

MODELLING COLLABORATION NETWORKS BASED ON NONLINEAR PREFERENTIAL ATTACHMENT

TAO ZHOU*, BING-HONG WANG and YING-DI JIN

*Nonlinear Science Center and Department of Modern Physics
University of Science and Technology of China
Hefei Anhui, 230026, P. R. China
zhutou@ustc.edu

DA-REN HE, PEI-PEI ZHANG, YUE HE and BEI-BEI SU

*College of Physical Science and Technology
Yangzhou University, Yangzhou Jiangsu, 225002, P. R. China*

KAN CHEN

*Department of Computational Science, Faculty of Science
National University of Singapore, Singapore 117543, Singapore*

ZHONG-ZHI ZHANG and JIAN-GUO LIU

*Institute of Systems Engineering
Dalian University of Technology, Dalian Liaoning
116024 P. R. China*

Received 5 June 2006

Accepted 13 June 2006

In this paper, we propose an alternative model for collaboration networks based on nonlinear preferential attachment. Depending on a single free parameter “preferential exponent”, this model interpolates between networks with a scale-free and an exponential degree distribution. The degree distribution in the present networks can be roughly classified into four patterns, all of which are observed in empirical data. And this model exhibits small-world effect, which means the corresponding networks are of very short average distance and highly large clustering coefficient. More interesting, we find a peak distribution of act-size from empirical data which has not been emphasized before. Our model can produce the peak act-size distribution naturally that agrees with the empirical data well.

Keywords: Complex networks; collaboration network model; nonlinear preferential attachment.

PACS Nos.: 89.75.Hc, 64.60.Ak, 84.35.+i, 05.40.-a, 05.50+q, 87.18.Sn.

1. Introduction

The last few years have witnessed a tremendous activity devoted to the characterization and understanding of complex networks,^{1–4} which arise in a vast number of natural and artificial systems, such as the Internet,^{5–7} the World Wide Web,^{8,9} social networks of acquaintance or other relations between individuals,^{10–12} metabolic networks,^{13–15} food webs^{16–19} and many others.^{20–26} Owing to the computerization of the data acquisition process and the availability of high computing powers, scientists have found that the networks in various fields have some common characteristics, which inspires them to construct a general model. Recently, some pioneer works have been done that bring us new eyes of the networks' evolution mechanism. For instance, Barabási and Albert introduced a scale-free network model (BA network),²⁷ which suggests that two main ingredients of self-organization of a network in a scale-free structure are growth and preferential attachment.

So far, BA model may be the most successful model to fit the empirical results of complex systems, but there are still a great number of real networks whose evolution mechanisms cannot be explained by the BA model. In fact, we should not ask for an all-powerful model which can explain the reason of a freewill real network coming into being, since many different networks have distinct underlying growth mechanisms. Therefore, it is meaningful to construct a microscopic suitable model aiming at a special kind of network.

A particular class of networks is the so-called collaboration networks, which is considered to be a kind of social network in the early studies. In the social sciences, a collaboration network is generally defined as a network of actors connected by common membership in groups of some sort, such as clubs, teams or organizations. Some empirical studies relevant to collaboration networks have been done, including scientific collaboration networks,^{28–32} board of directorships,^{33,34} movie actors collaboration networks,³⁵ social events attending networks for women,³⁶ and so on. It is worthwhile to point out that the extension of collaboration networks should not be restricted within social networks. One instance is the software collaboration networks,³⁷ and more examples not related to social networks will be shown in the following text.

Ramasco, Dorogovtsev and Pastor-Satorras have proposed a model for collaboration networks (RDP model for short).³⁸ In the RDP model, they found the power law behavior in degree distribution, the nontrivial clustering-degree correlation and nontrivial degree-degree correlation. By introducing the mutual selection mechanism,³⁹ Li *et al.* have established a model for weighted collaboration networks in which both the power law weight distribution and degree distribution are obtained.⁴⁰ Zhang *et al.* proposed a collaboration network model based on the competition of preferential and random attachment.⁴¹ Very recently, González, Lind and Herrmann proposed a network model consisting of mobile agents, which provides us the possible mechanism of social networks' evolution.^{42,43}

In this paper, we propose an alternative model for collaboration networks based on nonlinear preferential attachment. Depending on a single free parameter, the

preferential exponent, this model interpolates between networks with a scale-free and an exponential degree distribution. The degree distributions of the present networks can be roughly classified into four patterns, all of which are observed in empirical data. And this model exhibits small-world effect, which means that the corresponding networks are of very short average distance and highly large clustering coefficient. More interestingly, we find a peak distribution of act-size from empirical data which has not been emphasized before. Our model can produce the peak act-size distribution naturally that agrees with the empirical data well.

The present paper is organized as follows. In Sec. 2, we introduce our model. In Sec. 3, we show the small-world effect exhibited by this model. In Sec. 4, we display the simulation results of degree distribution, and demonstrate that the degree distribution is approximated to a stretched exponential distribution⁴⁷ with adjustable parameter c . In Sec. 5, we show the simulation results and some empirical data of act-size distribution. The comparison and qualitative discussion are also included. Finally, in Sec. 6, we draw the main conclusion of our work.

2. The Model

Our network starts with m_0 nodes which are fully connected. Then, at each time step, we add a new node into the network that will have a collaboration with some existing nodes. We assume that the probability that an existing node x is chosen to be an actor in the collaboration is proportional to k^α ($\alpha \geq 0$), where k is the degree of x and α is the so-called “preferential exponent” denoting the degree of preferential attachment. For $\alpha > 0$, we have preferential attachment. This mechanism, named nonlinear preferential attachment, is firstly proposed by Krapivsky *et al.* to explain the exponential cut-off in the tail of power-law degree distribution.^{44,45} All the existing nodes which are chosen to be collaborated will link to the new node, and if two chosen old nodes have never collaborated so far, there will be a new edge added connecting them.⁴⁶ This is the main difference between our model and Krapivsky’s, since in Krapivsky’s model, the old nodes will not connect to each other. It is obvious that this model can be stretched to a weighted one by using the times of collaborations between the corresponding two nodes as edge weight. Since the aim of this paper is to introduce the characteristics of non-weighted networks, the simulation and analysis relevant to weighted networks will not be included, as they will be published elsewhere.

It should be noted that the act-size is not fixed at each time step since whether a certain node is chosen will not affect other nodes. We suppose that a node with degree k will be chosen with the probability

$$\pi(k) = \frac{\lambda k^\alpha}{\sum_i k_i^\alpha}, \quad (1)$$

in which λ is a constant, and $\sum_i k_i^\alpha$ is the normalization factor. Using $\langle s \rangle$ to represent the average value of act-size such as the mean number of authors per paper, we will conclude that $\langle s \rangle \approx \lambda + 1$, since the number of nodes chosen each time

has the expecting value $\Sigma\pi(k) = \lambda$. Thus, λ is a parameter which can be used to control the average act-size of the whole network. Therefore, when we simulate an idiographic network of known average act-size, the parameter λ is not free.

