Available online at www.sciencedirect.com
SCIENCE DIRECT
Physica A 360 (2006) 599–616
PHYSICA A
www.elsevier.com/locate/physa

Model and empirical study on some collaboration networks

Pei-Pei Zhang[a], Kan Chen[b], Yue He[a], Tao Zhou[c], Bei-Bei Su[a], Yingdi Jin[c], Hui Chang[a], Yue-Ping Zhou[a], Li-Cheng Sun[d], Bing-Hong Wang[e,b,c], Da-Ren He[e,a,b,*]

[a]*College of Physics Science and Technology, Yangzhou University, Yangzhou 225002, China*
[b]*Department of Computational Science, Faculty of Science, National University of Singapore, 117543 Singapore*
[c]*Nonlinear Science Center and Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China*
[d]*Department of Economics, National Chengchi University, Taipei 11623, China*
[e]*CCAST(World Laboratory), PO Box 8730, Beijing 100080, China*

Available online 14 June 2005

Abstract

In this paper we present an empirical study of a few practical systems described by cooperation networks, and propose a model to understand the results obtained. We study four non-social systems, which are the Bus Route Networks of Beijing and Yangzhou, the Travel Route Network of China, Huai-Yang recipes of Chinese cooked food, and a social system, which is the Collaboration Network of Hollywood Actors. In order to explain the results related to the degree distribution, act-degree distribution and act-size distribution (especially about the degree distribution, which may be better fitted using a stretched exponential distribution (SED)), we suggest a simple model to show a possible evolutionary mechanism for

*Corresponding author. College of Physics Science and Technology, Yangzhou University, Yangzhou, 225002 China.
E-mail address: darendo10@yahoo.com.cn (D.-R. He).

0378-4371/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2005.05.044

the emergence of such networks. The analytic and numerical results obtained from the model are in good agreement with the empirical results.
© 2005 Elsevier B.V. All rights reserved.

1. Introduction

The recent years have witnessed an upsurge in the study of complex systems. Many of the systems can be effectively modeled by networks in which the units of the systems are modeled as vertices, and the interactions among the units are modeled concisely as edges (with or without weights) between the vertices. This represents a new framework for building models of complex systems [1,2].

In a network the number of edges from a vertex or the number of all other vertices directly connected to it is called its vertex degree, or briefly degree, which is often denoted by "k". The probability of a vertex with degree k (or the degree distribution), $P(k)$, is the most important topological property of the network. We use the standard definition of degree distribution [3]: $P(k)$ stands for the number of vertices with degree k in the network. Other distributions are defined similarly. In a seminal work Barabasi and Albert investigated the statistics of many empirical networks, including the collaboration network of Hollywood actors, which will also be discussed in this paper, and showed that the degree distributions of the networks obey or approximately obey a power law, $P(k) \sim k^{-\gamma}$. They proposed a model (BA model), which illustrates a key main mechanism for the emergence of this scale-free distribution [1,4]. In the BA model, there are two main constituents: first, the number of vertices in a network increases one by one; second, the new vertex links to the old ones with a linear preferential attachment rule (so that "rich gets richer") [4]. The power-law distribution of the BA model is confirmed both analytically and numerically. Barabasi and Albert also explained analytically that if the new vertex links to the old ones randomly, the degree distribution will become an exponential distribution, $P(k) \sim \exp(-k/m)$, where m is a constant [1,4]. We believe that the use of the linear preferential attachment rule reflects the state of the system, which is far away from equilibrium. In such a state there is a lack of global coherence and the information utilization is mainly local (as in the fast nonequilibrium growth of a crystal). The non-equilibrium growth can easily enlarge the small initial differences (such as in the growth of DLA [5]).

