Growing networks with mixed attachment mechanisms

2006 J. Phys. A: Math. Gen. 39 2035

(http://iopscience.iop.org/0305-4470/39/9/004)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **39** (2006) 2035–2042 doi:10.1088/0305-4470/39/9/004

Growing networks with mixed attachment mechanisms

Zhi-Gang Shao, Xian-Wu Zou, Zhi-Jie Tan and Zhun-Zhi Jin

Department of Physics, Wuhan University, Wuhan 430072, People's Republic of China

E-mail: xwzou@whu.edu.cn

Received 15 November 2005, in final form 26 December 2005
Published 15 February 2006
Online at stacks.iop.org/JPhysA/39/2035

Abstract

Networks grow and evolve when new nodes and links are added in. There are two methods to add the links: uniform attachment and preferential attachment. We take account of the addition of links with mixed attachment between uniform attachment and preferential attachment in proportion. By using numerical simulations and analysis based on a continuum theory, we obtain that the degree distribution $P(k)$ has an extended power-law form $P(k) \sim (k + k_0)^{-\gamma}$. When the number of edges k of a node is much larger than a certain value k_0, the degree distribution reduces to the power-law form $P(k) \sim k^{-\gamma}$; and when k is much smaller than k_0, the degree distribution degenerates into the exponential form $P(k) \sim \exp\left(-\frac{\gamma k}{k_0}\right)$. It has been found that degree distribution possesses this extended power-law form for many real networks, such as the movie actor network, the citation network of scientific papers and diverse protein interaction networks.

PACS numbers: 89.75.Hc, 89.75.Fb, 87.23.Ge, 05.10.−a

1. Introduction

The complexity of numerous social, biological and communication systems is rooted in the rather interwoven web defined by the system's components and their interactions. For example, the cell is best described as a complex network of chemicals connected by chemical reactions [1]. Similarly, society is characterized by a huge social network whose nodes are individuals or organizations, connected by social interactions [2]. The World Wide Web is an enormous virtual network of Web pages connected by hyperlinks [3]. To model the interwoven web, networks have recently been an object of intensive investigation [4–8]. To understand properties of networks, many mechanisms have been investigated [9–19].

Many quantities are proposed to characterize the macroscopic properties of networks [4, 5], among which the degree distribution $P(k)$, i.e. the probability with which a node in the network possesses k edges, is of particular importance. In the random graph model of Erdös and Rényi (ER) [6] and the small-world model of Walts and Strogatz (WS) [7, 8], links of the networks are connected with uniform attachment and the degree distribution $P(k)$ decays exponentially. In contrast, in the scale-free model of Barabási and Albert (BA) [9], links of the networks are connected with preferential attachment and the degree distribution $P(k)$ decays as a power law, following $P(k) \sim k^{-\gamma}$. The uniform attachment and preferential attachment are two kinds of fundamental mechanisms of network growth. For networks growing with uniform attachment mechanism $P(k)$ decays exponentially; and for networks growing with preferential attachment mechanism $P(k)$ decays as a power law [9, 11]. As a matter of fact, in realistic systems the degree distribution is neither exponential nor power law. For example, in the citation network of scientific papers and authors [20–22], for a relatively large value of k, $P(k)$ decays as a power law, whereas for a relatively small value of k, $P(k)$ declines exponentially. Therefore, to obtain a better understanding of the structure and evolution of complex networks, the growing mechanism of networks could not be described by pure uniform attachment or pure preferential attachment. Recently, we noticed that in [18] the authors gave a complex network model with both preferential and uniform attachments with the weight p. They obtained the probability that a new link pointing to a given node i was expressed as

$$\Pi_i = \frac{(1-p)k_i + p}{\sum_j \left[(1-p)k_j + p\right]},$$

i.e.

$$\Pi_i \sim k_i + \frac{p}{1-p}.$$

This result is similar to the probability ($\Pi_i \sim A + k_i$) of the initial attractiveness model described in [19]. The scaling exponent of degree distribution γ obtained in [18] depends on an average degree m.

We extend the model of scale-free networks proposed by Barabási and Albert in this paper. The growth mechanism of the network is mixed attachment, which is randomly chosen between uniform and preferential attachments in proportion. This model gives the degree distribution with an extended power-law form $P(k) \sim (k + k_0)^{-\gamma}$, γ is independent of the average degree m. When k is much larger than k_0, it is reduced to the power-law form $P(k) \sim k^{-\gamma}$; and when k is much smaller than k_0, it degenerates into the exponential form $P(k) \sim \exp\left(-\frac{\gamma k}{k_0}\right)$. There are a lot of real networks whose degree distributions can be satisfactorily fitted with this extended power-law form, such as the movie actor network [7, 10], the citation network of scientific papers [20–22], and diverse protein interaction networks [23–25]. Therefore, this mixed attachment exists in the real network evolution.