3. Small World Effect

In a network, the distance between two nodes is defined as the number of edges along the shortest path connecting them. The average distance L of the network, then, is defined as the mean distance between two nodes, averaged over all pairs of nodes and often considered to be one of the most important parameters to measure the efficiency of communication networks. The clustering coefficient $C(x)$ of node x is the ratio between the number of edges among $A(x)$ and the total possible number, where $A(x)$ denotes the set of all the neighbors of x . The clustering coefficient C of the whole network is the average of $C(x)$ over all x . Empirical studies indicate that most real-life networks have much smaller average distance (as $L \sim \ln N$ where N is the number of nodes in the network) than the completely regular networks and a much greater clustering coefficient than those of the completely random networks. And these two properties, small average distance and large clustering coefficient, make up the so-called “small world effect”.

Inspired by the empirical studies on real-life networks, Watts and Strogatz proposed a one-parameter model (WS model) that interpolates between an ordered finite-dimensional lattice and a random graph by randomly rewiring each edge of the regular lattice with probability p .³⁵ In the WS model, L scales logarithmatically with N , and the clustering coefficient is much larger than that of the random network, which is in excellent agreement with the characteristics of real networks. The pioneering article of Watts and Strogatz started an avalanche of research on the properties of small-world networks. In this section, we would like to demonstrate that the networks generated by the present rules display small-world effect.

At first, we study the average distance of the present model using the approach similar to that in Refs. 48 and 49. Using symbol $d(i, j)$ to represent the distance between nodes i and j , the average distance of present networks with order N , denoted by $L(N)$, is defined as:

$$L(N) = \frac{2\sigma(N)}{N(N-1)}, \quad (2)$$

where the total distance is:

$$\sigma(N) = \sum_{1 \leq i < j \leq N} d(i, j). \quad (3)$$

The distance between two existing nodes will not increase with the increasing of network size N , thus we have:

$$\sigma(N+1) \leq \sigma(N) + \sum_{i=1}^N d(i, N+1). \quad (4)$$

Assume that h existing nodes, x_1, x_2, \dots, x_h , are selected to collaborate with the new node $N + 1$, then:

$$d(i, N + 1) = \min\{d(i, x_j) | j = 1, 2, \dots, h\} + 1. \tag{5}$$

In a rough version, the sum $\sum_{i=1}^N \min\{d(i, x_j)\}$ can be expressed approximately in terms of $L(N - h + 1)$:

$$\sum_{i=1}^N \min\{d(i, x_j)\} \approx (N - h)L(N - h + 1). \tag{6}$$

In order to avoid the network being unconnected, always set $\lambda > 1$ and compel $h \geq 1$, which leads to $(N - h)L(N - h + 1) \leq (N - 1)L(N)$.

Combining the results above, we have:

$$\sigma(N + 1) < \sigma(N) + N + \frac{2\sigma(N)}{N}. \tag{7}$$

Consider (7) as an equation, then the increasing tendency of $\sigma(N)$ is determined by the equation:

$$\frac{d\sigma(N)}{dN} = N + \frac{2\sigma(N)}{N}, \tag{8}$$

which leads to

$$\sigma(N) = N^2 \ln N + H, \tag{9}$$

where H is a constant. As $\sigma(N) \sim N^2 L(N)$, we have $L(N) \sim \ln N$. Since (7) is an inequality in fact, the precise increasing tendency of L is a little tardier than $\ln N$.

In Fig. 1, we report a typical simulation result on average distance of the present networks under parameters $\alpha = 1$ and $\lambda = 3$, which agrees with the analytical result well.

In succession, we discuss the clustering coefficient. As mentioned above, for an arbitrary node x , the clustering coefficient $C(x)$ is:

$$C(x) = \frac{2E(x)}{k(x)(k(x) - 1)}, \tag{10}$$

where $E(x)$ is the number of edges between any two nodes in the neighbor-set $A(x)$ of node x , and $k(x) = |A(x)|$ denotes the degree of node x . The clustering coefficient C of the whole network is defined as the average of $C(x)$ over all nodes.

Figure 2 exhibits the simulation results on clustering coefficient of the present networks versus network size. From Fig. 2, one can find that the present networks display a highly clustered property even for large N . Therefore, our model exhibits a completely different clustering structure from that of BA networks, in which the clustering coefficient is very small and decreases with the increasing of network size N , following approximately $C \sim (\ln^2 N/N)$.⁵¹

In addition, we plot the clustering coefficient as a function of α with fixed network size $N = 5000$ in Fig. 3. The two curves with different λ are almost

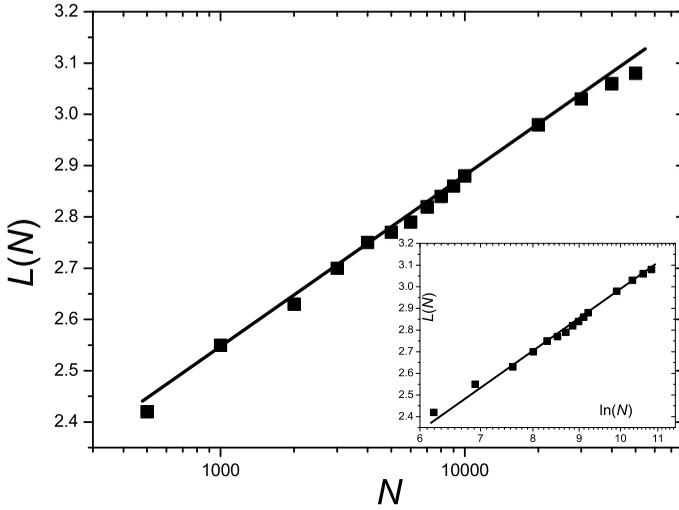


Fig. 1. The dependence between the average distance L and the network size N . One can see that L increases very slowly as N increases. The main plot exhibits the curve where L is considered as a function of $\ln N$, which is well fitted by a straight line. The curve is under the fitting line for sufficiently large N , which indicates that the increasing tendency of L can be approximated as $\ln N$ and in fact a little slower than $\ln N$. The inset shows the average distance L vs $\ln \ln N$, the error of linear-fitting by form $\ln \ln N$ is smaller than $\ln N$, indicating that the networks may be considered as ultrasmall world networks.⁵⁰ All the data are obtained by 10 independent simulations with parameters $\alpha = 1$ and $\lambda = 3$.

the same, thus the clustering coefficient is influenced little by λ . Both of the two curves increase monotonically with α , since α represents the degree of preferential attachment, this phenomenon reveals that the larger difference between attraction of preponderant and puny individuals will lead to greater clustering behavior.