A number of modified BA models have been proposed recently (see e.g., Refs. [6–10]). In this paper we describe three of them [7,9,10], which are relevant to our study. The first one was suggested by Liu and Lai [7]. Their model includes a competition between linear preference attachment and random attachment so that the degree distribution is between a power law and an exponential decay; they have used this to explain the empirical results on the scientific collaboration networks. The second one, proposed by Li and Chen [9], modeled the competition between linear preference attachment and random attachment in a more concrete and fundamental way. They used the World Trade Web as an example to illustrate that within the "local worlds" there exist "preferential" interactions among network units in many practical systems. Thus the rule of linear preference can only be applied within the local worlds, which are formed by choosing some seed units (vertices) randomly. The Li-Chen model also highlights the important role of random attachment. The local worlds are usually contained in the ranges where the units of the network have dense connections. The hubs in the network, which have the largest degree values, can usually have an influence within the local worlds they belong to. Outside the local worlds the units lack the extra information about "who is rich", and therefore make connections, very possibly, randomly. We believe that, although the "rich gets richer" rule is general in the systems far away from equilibrium, the mechanism of random attachment (choice) is usually present to compete with it. Therefore, according to these two models, the general degree distribution in practical systems should be between a power law and an exponential decay. Exact power law or exponential decay, which is induced by only one mechanism or tendency, should be an extreme case which occurs rarely. In the current paper, we shall present some empirical results to support this idea.

The third model we will discuss has been proposed by Ramasco, Dorogavtsev and Pastor-Satorras (RDP). The model focuses on (social) collaboration networks [10], which are the type of networks we study in this paper. But we shall first introduce collaboration networks.

There have been considerable of interest in the study of a special class of social networks, called social collaboration networks [11]. These include movie actor collaboration networks and scientist collaboration networks. This kind of networks can be described using bigraphs (bipartite graphs) as shown in Fig. 1 [1,10]. One type of nodes can be called "actor" such as movie actors or scientists, which are shown in the bottom row, indexed with $L_i,\ i=1,2,\ldots$. Another type of nodes can be called "act" such as movies or scientific papers, which are shown in the top row ($U_i,\ i=1,2,\ldots$). In these graphs, only undirected edges between different types of nodes are considered. We draw them in Fig. 1 with a solid line $e_i, i=1,2,\ldots$. An

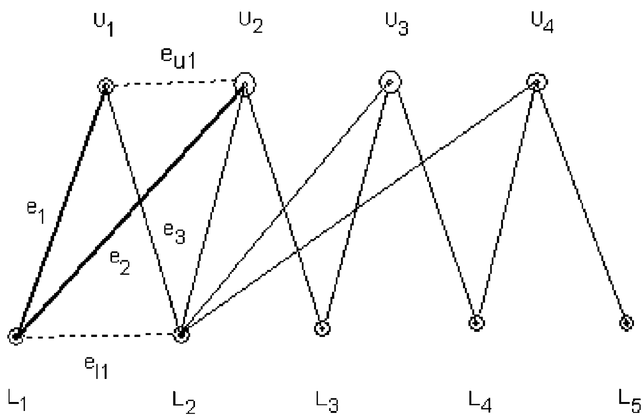

Fig. 1. A bi-partite graph for describing collaboration networks.