2. Model and method

Our model is an extension of the scale-free model proposed by Barabási and Albert [9]. In our model, the growing network is constructed with mixed attachment. In each step the attachment is chosen randomly between preferential and uniform attachments in proportion. The probability of preferential attachment is q, and the probability of uniform attachment is $1 - q$. Growth and evolution of the networks in our model are described as follows.

At the beginning, the network consists of a small number m_0 nodes. Then, at every time step, a new node is added in. The new node is connected to m nodes ($m \leqslant m_0$) among the nodes existing in the network. After t time steps, this procedure results in a network with $N = m_0 + t$ nodes and mt edges.

At the time step t, when a new node is added in, the connected probability of a node i connected to the new node is

$$\Pi_i = q\Pi_i^{\text{pref}} + (1-q)\Pi_i^{\text{unif}}, \tag{1}$$

where q is the fraction of preferential attachment, Π_i^{pref} and Π_i^{unif} are the connected probability of a node i in the cases of preferential and uniform attachment respectively.

According to uniform attachment mechanism, the new node is connected to any node existed in the network with equal opportunity. Thus, at the time step t the connected probability of a node i by the uniform attachment mechanism can be expressed as

$$\Pi_i^{\text{unif}} = \frac{1}{m_0 + t - 1}. \tag{2}$$

In the light of preferential attachment mechanism, the connected probability Π_i^{pref} is proportional to the fraction of k_i (the degree of the node i) at that time. Therefore, we have

$$\Pi_i^{\text{pref}} = \frac{k_i}{\sum_j k_j}, \tag{3}$$

where j extends all nodes over the network except the new one added in. Thus the sum in the denominator is $\sum_j k_j = 2mt - m$ [4].

It can be seen from equation (1) that if $q = 0$, the present model reduces to the network growth model with uniform attachment and the degree distribution $P(k)$ decays exponentially as described in 11; if $q = 1$, the present model degenerates into the BA model in which the network grows by way of preferential attachment and $P(k)$ decays as a power law as shown in [9, 11].

3. Formulation of the degree distribution

The degree distribution $P(k)$ is the probability with which a node in the network possesses k edges. It is an important quantity to characterize structure property and evolution of networks. We have deduced the degree distribution $P(k)$ for present model by using the continuum theory. In the interval from $t-1$ to t, the degree of a node i increases by $m\Pi_i$. Adopting continuity approximation [9–11], the rate of change of k_i can be written as

$$\frac{\partial k_i}{\partial t} = m\Pi_i = mq\frac{k_i}{\sum_j k_j} + m(1-q)\frac{1}{m_0 + t - 1}. \tag{4}$$

Taking into account that $\sum_j k_j = 2mt - m$ and $t \gg m_0$, equation (4) changes into

$$\frac{\partial k_i}{\partial t} = q\frac{k_i}{2t} + m(1-q)\frac{1}{t}. \tag{5}$$

The initial condition is that at time t_i the node i is added in and it has the degree $k_i(t_i) = m$. Thus, the solution of equation (5) has the form

$$k_i = \left(\left[\frac{mq}{2} + m(1-q)\right]\left(\frac{t}{t_i}\right)^{\frac{q}{2}} - m(1-q)\right)\frac{2}{q}. \tag{6}$$

By using equation (6) the inequality $k_i < k$ changes into

$$t_i > \left[\frac{\frac{mq}{2} + m(1-q)}{\frac{kq}{2} + m(1-q)}\right]^{\frac{2}{q}} t. \tag{7}$$

Therefore, the probability that the node i has the degree $k_i(t)$ smaller than k, can be written as

$$P(k_i(t) < k) = P\left(t_i > \left[\frac{\frac{mq}{2} + m(1-q)}{\frac{kq}{2} + m(1-q)}\right]^{\frac{2}{q}} t\right) = 1 - P\left(t_i \leqslant \left[\frac{\frac{mq}{2} + m(1-q)}{\frac{kq}{2} + m(1-q)}\right]^{\frac{2}{q}} t\right). \tag{8}$$

