within mean field theory. Numerically, we fit the degree-distribution function to the form $k^{-\gamma(p)}/\ln(k/m)$ and extract the exponent $\gamma(p)$ directly from results of numerical simulations for each value of p in steps of 0.1. However, as $p\to 1$, one should be more careful in handling Eq. (15) in numerical extraction of the exponent as the prefactor $1/(1-p)$ diverges and the $\ln(k/m)$ term becomes unimportant due to cancellation effect from the terms in the parentheses. In this case, it is more convenient to work from the $p=1$ limit and extract a functional form from Eq. (15) that is valid for $p\to 1$. The result is $P(p,k)\sim[p/2+(1-p)/C](k/m)^{-\gamma(p)}/mp^2$, where $\gamma(p)$ is again given by Eq. (16). This form is used to extract the exponent $\gamma(p)$ for $p=0.9$ and 1.0. For $p\leqslant 0.8$, using either functional form extracts the same value of γ. Figure 2 shows the exponent $\gamma(p)$ numerically extracted from simulations, together with the analytic result. The two sets of results are found to be in good agreement.

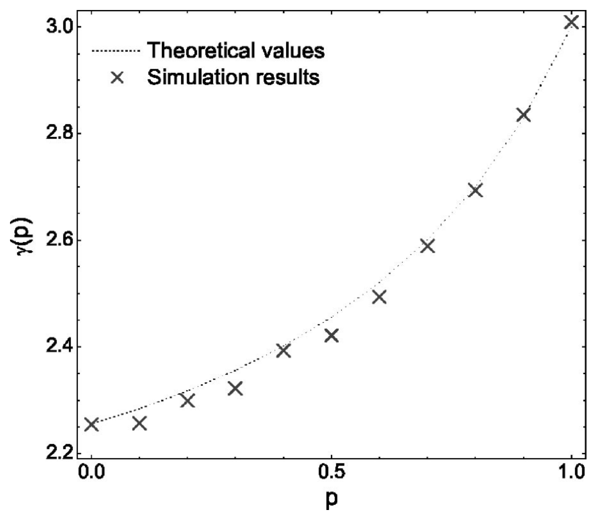

FIG. 2. The exponent $\gamma(p)$ characterizing the degree-distribution function as obtained by fitting to numerical simulation results (symbols) for networks of size 10^7 nodes for values of p from $p=0$ to $p=1$ in steps of 0.1, and by the mean field theory (dotted line) given by $\gamma(p)=1+1/[p/2+(1-p)/C(p)]$ with $C(p)$ given by Eq. (8).

In summary, we proposed and studied a model in which the nodes are inhomogeneous. The model combines popularity-driven and fitness-driven preferential attachments in growing networks. Extensive numerical simulations were carried out and a mean field theory was developed. The degree-distribution function shows a predominant power-law behavior. The exponent takes on a model-dependent, hence tunable, value depending on the concentration p of nodes for which the links are established by a popularity-driven mechanism. The exponent $\gamma(p)$ takes on values between 2.255 and 3, which lie within the range of values of γ observed in many real-world networks. Analytic expressions for the degree-distribution function and the exponent $\gamma(p)$ were derived. Results of mean field theory were found to be in good agreement with results obtained by numerical simulations.

We wish to thank Oliver K.H. Chung for useful discussions on numerical methods on setting up large networks efficiently, and Professor DaFang Zheng for useful discussions on fitness-driven growing networks.

[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] D. J. Watts, *Small Worlds* (Princeton University Press, Princeton, 1999).
[5] D. J. Watts, *Six Degrees: The Science of a Connected Age* (William Heinemann, London, 2003).
[6] P. Erdös and P. Rényi, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).
[7] B. Bollobás, *Random Graphs* (Academic Press, London, 1985).
[8] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[9] B. A. Huberman and L. A. Adamic, Nature (London) **401**, 131 (1999).
[10] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajalopagan, R. Stata, A. Tomkins, and J. Wiener, Comput. Netw. **33**, 309 (2000).
[11] S. Redner, Eur. Phys. J. B **4**, 131 (1998).
[12] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, Nature (London) **407**, 651 (2000).
[13] H. Jeong, S. Mason, A.-L. Barabási, and Z. N. Oltvai, Nature (London) **411**, 651 (2001).
[14] S. Maslov and K. Sneppen, Science **296**, 910 (2002).
[15] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Åberg, Nature (London) **411**, 907 (2001).
[16] M. Granovetter, Am. J. Sociol. **78**, 1360 (1973).
[17] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
[18] E. L. Berlow, Nature (London) **398**, 330 (1999).
[19] J. Travers and S. Milgram, Sociometry **32**, 425 (1969).
[20] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[21] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[22] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[23] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[24] R. Albert and A.-L. Barabási, Phys. Rev. Lett. **85**, 5234 (2000).
[25] P. L. Krapivsky, G. J. Rodgers, and S. Redner, Phys. Rev. Lett. **86**, 5401 (2001).
[26] G. Bianconi and A.-L. Barabási, Europhys. Lett. **54**, 436 (2001).
[27] G. Ergun and G. J. Rodgers, Physica A **303**, 261 (2002).