edge represents an actor taking part in an act. If we consider one type of nodes only, two edges sharing a common vertex in the bigraph are projected onto an edge between the two nodes of the same type. Take, for example, the movie actor collaboration network. Sometimes, we need to consider only the collaboration between actors. In this situation, edge e_{l1} between L_1 and L_2 in Fig. 1, which is obtained by projecting e_1 and e_3 to the bottom row, shows their collaboration in the same film U_1. If two actors cooperate in more than one film, the relation can be expressed by multiple edges between them. On the other hand, we can define an edge between two films (two U_i vertices), which indicates that the same common actor takes part in both films. The edge e_{u1} between U_1 and U_2 in Fig. 1, which is a projection of e_1 and e_2 to the top row, indicates that L_1 takes part in these two films. If two films involve more than one common actor, the relation can also be expressed by multiple edges. The larger the number of edges between two films is, the more similar characteristics these films share. Newman and Li used connection weights to denote multiple edges, and studied the resulting weighted networks in Refs. [12,13]. But we prefer to retain multiple edges to make our model clearer. To consider how many films actor i acted in all, we define a quantity h_i, "act-degree of actor i", which is equal to the number of U nodes linked to L_i in the bigraph, such as the four thin lines emitted from L_2 in Fig. 1. Obviously, the four U nodes U_1, U_2, U_3 and U_4 form a complete graph in the up-projected graph consisting of only U nodes. Similarly, if we have to consider how many actors are taking part in film j, we can define a quantity T_j, "act-size", which stands for the number of actors in act j, and it is equal to the number of L nodes linked by the node U_j in the bigraph. Again, these L nodes form a complete graph in the down-projected graph consisting of only L nodes. Each node has a degree value $T_j - 1$. Of course, two complete graphs may share one or more edges in the down-projected graph, however, the conclusion that the degree of each node equals $T_j - 1$ still holds when multiple edges are present. If we extend the concept of a "complete graph" to the situation where multiple edges are considered, i.e., define a graph where each pair of nodes are connected by edges (including multiple edges) as a complete graph, it is easy to verify that such a down-projected network is still a set of complete graphs.

The RDP model suggests organizing an "act", which includes T "actors" (T is the "act-size") in each step of time evolution. Among the T actors m are new participators, the other $T - m$ ones can be selected from the old ones with a probability proportional to the aforementioned "act-degree" h of each old vertex. In a simplified situation when T and m are both constants and the multiple edges between a pair of vertex are not considered, they analytically proved that, in the projected actor collaboration graph, the degree distribution $P(k)$ and the act degree distribution $P(h)$ are both exact power functions, and the scaling exponents are exactly the same, equal to $(2 + m)/(T - m)$. The simulation results from this model are in rather good agreement with the empirical results, which RDP obtained for a scientific co-authorship network and Hollywood movie actor collaboration network [10]. To construct our model, we extend the essential idea of RDP, but apply it to more complicated situations.

Many empirical studies have shown that the degree distribution is mostly between a power law and an exponential decay [6–9]. The degree distributions of these networks have been described as "having a power-law tail", "truncated power-law", "truncated exponential", "double power-law", etc. In this paper, we show that these distributions are better described using stretched exponential distribution (SED), as was proposed by Laherrete and Sornette in 1998 [14]. The distribution is given by $P(x)\,\mathrm{d}x = \mu(x^{\mu-1}/x_0^{\mu})\exp(-(x/x_0)^{\mu})\,\mathrm{d}x$ and its accumulative distribution is $P(x) = \exp(-(x/x_0)^{\mu})$, which can be conveniently stated as $\ln P(x) \sim x^{\mu}$, or $\ln(\ln P(x)) \sim \mu \ln x$. The latter form indicates a linear relation between $\ln(\ln P(x))$ and $\ln x$. Obviously, SED degenerates to an exponential distribution when μ is close to 1 and to a power law when μ is close to 0. When μ is between 0 and 1, the degree distribution is between a power law and an exponential function. The center region of an empirical distribution in a log–log plot may appear to have a linear part (scale-free region). However, there are typically curvatures in the head and tail parts. The closer the value of μ to 0, the wider the scale-free region [15,16].

In this paper we shall present, based on the BA, Liu-Lai, Li-Chen and RDP models, a model describing a collaboration network evolution, and suggest extension the concept "collaboration network" to non-social systems. The model will be presented in the second section. We will show that the model can be analytically solved in both extreme cases, but it can only be studied numerically in the more complicated cases in between. Our analytic and numerical analyses show that the degree distribution can take different forms: a power law or an exponential decay in the extreme cases, but an SED in between. In the third section, we shall present empirical studies of a number of "non-social collaboration networks", including Bus Route networks of Beijing (the capital and the largest city of China) and Yangzhou (a small city in China), Travel Route networks in China and a Recipe Network of Chinese cooked food, to support the conclusions of our model. We will also present our empirical study on the collaboration network of Hollywood movie actors in order to show that its degree distribution is better fitted by an SED function. The last section contains the summary.