Assuming that we add the nodes to the system at equal time intervals, the probability density of t_i is [9–11]

$$P_i(t_i) = \frac{1}{m_0 + t} \approx \frac{1}{t}. \tag{9}$$

Thus,

$$P\left(t_i \leqslant \left[\frac{\frac{mq}{2} + m(1-q)}{\frac{kq}{2} + m(1-q)}\right]^{\frac{2}{q}} t\right) = P_i(t_i)\left[\frac{\frac{mq}{2} + m(1-q)}{\frac{kq}{2} + m(1-q)}\right]^{\frac{2}{q}} t = \left[\frac{\frac{mq}{2} + m(1-q)}{\frac{kq}{2} + m(1-q)}\right]^{\frac{2}{q}}. \tag{10}$$

Substituting equation (10) into equation (8) we obtain

$$P(k_i(t) < k) = 1 - \left[\frac{\frac{mq}{2} + m(1-q)}{\frac{kq}{2} + m(1-q)}\right]^{\frac{2}{q}}. \tag{11}$$

The degree distribution $P(k)$ can be obtained from equation (11)

$$P(k) = \frac{\partial P\,(k_i(t) < k)}{\partial k} = \frac{2}{q}\left[\frac{m(2-q)}{q}\right]^{\frac{2}{q}}\left[k + \frac{2m(1-q)}{q}\right]^{-(1+\frac{2}{q})}. \tag{12}$$

Equation (12) possesses the extended power-law form as

$$P(k) \sim (k + k_0)^{-\gamma}, \tag{13}$$

where $k_0 = \frac{2m(1-q)}{q}$ and $\gamma = 1 + \frac{2}{q}$.

When k is much larger than k_0, equation (13) reduces to the power-law form $P(k) \sim k^{-\gamma}$. Conversely, when k is much smaller than k_0, we have

$$\ln(P(k)) \sim -\gamma \ln(k + k_0) = -\gamma\left[\ln\left(1 + \frac{k}{k_0}\right) + \ln(k_0)\right] \sim -\gamma\left[\frac{k}{k_0} + \ln(k_0)\right], \tag{14}$$

then,

$$P(k) \sim \frac{1}{k_0^{\gamma}} \exp\left(-\frac{\gamma k}{k_0}\right), \tag{15}$$

so, equation (13) degenerates into the exponential form $P(k) \sim \exp\left(-\frac{\gamma k}{k_0}\right)$.

4. Compared with numerical simulations

We have performed extensive numerical simulations to investigate behaviours of the degree distribution $P(k)$ in growing networks with mixed attachment between uniform attachment and preferential attachment. To reduce the effect of fluctuation on simulation results, for every system with a certain size the simulation results are average over ten network realizations.

In the present model there are two important parameters: the preferential attachment fraction q and the number of nodes connected by a new node m. First, we investigate the dependence of the degree distribution $P(k)$ upon q. Figure 1(a) shows the simulated degree

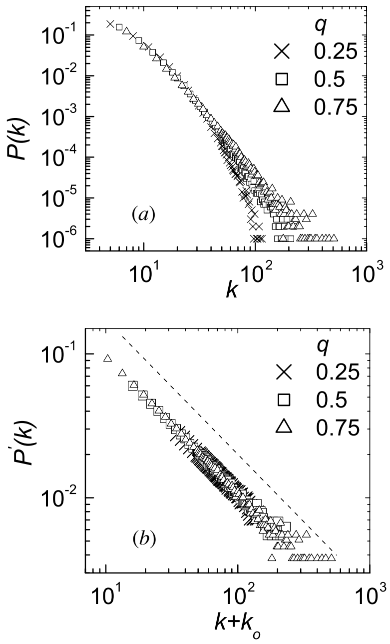

Figure 1. The simulated degree distribution $P(k)$ (a) and rescaled degree distribution $P'(k) = (P(k)/\frac{2}{q}[\frac{m(2-q)}{q}]^{\frac{2}{q}})^{\frac{q}{q+2}}$ (b) for a set of preferential attachment fraction q. The number of nodes connected by a new node $m = m_0 = 5$ and the size of system $N = 10^5$. $k_0 = \frac{2m(1-q)}{q}$ and the slope of the dashed line is -1.

distribution $P(k)$ for a set of preferential attachment fraction q=0.25, 0.5 and 0.75 in the case of the system size $N = 10^5$ and $m = m_0 = 5$. To make a comparison between simulated degree distributions and theoretical values, equation (12) is rewritten to the rescaled degree distribution

$$P'(k) = \left(\frac{P(k)}{\frac{2}{q}\left[\frac{m(2-q)}{q}\right]^{\frac{2}{q}}} \right)^{\frac{q}{q+2}} = (k + k_0)^{-1}. \tag{16}$$

Figure 1(b) shows simulated results of the rescaled degree distribution. In log–log plot of $P'(k)$, these simulated data fall into a straight line with the slope of -1. It agrees with the theoretical expression equation (16).