Even when the network grows without preferential attachment (i.e., $\alpha = 0$), the clustering coefficient of our model is much greater than completely random networks, because of its special linking mode proposed here. For $\alpha > 1.5$, the clustering coefficient is approximate to 1, and the structure of corresponding networks is similar to a star in topology.^{52,53} The difference is that in our networks with very large α , the central part is not one node like star, but many nodes almost fully connected to each other. Since the structure for networks with $\alpha > 1.5$ is much different from reality, we will not discuss their characteristics hereinafter.

Summing up, the present networks possess both very large clustering coefficient and very small average distance which agree with previous empirical studies well.

4. Degree Distribution

The degree distributions of real-life networks are varied.^{54,55} Some of them, such as the acquaintance network of Mormons⁵⁶ are Gaussian; some such as the power-grid of southern California are exponential³⁵ and some such as the network of the World Wide Web are power-law.^{8,9} However, the degree distributions of most

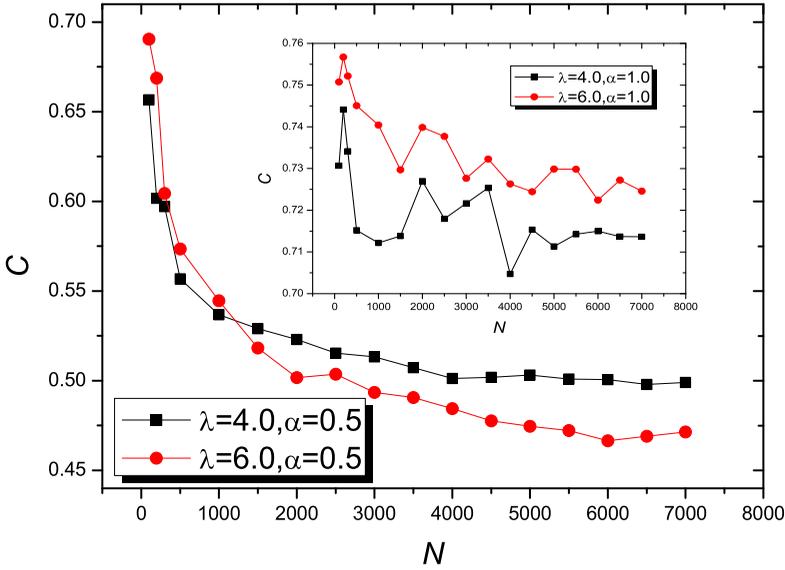


Fig. 2. The clustering coefficient vs network size N . The main plot and inset exhibit the dependence between the clustering coefficient C and network size N . One can see clearly that the clustering coefficient of the present networks is sufficient large even for big N .

real-life networks do not obey these simple forms above, they may interpolate between Gaussian and exponential ones such as the network of world airports,⁵⁴ or interpolate between exponential and power-law ones such as citation networks in high energy physics,⁵⁷ or in another form.⁵⁸

In this section, we focus on the empirical results of collaboration networks. In 2001, Newman investigated the statistic properties of scientific collaboration networks. He demonstrated that the degree distribution can be well fitted by an truncated power-law in the form $p(k) \sim k^{-\tau} e^{-(k/k_c)}$,²⁹ or considered as a double power-law.²⁸ Csányi and Szendrői have investigated the acquaintance networks from WIW project where the double power-law is also detected.³² In fact, Lehmann *et al.* have shown an example where the observed double power-law can be well fitted by a stretched exponential form.⁵⁷ In Appendix A, the details about stretched exponential distribution (SED) are shown, including the definition of SED, the basic properties of SED, the relations between SED and exponential distribution as well as power-law distribution, and the reason why we use SED in this paper. Figure 4 shows the degree distribution of the scientific collaboration network proposed by Newman^{28,29} which can be well fitted by SED with $c = 0.73$ indicating this distribution is more approximated to exponential form rather than power-law form. Another famous example is the collaboration network of movie actors,³⁵ which displays power-law only in its middle region. This distribution is also consistent with a stretched exponential form with $c = 0.45$ (see Fig. 5).

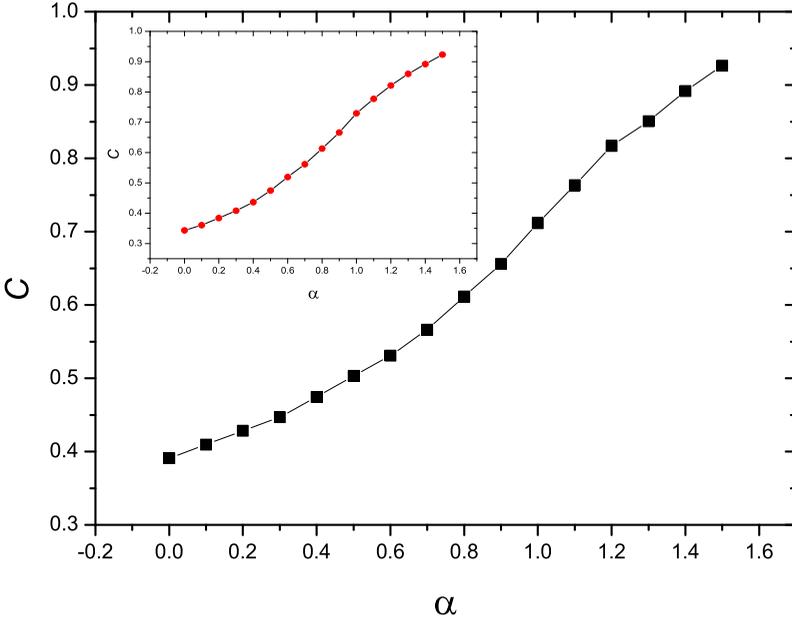


Fig. 3. The clustering coefficient versus preferential exponent α . The two curves can be considered as the clustering coefficient as a function of α with fixed network size $N = 5000$, which increase monotonically with the increasing of α . The main plot and inset are of $\lambda = 4.0$ and $\lambda = 6.0$ respectively. It is clear that the clustering coefficient is sensitive to the preferential exponent but influenced little by λ .