2. The model

Suppose there are m_0 nodes at $t = 0$, which are connected and form some complete graph representing a number of acts. In each time step a new node is added. It connects to $T - 1$ old nodes selected according to a specified rule (which will be described shortly); a complete graph is formed consisting of these $T - 1$ old nodes and the new node by connecting all the possible edges. Below we discuss the degree distribution corresponding to different rules.

First we consider the rule of selecting $T - 1$ old nodes (T is a constant) with a probability proportional to the act-degree h_i of each old node i. This is the "act-degree linear preference rule", which means that, in the case of a network of movie actors, selecting a movie actor according to how many films he has acted in. We can, very similar to what BA did [1,4], obtain the evolution equation of the act-degree, h_i, using a quasi-continuous approximation

$$\frac{\partial h_i}{\partial t} = (T-1)\frac{h_i}{\sum_j h_j} \ . \tag{1}$$

The equation can be solved analytically using the method stated in Ref. [1,4]. The solution of the act-degree distribution is

$$P(h) = \frac{t}{(m_0+t)\beta}\, h^{-(1/\beta+1)} \ . \tag{2}$$

The act-degree distribution thus follows a power law with the scaling exponent, γ, equals $1/\beta + 1 = (2T-1)/(T-1)$. γ decreases as the act-size, T, increases. It tends to 2 when T tends to infinity. As mentioned in the last section, we have $k_i = h_i(T-1)$ when considering multiple edges; we can then obtain the degree distribution (with multiple edges counted) as

$$P(k) \sim k^{-\nu} \ , \tag{3}$$

where $\nu = \gamma$. Thus the degree distribution $P(k)$ and the act-degree distribution $P(h)$ are both exact power functions with the same scaling exponent. This is the same conclusion RDP obtained from their model.

Second, we consider the rule of selecting $T-1$ old nodes randomly. We can write down, following BA [1,4], the evolution equation of the act-degree as

$$\frac{\partial h_i}{\partial t} = (T-1)\frac{1}{m_0+t-1} \ . \tag{4}$$

It can be analytically solved with a similar method, and we obtain

$$P(h) = Be^{(1-h)/(T-1)} \tag{5}$$

or

$$\ln P(h) = \ln B + \frac{1}{T-1} - \frac{h}{T-1} \ , \tag{6}$$

where $B = (t+m_0-1)/[(T-1)(m_0+t)]$. Thus the act-degree distribution follows an exponential decay with a slope $\gamma = -1/(T-1)$ in a log–linear plot. γ also changes as the act-size, T, varies. It tends to 0 when T tends to infinity. Similarly, we can obtain the degree distribution (with multiple edges counted) as

$$\ln P(k) \sim -k/(T-1)^2 \ . \tag{7}$$

This means that the degree distribution $P(k)$ and the act-degree distribution $P(h)$ are both exact exponential functions, but on a log–linear plot, the slopes of the lines fitting the data are different. The slope for the degree distribution is the square of the slope for the act-degree distribution. This limit case has not been discussed in the context of cooperation networks previously. Our numerical simulation results based on this model show very good agreement with the analytic results obtained in these two extreme situations.

To interpolate between the above two extreme cases, we consider the rule of selecting the $T-1$ old nodes (T is a constant) randomly with a probability p, and using the act-degree linear preference rule with probability $1-p$. It is not difficult to imagine that the resulting degree distribution $P(k)$ and act-degree distribution $P(h)$ should be between a power law and an exponential function; SED is a good candidate for describing such a distribution. We have performed numerical investigations on the cases when p takes values between 0.1 and 0.9, all the results show that the degree distribution $P(k)$ and the act-degree distribution $P(h)$ can be fitted well using SED. The key parameter, μ, in these two SED functions are exactly the same. The details of the computation will be presented elsewhere.