Second, we study variation of the degree distribution $P(k)$ with the number of nodes connected by a new node m. Figure 2(a) shows the simulated degree distribution $P(k)$ for a set of the number of nodes connected by a new node $m = m_0 = 3$, 5, and 7 in the case of the system size $N = 10^5$ and the preferential attachment fraction $q = 0.5$. In this case equation (12) becomes

$$P(k)/m^4 \propto (k + k_0)^{-\gamma}.(k_0 = 2m, \gamma = 5). \tag{17}$$

Figure 2(b) plots the simulated values of $P(k)/m^4$ as a function of $(k + k_0)$. It can be found by transforming $P(k)$ into $P(k)/m^4$, the curves with different m in figure 2(a) into a straight line with the slope of -5. It is just the thing described by the theoretical expression equation (17).

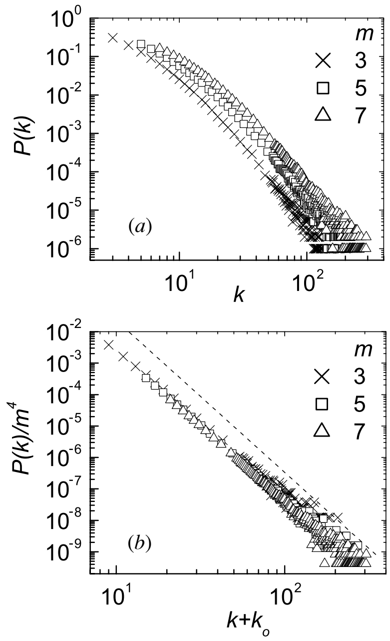

Figure 2. The simulated degree distribution $P(k)$ (a) and rescaled degree distribution $P(k)/m^4$ (b) for a set of number of nodes connected by a new node. The size of system $N = 10^5$ and the preferential attachment fraction $q = 0.5$. $k_0 = 2m$ and the slope of the dashed line is -5.

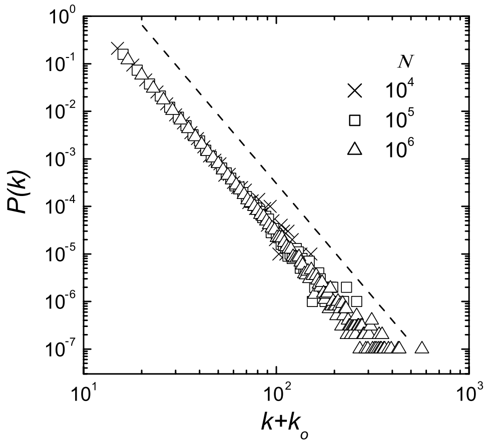

Figure 3. The simulated degree distribution $P(k)$ as a function of $(k + k_0)$ for a set of system size N, where $k_0 = 2m$. The number of nodes connected by a new node $m = m_0 = 5$ and the preferential attachment fraction $q = 0.5$. The slope of the dashed line shows -5.

The present model generates the growing networks. To check the variation of the degree distribution $P(k)$ with the system size N, a set of simulations with different system size have been done. Figure 3 shows the simulated degree distribution $P(k)$ with the system size $N = 10^4$, 10^5 and 10^6, $m = m_0 = 5$, and $q = 0.5$. These simulated data comply with the power law $P(k) \sim (k + k_0)^{-\gamma}$ with $k_0 = 2m = 10$ and $\gamma = 5$, that is just as the expected results of equation (12). It follows that the characteristics of these networks are independent of the system size N. It indicates that despite of continuous growth, the system organizes itself into a stationary state.

In former works, the degree distributions of some real networks have been found to follow form $P(k) \sim (k + k_0)^{-\gamma}$, such as movie actor network [7, 10], citation network of scientific paper [20–22] and diverse protein interaction networks [23–25]. Therefore, the degree distribution with the extended power-law form $P(k) \sim (k + k_0)^{-\gamma}$ is general in real world. It is conceivable that this kind of degree distribution form should be caused by the existence of mixed attachment between uniform attachment and preferential attachment in the real network evolution.