We did some empirical as well as theoretical works on collaboration networks and found that the degree distributions of many real-life collaboration networks in various fields obey the stretched exponential form approximately.⁵⁹ For example, if we consider the traveling sites as actors and the traveling routes that contain several sites as acts, then the recommended traveling routes from the web *Walkchina* and *Chinavista* in the year 2003 will form a Chinese touristry collaboration network, whose degree distribution is accurately consistent with SED of $c = 0.50$.⁵⁹

In succession, we will discuss the degree distribution of the networks generated from our model. Since the number of both the selected nodes and new edges are unfixed during each time step, it is hard for us to obtain the analytic results. For comparison, we will give analytic results for a special case of this model in Appendix B, and here, only the numerical results are shown. In Fig. 6, we report a typical simulation result with $\alpha = 1.0$ and $\lambda = 4$. The degree distribution is similar to that of movie actors collaboration network and can be well fitted by the stretched exponential form of $c = 0.34$. We have also investigated how the two parameters affect the degree distribution. In Fig. 7, one can see that the parameter c of SED monotonously decreases from 0.98 to 0.14 with α , the smaller α corresponds to a “more exponential” network while the larger one corresponds to a “more power-law”

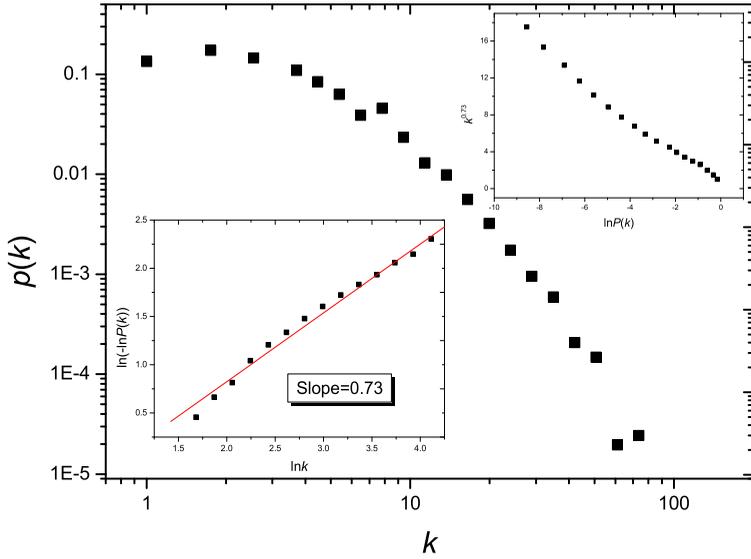


Fig. 4. The degree distribution of scientific collaboration network, where $p(k)$ denotes the probability a randomly selected node is of degree k and $P(k) = \int_k^\infty p(k)dk$ is the cumulation probability. The main plot is the degree distribution. The left-down inset shows how the quantity $\ln(-\ln P(k))$ behaves as a function of $\ln k$, which can be approximately fitted by a straight line of slope 0.73 ± 0.02 , thus these data obey SED of $c \approx 0.73$ (see Eq. (A3)). The right-up inset exhibits k^c versus $\ln P(k)$, which approximates to a line with negative slope.

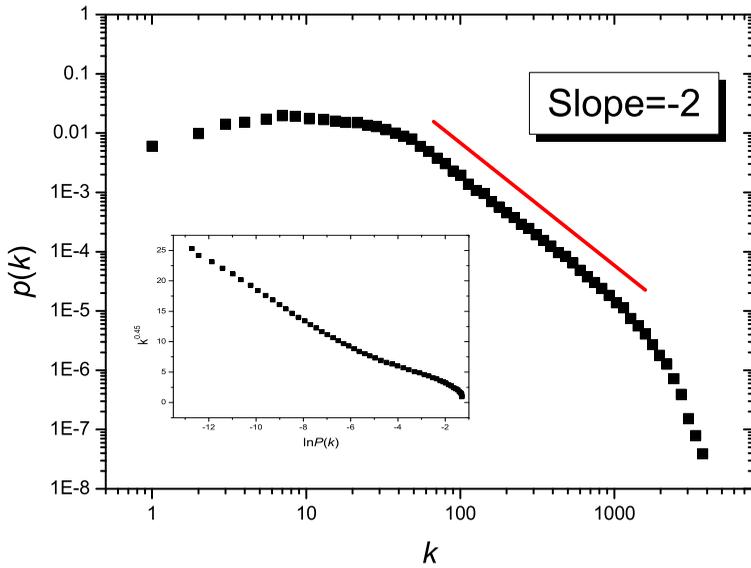


Fig. 5. The degree distribution of actors collaboration network. The main plot is the degree distribution that displays power-law only in the interval of about $k \in [50, 1000]$. The solid line is of slope -2 for comparison. These data can be approximately fitted by SED of $c = 0.45 \pm 0.02$. The inset exhibits k^c vs $\ln P(k)$.

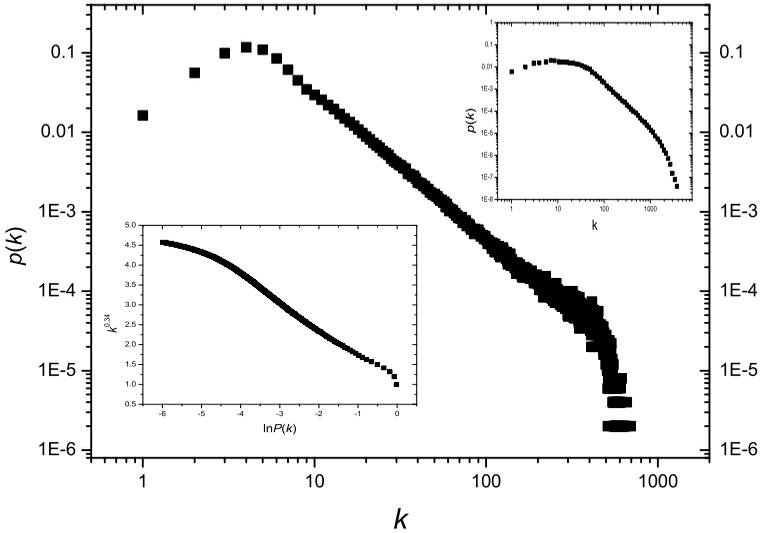


Fig. 6. A typical simulation result of degree distribution with $N = 5000$, $\alpha = 1.0$ and $\lambda = 4$. The main plot is the average of 100 independent simulations. The degree distribution exhibits observed power-law behavior in its middle region, which is similar to the case of movie actors collaboration network (see Fig. 5 or the right-up inset for comparison). The left-down inset shows $k^{0.34}$ versus $\ln P(k)$, which is approximated to a negative line indicating that the corresponding degree distribution can be well fitted by the stretched exponential form of $c = 0.34$.