The empirical investigation presented in the next section will show that T is in general not a constant. In all the systems we investigated, the distribution of act-size, $P(T)$, shows a functional form with an asymmetric peak. It seems difficult to find a common function to fit $P(T)$. However, the accumulative distribution of the act-size, $P(T \geqslant T_c)$, has less fluctuations and is more stable. The distribution of all the systems can be well fitted by a shifted Poisson distribution, given by $P(T) = (\lambda^{T+b}e^{-\lambda})/[A(T+b)!]$ in which λ is the value of T corresponding to the maximum of $P(T)$, T is an integer larger or equal to λ, and $1/A$ is the normalization factor.

We can easily extend our model to allow T chosen randomly from a probability distribution $P(T)$. At each time step we select T with probability $P(T)$. For the case of selecting $T-1$ old nodes with the probability proportional to their act-degrees $\{h_i\}$ to form a complete graph, the numerical simulation results of our model are shown in Fig. 2. In our simulation we take $\lambda = 8.5$ and $b = 0$. One can see from the figure that the data of both the degree distribution $P(k)$ and the act-degree distribution $P(h)$ are well fitted by straight lines on log–log plots; they are power functions with almost exactly the same exponent. For the case of selecting $T-1$ nodes randomly (after T is selected randomly with probability $P(T)$), the numerical simulation results of our model, shown in Fig. 3, confirm that both the degree

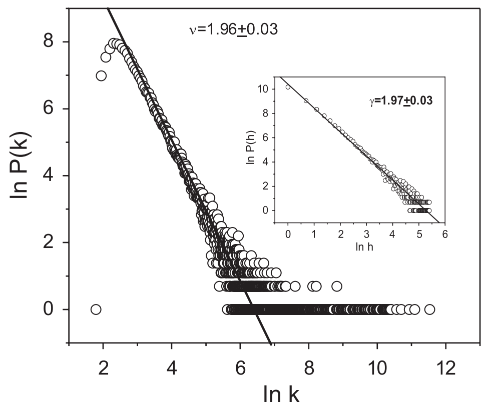

Fig. 2. The simulation results of degree distribution (multiple edges are counted) when the act-size follows a $P(T)$ distribution ($\lambda = 8.5, b = 0$) and act-degree linear preference rule is used. The inset shows an act-degree distribution in the same situation. The parameter in the model is chosen as $m_0 = 100$, $N = 50{,}000$ nodes were generated.

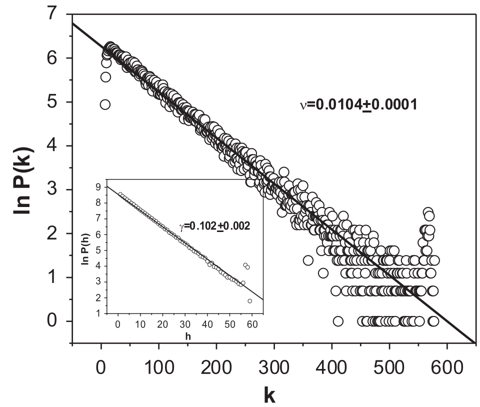

Fig. 3. The simulation results of degree distribution (multiple edges are counted) when the act-size follows a $P(T)$ distribution ($\lambda = 8.5, b = 0$) and random choice is used. The inset shows an act-degree distribution in the same situation. The parameter in the model is chosen as $m_0 = 100$, $N = 50{,}000$ nodes were generated.