Finally, we stress the differences between our model and the model analysed in [18]. First, the fundamental attachment in growth mechanisms is different between these two models. In [18], authors think about the network grows by preferential and uniform attachments with a weight, but we consider the attachment to be randomly chosen between preferential and uniform ways in proportion. Thus, in [18], the probability that a new link points to a given node i is

$$\Pi_i = \frac{(1-p)k_i + p}{\sum_j[(1-p)k_j + p]};$$

but in our model, $\Pi_i = q\Pi_i^{\text{pref}} + (1-q)\Pi_i^{\text{unif}} = q\frac{k_i}{\sum_j k_j} + (1-q)\frac{1}{m_0+t-1}$. Second, for these two models the degree distribution has an extended form $P(k) \sim (k+k_0)^{-\gamma}$, but the character of the scaling exponent γ is dissimilar. In [18], γ depends on the average degree m, and in our model γ is independent on m. It seems to be acceptable that γ is independent on m. In addition, our model is easy to get the asymptotic form: when k is much larger than k_0, the degree distribution reduces to the power-law form $P(k) \sim k^{-\gamma}$; when if k is much smaller than k_0, the degree distribution degenerates into the exponential form $P(k) \sim \exp\left(-\frac{\gamma k}{k_0}\right)$.

5. Conclusions

We extend the model of scale-free networks proposed by Barabási and Albert. The character of the present model is that the attachment is mixed attachment, which is chosen randomly between preferential and uniform attachments in proportion. The probability with preferential attachment is q, and the probability with uniform attachment is $1-q$.

By using numerical simulations and analysis based on a continuum theory, we obtain that the degree distribution $P(k)$ has an extended power-law form $P(k) \sim (k+k_0)^{-\gamma}$, where $k_0 = \frac{2m(1-q)}{q}$ and $\gamma = 1 + \frac{2}{q}$. Then, we deduce the asymptotic form of the degree distribution: when $k \gg k_0$, the extended scale-free form reduces to the power-law form $P(k) \sim k^{-\gamma}$; and when $k \ll k_0$, it degenerates into the exponential form $P(k) \sim \exp\left(-\frac{\gamma k}{k_0}\right)$. This model can be used to describe the evolution of a kind of real networks, such as movie actor network, citation network of scientific paper and diverse protein interaction networks.

Acknowledgments

This work was supported by the National Natural Science Foundation of China No. 10374072 (XWZ) and No. 10274056 (ZJT). It was also supported by the Specialized Research Fund for the Doctoral Program of Higher Education No. 20020486009.

References

[1] Jeong H, Tombor B, Albert R, Ottvai Z N and Barabási A L 2000 *Nature* **407** 651

[2] Wasserman S and Faust K 1994 *Social Network Analysis* (Cambridge: Cambridge University Press)
[3] Albert R, Jeong H and Barabási A L 1999 *Science* **284** 92
[4] Albert R and Barabási A L 2002 *Rev. Mod. Phys.* **74** 47
[5] Dorogovtsev S N and Mendes J F F 2002 *Adv. Phys.* **51** 1079
[6] Bollobás B 1985 *Random Graphs* (London: Academic Press)
[7] Watts D J and Strogatz S H 1998 *Nature* **393** 440
[8] Watts D J 1999 *Small Worlds: The Dynamics of Networks Between Order and Randomness* (New Jersey: Princeton University Press)
[9] Barabási A L and Albert R 1999 *Science* **286** 509
[10] Barabási A L and Albert R 2000 *Phys. Rev. Lett.* **85** 5234
[11] Barabási A L, Albert R and Jeong H 1999 *Physica* A **272** 173
[12] Callaway D S, Hopcroft J E, Keinberg J M, Newman M E J and Strogatz S H 2001 *Phys. Rev.* E **64** 041902
[13] Deng Ke and Tang Yi 2004 *Chin. Phys. Lett.* **21** 1858
[14] Amaral L A N, Scala A, Barthélé M and Stanley H E 2000 *Proc. Natl. Acad. Sci. USA* **97** 11149
[15] Dorogovtsev S N and Mendes J F F 2000 *Phys. Rev.* E **62** 1842
[16] Dorogovtsev S N and Mendes J F F 2001 *Phys. Rev.* E **63** 056125
[17] Krapirsky P L, Redner S and Leyvraz F 2000 *Phys. Rev. Lett.* **85** 4629
[18] Liu Z, Lai Y C, Ye N and Dasgupta P 2002 *Phys. Lett.* A **303** 337
[19] Dorogovtsev S N, Mendes J F F and Samukhin A N 2000 *Phys. Rev. Lett.* **85** 4633
[20] Redner S 1998 *Eur. Phys. J.* B **4** 131
[21] Tsallis C and de Albuquerque M P 2000 *Eur. Phys. J.* B **12** 777
[22] Laherrere J and Sornette D 1998 *Eur. Phys. J.* B **2** 525
[23] Salwinski L, Miller C S, Smith A J, Pettit F K, Bowie J U and Eisenberg D 2004 *Nucleic Acids Res.* **32** 449
[24] Li S *et al* 2004 *Science* **303** 540
[25] Giot L *et al* 2003 *Science* **302** 1727