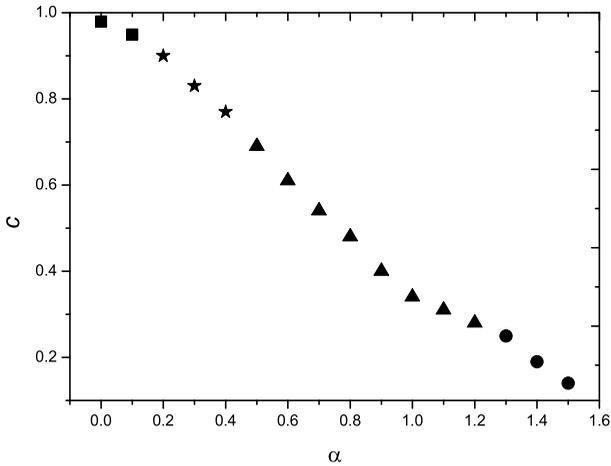


Fig. 7. The parameter c of SED as a function of preferential exponent α . The value of c monotonously decreases from 0.98 to 0.14 as α increases. All the data are the average of 100 independent simulations, where $N = 5000$ and $\lambda = 4$ are fixed. The pattern of degree distributions for different α can be roughly classified into four types: exponential (■), non-power (★), semi-power law (▲) and power law (●).

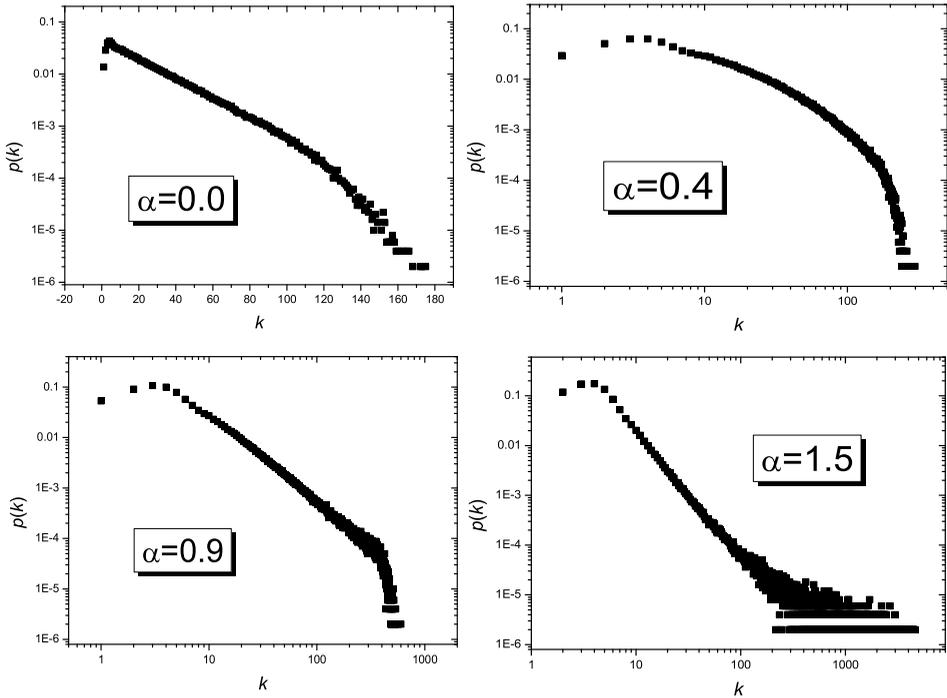


Fig. 8. The representative instances for the four patterns of degree distribution. **Exponential** ($\alpha = 0.0$), the degree distribution obeys exponential form except its tail; **non-power** ($\alpha = 0.4$), the degree distribution does not exhibit observed exponential or power law; **semi-power law** ($\alpha = 0.9$), the degree distribution exhibits observed power-law behavior only in its middle region; **power law** ($\alpha = 1.5$), the degree distribution displays power law in all the regions except a ridgy head and a fat tail.

one. As we mentioned above, for $\alpha > 1.5$, the networks are star-like in which the degree of hub node (i.e., the node of maximal degree) will exceed half of the network size, which has not been observed in the real-life collaboration networks, and will not be discussed hereinafter. To have an intuitionistic sight into the degree distribution of the present networks, we roughly classify those distributions into four patterns. They are **exponential**, **non-power**, **semi-power law** and **power law**. In Fig. 8, we show the representative instances for the four patterns. There are no unambiguous borderline between two neighboring patterns. We have also checked that the parameter λ affects the holistic property of degree distribution only a little; the larger λ only makes the head larger for sufficiently big N .

In a nutshell, many real-life collaboration networks are of degree distribution lying between exponential and power-law ones that can be well fitted by stretched exponential form, and the present model can generate networks of degree distribution from “almost exponential” to “almost power-law” containing four patterns.

5. Act-Size Distribution

Act-size distribution is another characteristic distribution besides degree distribution for collaboration networks, which is a particular distribution of collaboration networks. In many real-life cases, this distribution is single-peaked, and decays exponentially. One famous instance is the networks of corporate directors⁶⁰ in which the act-size distribution, defined as the number of directors per board, is single-peaked (see Fig. 8 in Ref. 55). We have also done some empirical works about act-size distribution of collaboration networks.⁵⁹ All of these networks, including Chinese touristry collaboration network, bus route network, scientific collaboration network, and so on, exhibit single-peaked act-size distribution. In Fig. 9, we show two examples, the Chinese touristry collaboration network and the scientific collaboration network of *Physical Review Letters*, respectively.