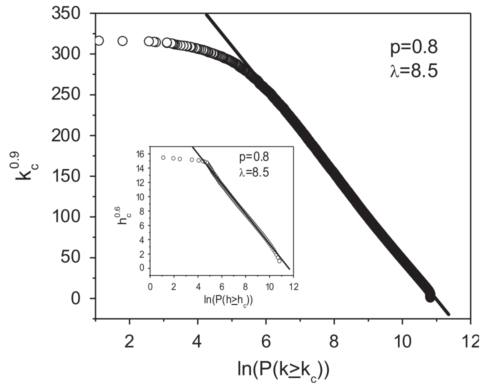

Fig. 4. The simulation results of degree distribution (multiple edges are counted) when the act-size follows a $P(T)$ distribution ($\lambda = 8.5, b = 0$) and random choice is used with the probability p and the act-degree linear preference rule is used with probability $1 - p$. The inset shows an act-degree distribution in the same situation. The parameter in the model is chosen as $m_0 = 10$, $p = 0.8$, $N = 100{,}000$ nodes were generated.

distribution $P(k)$ and act-degree distribution $P(h)$ are exponential functions. The slope in the linear–log plot for the degree distribution is the square of the slope for the act-degree distribution. For the case of selecting the $T - 1$ old nodes randomly with a probability p, and using the act-degree linear preference rule with probability $1 - p$, our numerical results show that both the degree distribution $P(k)$ and act-degree distribution $P(h)$ follow SED. However, the key parameter values, μ, are not exactly the same. Fig. 4 shows the numerical results when $p = 0.8$. The results obtained with other values of p are qualitatively the same. These results show that there is little qualitative difference in the model with a fixed T and the model with T

chosen from the probability distribution $P(T)$. It is important to know how the acts are being organized in the revolution process since the act-degree distribution determines the degree distribution. But the act-size distribution is rather unimportant.

3. Empirical study on some collaboration networks

In this section we present our empirical investigations on some practical systems, which can be described by collaboration networks. In order to characterize the collaboration among vertices (actors, stations in bus route, scenic spots, medicinal herbs and foodstuffs, etc.), we define the edges to denote their collaborations in one act (bus routes, travel routes, prescriptions in Traditional Chinese Medicine and recipes of cooked foods). In this way, two vertices connected by an edge "collaborate" as in a social collaboration network. Actors in one act also form a complete graph, and the whole network consists a set of complete graphs [15,16].

As the first example, a travel route can be thought of as a system of scenic spots or activities put together by a travel agency to satisfy the needs of tourists. Usually, the scenic spots in one travel route complement one another in scenery, conveyances, service, amusement, shopping, etc. for attracting tourists. Each scenic spot collaborates with others, contributes its own specialty and also shares the profit. If we denote the scenic spots as vertices and the collaboration between two scenic spots in a route as an edge, one route will form a complete graph. The whole travel route network consists of a set of such complete graphs. We choose 240 routes from the year 2003; there are a total of 171 vertices and 719 edges in the network. These data are small compared to classic social collaboration networks such as the Hollywood actors collaboration network and scientists collaboration network, but it represents a network of very different origin. There have been studies of small-size networks, such as the Jazz instrumentalists collaboration network [17], the underground railway routes network in Boston [18], the airline networks of China and US [19], the power grid network of China and US [4,20], the Indian railway network [21] and the company board member collaboration network [2,22]. One can still obtain valuable statistics about these networks, if the analysis of the data is done properly.

The inset of Fig. 5 shows the act-size distribution of the travel route network of China, which exhibits a single peak. As seen from the figure, the number of travel routes with two scenic spots is the largest, the number of routes with more than two scenic spots decrease gradually. Fig. 5 shows a typical example of an accumulative act-size distribution of the travel route network of China, which is a smooth function decreasing monotonously. As mentioned in the last section, we can describe the distribution with a shifted Poisson distribution, given by $P(T) = (\lambda^{T+b}\mathrm{e}^{-\lambda})/[A(T+b)!]$ in which λ is the value of T corresponding to the maximum of the function, T is an integer larger or equal to λ, and $1/A$ is the normalization factor. The large crosses in the figure represent the fitting by the $P(T)$ function.