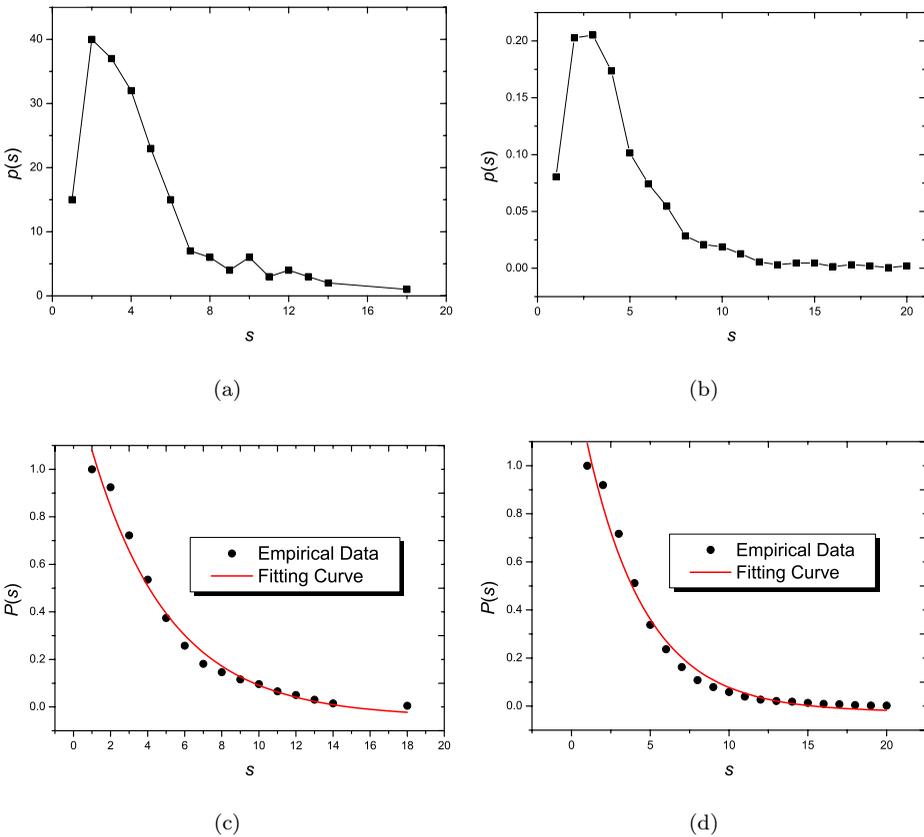


Fig. 9. Empirical results about act-size distribution. Figures (a) and (b) show the act-size distribution of Chinese touristry collaboration network and the scientific collaboration network of *Physical Review Letters*, respectively. Both the two distributions display obviously single-peaked behavior. Figure (c) and (d) are the corresponding cumulation distributions for those two networks. The red solid curves are the fitting curves of exponential form. In these four plots, the symbol s , $p(s)$ and $P(s) = \int_s^\infty p(s)ds$ denote act-size, the probability that a randomly selected act are of size s , and the cumulation probability, respectively.

oration network. In the former case, the act-size is the number of traveling sites per traveling route; the latter one only contains the 2062 papers in Vol. 93 of *Physical Review Letters*, where each paper is considered as an act and the act-size is the number of authors. There are 98 papers having authors more than 20, which have not been shown in Fig. 9. Both of the two distributions are single-peaked and in an approximately exponential form.

However, the act-size distribution seems not as attractive as degree distribution, thus the observed peaked behavior has not been emphasized before. It is always ignored,⁴⁰ or only considered as an extrinsic factor,³⁸ having nothing to do with and thus not being affected by the evolutionary mechanism of networks. In our model, the act-size distribution is not generated based on a static perspective like the degree distribution of configuration model,⁶¹ but an indiscerptible part of the dynamical mechanism of network evolution. It is clear that, when $\alpha = 0$, for sufficient large N , the act-size distribution is Poissonian distribution, single-peaked and decaying approximately exponentially. Contrary to the case of degree distribution, the numerical study indicates that the act-size distribution is insensitive to α .

A typical simulation result is shown in Fig. 10, one can compare this to the empirical data for scientific collaboration networks (see Figs. 9(b) and 9(d)). We set $\alpha = 0.4$ since it makes the parameter c of the two networks pretty much the same thing. Clearly, the act-size distribution generated by our model is consistent with the real-life one qualitatively.

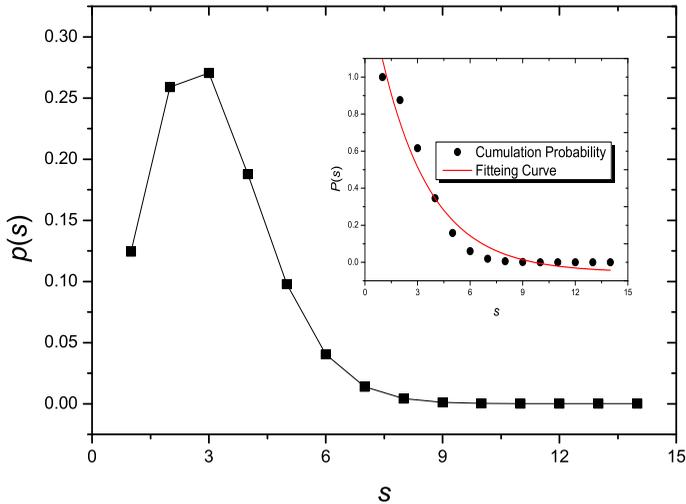


Fig. 10. A typical simulation result on act-size distribution with $N = 5000$, $\alpha = 0.4$ and $\lambda = 2.0$. The data are the average of 100 independent simulations. The main plot exhibits obviously single-peaked behavior. The inset shows the corresponding cumulation distribution, which can be well fitted by an exponential function (see the red solid curve).

6. Conclusion

In summary, we have constructed a general model for collaboration networks, the basic constituents of which are nonlinear preferential attachment and particular selecting and linking rules aiming at collaboration networks. The present networks are both of very large clustering coefficient and very small average distance, which is consistent with the previous empirical results that collaboration networks display small-world effect. We argue that, the degree distribution of many real-life collaboration networks may appropriately be fitted by stretched form. Numerical study indicates the degree distribution of the present networks can be well fitted by stretched form with the parameter c decreasing from 0.98 to 0.14 as the increasing of α . We roughly classify the degree distribution of our model into four patterns, **exponential** (bus route network in Beijing⁵⁹), **non-power** (scientific collaboration network^{28,29}), **semi-power law** (movie actors collaboration network³⁵ and Chinese touristry collaboration network⁵⁹) and **power law** (bus route network in Yangzhou⁵⁹), all of which are observed in the empirical data. More intriguing, we find that the act-size distribution is single-peaked and decaying exponentially, which can be reproduced by our model naturally.

Although this model is too simple and rough, it offers a good starting point to explain the existing empirical data and can be easily extended when more factors that may affect network evolution are considered. In addition, it is obvious that this model can be stretched to the weighted network model if the edge weight is used to represent the times of collaborations between the corresponding two nodes. The further statistical properties of the present networks, such as the degree-degree correlation, the clustering-degree correlation and so on have also been investigated, which will be published elsewhere.

Acknowledgments

We wish to thank Jun Liu for her help in preparing this manuscript and Prof. Zeng-Ru Di for his irradiative talk. B.H. Wang acknowledges the support by the National Natural Science Foundation of China (NNSFC) under Nos. 10532060, 10472116 and 10547004, the Special Research Funds for Theoretical Physics Frontier Problems under Grant No. A0524701, and Specialized Program under the Presidential Funds of the Chinese Academy of Science. D.R. He acknowledges the support by NNSFC under Nos. 70371071 and 10635040. K. Chen acknowledges the support by the National University of Singapore research grant R-151-000-028-112. T. Zhou acknowledges the support by NNSFC under Nos. 70471033 and 70571074, and the Graduate Student Foundation of University of Science and Technology of China under Grant Nos. KD2004008 and KD2005007.