Power-law distribution, often associated self-similarity, is a signature of a class of complex networks, called scale-free networks. However, many real complex

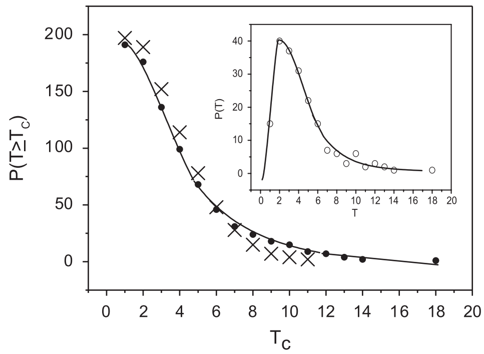

Fig. 5. Accumulative act-size distribution of Chinese travel routes network in 2003. The inset shows the corresponding act-size distribution. The solid or hollow circles represent the empirical data. The black curves only represent possible smooth connections of the data. The large crosses represent the fitting by the $P(T)$ function as explained in the text. The fitting parameters are: $\lambda = 7.5, b = 6$.

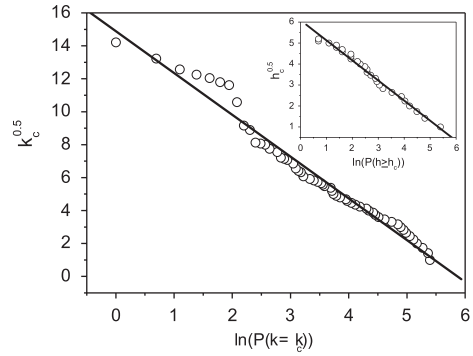

Fig. 6. The empirical results on the accumulative degree distribution (multiple edges are counted) of Chinese travel routes network in 2003. The inset shows the results of an accumulative act-degree distribution. The solid lines represent the least square fitting of the data.

networks are not strictly scale free. In the analysis of the empirical data, one has to be careful in dealing with a distribution, which might only resemble a power law in a limited region. Note that in the BA model linear preferential attachment gives rise to a power-law distribution, but random attachment gives rise to an exponential distribution. As illustrated in our model, most empirical distributions should be between these two extreme cases. If these distributions can be described using SED (which interpolates between these two extremes), then the value of μ is related to the proportion of preferential to random attachments.

Fig. 6 shows the accumulative degree distribution (with multiple edges counted) of the travel route network of China. The distribution can be well described using SED with $\mu = 0.5$. It reveals that there are both preferential and random mechanisms in the process of network evolution (Travel agencies tend to choose famous scenic spots, and also consider special conditions for different reasons; many different uncorrelated considerations give rise to random attachments). The inset of Fig. 6 shows the accumulative act-degree distribution of the travel route network of China. Compared with Fig. 6, we find that if the degree distribution is an SED function, the corresponding act-degree distribution shows the same form. In addition, the value of $\mu = 0.5$ is the same in SED fitting of the degree and the act-degree distributions for the travel route network of China. Our investigation shows that most of our empirical networks, such as Traditional Chinese Medicine network (will be presented elsewhere soon) and the network of Huai Yang recipes of Chinese cooked food (will be presented in the next paragraph), show similar qualitative results.

We choose 329 recipes of the Huai-Yang system (Huai-Yang denotes two geographical regions located in the middle-eastern part of China) of Chinese cooked food (modeled as complete graphs representing recipes or "acts"); there are a total of 242 vertices (foods or "actors") and 1713 edges in the network. Similarly, the inset of Fig. 7 shows the act-size distribution of the Huai-Yang recipe network, which exhibits a similar single peak. Fig. 7 shows the accumulative act-size distribution of the network, which can be described by the $P(T)$ function. Fig. 8 shows the accumulative act-degree distribution of the Huai-Yang recipe network of Chinese cooked food. The distribution can be well described using SED with $\mu = 0.2$. The inset of Fig. 8 shows the accumulative degree distribution of the network (with multiple edges counted). It is also of the SED form with a value of $\mu = 0.3$. These results are in good agreement with the conclusions deduced from our model.