Appendix A. Power Law and Stretched Exponential Distribution

Frequency or probability distribution functions (PDF) that decay as a power law have acquired a special status in the last decade. A power law distribution $p(x)$

characterizes the absence of a characteristic size: independently of the value of events x . In contrast, an exponential for instance or any other functional dependence does not enjoy this self-similarity. In other words, a power law PDF is such that there is the same proportion of smaller and larger events, whatever the size one is looking at within the power law range. Since the power law distribution has repeatedly been claimed to describe many natural phenomena and been proposed to apply to a vast set of social and economic statistics,^{62–68} it is considered as one of the most striking signatures of complex dependence. Empirically, a power law PDF is represented by a linear dependence in a double logarithmic axis plot of the frequency or cumulative number as a function of size. However, logarithms are notorious for contraction data and the qualification of a power law is not as straight-forwards as often believed. In addition, log-log plots of data from natural phenomena in nature and economy often exhibit a limit linear regime followed by a signature curvature. Latherrère and Sornette⁴⁷ explore and test the hypothesis that the curvature observed in log-log plots of distribution of several data sets taken from natural and economic phenomena might result from a deeper departure from the power law paradigm and might call for an alternative description over the whole range of the distribution. Thus, they propose a **stretched exponential distribution** (SED):

$$p(x)dx = c \left(\frac{x^{c-1}}{x_0^c} \right) \exp \left[- \left(\frac{x}{x_0} \right)^c \right] dx, \tag{A.1}$$

such that the cumulation distribution is

$$P(x) = \exp \left[- \left(\frac{x}{x_0} \right)^c \right]. \tag{A.2}$$

Stretched exponentials are characterized by an exponent c smaller than one. The borderline $c = 1$ corresponds to the usual exponential distribution. For c smaller than one, the distribution presents a clear curvature in a log-log plot. Based on the reasons discussed above, we use the SED to fit the degree distribution of our model. In the simulations, using the frequency $p(k)$ of degree k as x -axis and k^c as y -axis, we will obtain a line with negative slope, if the degree satisfy strict Stretched Exponential Distribution.

In the numerical case, write down the equivalent form of Eq. (A.2):

$$\ln(-\ln P(k)) = c \ln k - c \ln k_0. \tag{A.3}$$

Using $\ln k$ as x -axis and $\ln(-\ln P(k))$ as y -axis, if the corresponding curve can be well fitted by a straight line, then the slope will be the value of c .

Appendix B. A Special Model for Collaboration Networks of Fixed Act-Size

Under a very special case, the act-size is fixed. For example, if the four players in a bridge game are considered as actors in one act, then the act-size is 4. For comparison, in this appendix, we introduce a resolvable model for this special case.

This model starts with a m -complete network,^a where $m \geq 2$. At each time step, a new node is added and linked to all the nodes of a randomly selected m -complete network. Under these rules, not only the act-size, but also the number of new edges in each time step is fixed, which makes the model very easy to be analyzed.

Since after a new node is added to the network, the number of K_m increases by m , thus when the network is of order N , the number of K_m is $N_m = Nm - m^2 + m$. Note that, when a given node's degree increases by one, the number of K_m containing this node increases $m - 1$, hence for any node with degree k , it belongs to $\phi_k = km - k - m^2$ m -complete networks. Let $n(N, k)$ be the number of nodes with degree k when N nodes are present, now when we add a new node to the network, $n(N, k)$ evolves according to the following rate equation:⁴⁴

$$n(N + 1, k + 1) = n(N, k) \frac{\phi_k}{N_m} + n(N, k + 1) \left(1 - \frac{\phi_{k+1}}{N_m} \right). \tag{B.1}$$

When N is large enough, $n(N, k)$ can be approximated to $Np(k)$, where $p(k)$ is the probability density function for the degree distribution. In terms of $p(k)$, the above equation can be rewritten as:

$$p(k + 1) = \frac{N}{N_m} [p(k)\phi_k - p(k + 1)\phi_{k+1}]. \tag{B.2}$$

Using the expression $p(k + 1) - p(k) = \frac{dp}{dk}$, we can get the continuous form of Eq. (B.2):

$$p(k + 1) + \frac{N}{N_m} [(km - k - m^2) \frac{dp}{dk} + (m - 1)p(k + 1)] = 0. \tag{B.3}$$

Under the case $N \geq k \geq m$, this equation leads to $p(k) \propto k^{-\gamma}$ with $\gamma = (2m - 1/m - 1) \in [2, 3]$. The simulation result accurately agrees with the analytic one for large network size N .

In addition, there exists a bijection from node's degree to clustering coefficient as:

$$C(k) = \frac{(m - 1)(2k - m)}{k(k - 1)}. \tag{B.4}$$

The clustering coefficient of the whole network can be obtained as the mean value of $C(k)$ with respect to the degree distribution $p(k)$:

$$C = \int_{k_{\min}}^{k_{\max}} C(k)p(k)dk, \tag{B.5}$$

where $k_{\min} = m$ is the minimal degree and $k_{\max} \gg k_{\min}$ is the maximal degree. Combine Eqs. (B.4) and (B.5), note that $p(k) = Ak^{(2m-1/m-1)}$ and $\int_{k_{\min}}^{k_{\max}} Ap(k)dk = 1$, we can get the analytical result of C by approximately treating k_{\max} as $+\infty$. For example, when $m = 2, 3, 4, 5$ the clustering coefficients are 0.739,

^aHere, m -complete network means the networks of m nodes fully connected to each other, which is denoted by K_m in mathematical literatures.

0.813, 0.851 and 0.875. Further more, many real-life networks are characterized by the existence of hierarchical structure,⁶⁹ which can usually be detected by the negative correlation between the clustering coefficient and the degree. The BA network, which does not possess hierarchical structure, is known to have the clustering coefficient $C(x)$ of node x independent of its degree $k(x)$, while the present network has been shown to have $C(k) \sim k^{-1}$, in accordance with the observations of many real networks.⁶⁹

References

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. M. E. J. Newman, *SIAM Reviews* **45**, 167 (2003).
4. X.-F. Wang, *Int. J. Bifurcation & Chaos* **12**, 885 (2002).
5. M. Faloutsos, P. Faloutsos and C. Faloutsos, *Comput. Commun. Rev.* **29**, 251 (1999).
6. R. Pastor-Satorras, A. Vázquez and A. Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001).
7. G. Caldarell, R. Marchetti and L. Pietronero, *Europhys. Lett.* **52**, 386 (2000).
8. R. Albert, H. Jeong and A.-L. Barabási, *Nature* **401**, 130 (1999).
9. B. A. Huberman, *The Laws of the Web* (MIT Press, Cambridge, 2001).
10. J. Scott, *Social Network Analysis: A Handbook* (Sage Publications, London, 2000).
11. S. Wasserman and K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, 1994).
12. F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley and Y. Åberg, *Nature* **411**, 907 (2001).
13. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A.-L. Barabási, *Nature* **407**, 651 (2000).
14. J. Padani, Z. N. Oltvai, B. Tombor, A.-L. Barabási and E. Szathmary, *Nature Genetics* **29**, 54 (2001).
15. J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster and E. D. Gilles, *Nature* **420**, 190 (2002).
16. S. L. Pimm, *Food Webs* (University of Chicago Press, Chicago, 2002).
17. J. Camacho, R. Guimerà and L. A. N. Amaral, *Phys. Rev. Lett.* **88**, 228102 (2002).
18. R. J. Williams, E. L. Berlow, J. A. Dunne, A.-L. Barabási and N. D. Martinez, *Proc. Natl. Acad. Sci. USA* **99**, 12913 (2002).
19. J. A. Dunne, R. J. Williams and N. D. Martinez, *Proc. Natl. Acad. Sci. USA* **99**, 12917 (2002).
20. Y. He, X. Zhu and D.-R. He, *Int. J. Mod. Phys. B* **18**, 2595 (2004).
21. T. Xu, J. Chen, Y. He and D.-R. He, *Int. J. Mod. Phys. B* **18**, 2599 (2004).
22. T. Zhou, B.-H. Wang, P.-M. Hui and K. P. Chan, *Physica A* **367**, 613 (2006).
23. P. Sen, S. Dasgupta, A. Chatterjee, P. A. Sreeram, G. Mukherjee and S. S. Manna, *Phys. Rev. E* **67**, 036106 (2003).
24. J. R. Banavar, A. Maritan and A. Rinaldo, *Nature* **399**, 130 (1999).
25. G. B. West, J. H. Brown and B. J. Enquist, *Science* **276**, 122 (1997).
26. G. B. West, J. H. Brown and B. J. Enquist, *Nature* **400**, 664 (1999).
27. A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
28. M. E. J. Newman, *Phys. Rev. E* **64**, 016131 (2001).
29. M. E. J. Newman, *Proc. Natl. Acad. Sci. USA* **98**, 404 (2001).
30. Y. Fan, M. Li, J. Chen, L. Gao, Z. Di and J. Wu, *Int. J. Mod. Phys. B* **18**, 2505 (2004).

31. M. Li, Y. Fan, J. Chen, L. Gao, Z. Di and J. Wu, *Physica A* **350**, 643 (2005).
32. G. Csányi and B. Szendrői, *Phys. Rev. E* **69**, 036131 (2004).
33. G. F. Davis and H. R. Greve, *Am. J. Sociol.* **103**, 1 (1997).
34. M. E. J. Newman, D. J. Watts and S. H. Strogatz, *Proc. Natl. Acad. Sci. USA* **99**, 2566 (2002).
35. D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
36. A. Davis, B. B. Gardner and M. R. Gardner, *Deep South* (University of Chicago Press, Chicago, 1941).
37. C. R. Myers, *Phys. Rev. E* **68**, 046116 (2003).
38. J. J. Ramasco, S. N. Dorogovtsev and R. Pastor-Satorras, *Phys. Rev. E* **70**, 036106 (2004).
39. W.-X. Wang, B. Hu, T. Zhou, B.-H. Wang and Y.-B. Xie, *Phys. Rev. E* **72**, 046140 (2005).
40. M. Li, J. Wu, D. Wang, T. Zhou, Z. Di and Y. Fan, *Physica A* **375**, 355 (2007).
41. P.-P. Zhang, Y. He, T. Zhou, B.-B. Su, H. Chang, Y.-P. Zhou, B.-H. Wang and D.-R. He, *Acta Physica Sinica* **55**, 60 (2006).
42. M. C. González, P. G. Lind and H. J. Herrmann, *Phys. Rev. Lett.* **96**, 088702 (2006).
43. M. C. González, P. G. Lind and H. J. Herrmann, *Eur. Phys. J. B* **49**, 371 (2006).
44. P. L. Krapivsky, S. Redner and F. Leyvraz, *Phys. Rev. Lett.* **85**, 4629 (2000).
45. P. L. Krapivsky and S. Redner, *Phys. Rev. E* **63**, 066123 (2001).
46. R. Guimerà, B. Uzzi, J. Spiro and L. A. N. Amaral, *Science* **308**, 697 (2005).
47. J. Laherrère and D. Sornette, *Eur. Phys. J. B* **2**, 525 (1998).
48. Z.-M. Gu, T. Zhou, B.-H. Wang, G. Yan, C.-P. Zhu and Z.-Q. Fu, *Dyn. Contin. Discret. Impuls. Syst. B* **13**, 505 (2006).
49. T. Zhou, G. Yan and B.-H. Wang, *Phys. Rev. E* **71**, 046141 (2005).
50. A. F. Rozenfeld, R. Cohen, D. ben-Avraham and S. Havlin, *Phys. Rev. Lett.* **89**, 218701 (2002).
51. K. Klemm and V. M. Eguíluz, *Phys. Rev. E* **65**, 036123 (2002).
52. B. Bollobás, *Modern Graph Theory* (Springer-Verlag Publishers, New York, 1998).
53. J.-M. Xu, *Theory and Application of Graphs* (Kluwer Academic Publishers, 2003).
54. L. A. N. Amaral, A. Scala, M. Barthélémy and H. E. Stanley, *Proc. Natl. Acad. Sci. USA* **97**, 11149 (2000).
55. S. H. Strogatz, *Nature* **410**, 268 (2001).
56. H. R. Bernard, P. D. Killworth, M. J. Evans, C. McCarty and G. A. Shelley, *Ethnology* **27**, 155 (1988).
57. S. Lehmann, B. Lautrup and A. D. Jackson, *Phys. Rev. E* **68**, 026113 (2003).
58. A. Scala, L. A. N. Amaral and M. Barthélémy, *Europhys. Lett.* **55**, 594 (2001).
59. P.-P. Zhang, K. Chen, Y. He, T. Zhou, B.-B. Su, Y.-D. Jin, H. Chang, Y.-P. Zhou, L.-C. Sun, B.-H. Wang and D.-R. He, *Physica A* **360**, 599 (2006).
60. G. F. Davis, *Corp. Govern.* **4**, 154 (1996).
61. M. E. J. Newman, S. H. Strogatz and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001).
62. B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, New York, 1983).
63. R. Mantegna and H. E. Stanley, *Nature* **376**, 46 (1995).
64. Y.-B. Xie, B.-H. Wang, B. Hu and T. Zhou, *Phys. Rev. E* **71**, 046135 (2005).
65. P. Bak, *How Nature Works: The Science of Self-Organized Criticality* (Freeman, New York, 1994).
66. B.-H. Wang and P.-M. Hui, *Eur. Phys. J. B* **20**, 573 (2001).
67. M. E. J. Newman, *Contemporary Physics* **46**, 323 (2005).
68. P.-L. Zhou, C.-X. Yang, T. Zhou, M. Xu, J. Liu and B.-H. Wang, *New Mathematics and Natural Computation* **1**, 275 (2005).
69. E. Ravasz and A.-L. Barabási, *Phys. Rev. E* **67**, 026112 (2003).