We discuss three non-social networks above as examples of collaboration networks. However, these examples are small-size networks. For comparison we also study the Hollywood actors collaboration network, which contains 392,304

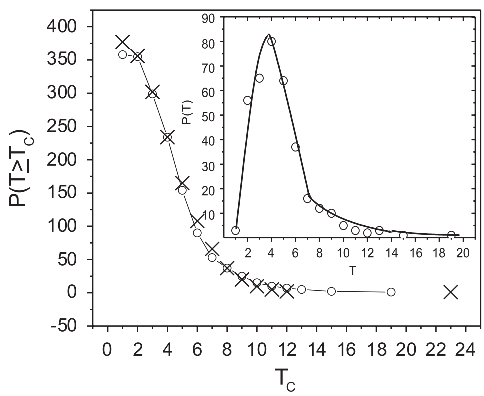

Fig. 7. Accumulative act-size distribution of Huai-Yang recipe network of Chinese cooked food. The inset shows the corresponding act-size distribution. The hollow circles represent the empirical data. The black curves only represent possible smooth connections of the data. The large crosses represent the fitting by $P(T)$ function as explained in the text. The fitting parameters are $\lambda = 8.5,\ b = 7$.

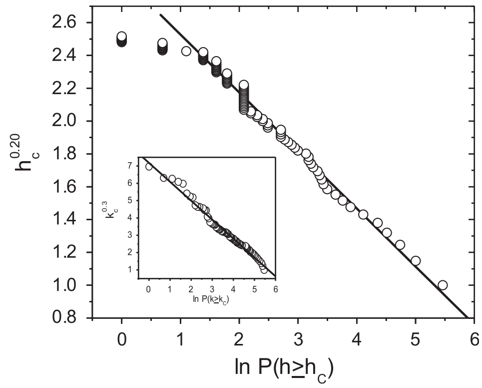

Fig. 8. The inset shows the empirical results of accumulative degree distribution (multiple edges are counted) of Huai-Yang recipe network of Chinese cooked food. The main figure shows the results of an accumulative act-degree distribution. The solid lines represent the least square fitting of the data.

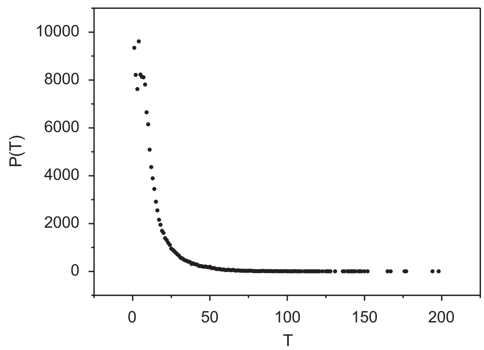

Fig. 9. The empirical results of act-size distribution of the Hollywood actors collaboration network.

vertices (film actors) and 181,455 films (acts). As was done in the previous studies, we linked two actors if they acted in the same film. So all actors in one film form a complete graph. Fig. 9 shows the act-size distribution of the Hollywood actors collaboration network. Similar to the inset of Figs. 5 and 7, it is a peak function with the peak at about $T = 10$. Most films have more than ten actors and the number of films having fewer actors decreases rapidly. The inset in Fig. 10 shows the degree distribution (with multiple edges counted) of Hollywood actors collaboration network in a log–log plot; there is a curvature in the tail as in the figure reported by RDP [10]. Fig. 10 shows that SED describes the data better. This result on the Hollywood actors collaboration network indicates that film producers select actors with both the preferential and random attachment rules. In fact, producers often choose famous actors without considering other reasons. Usually, the films involving famous actors have a better chance of success, because most people will be more likely to watch them than other films given limited information and chances of

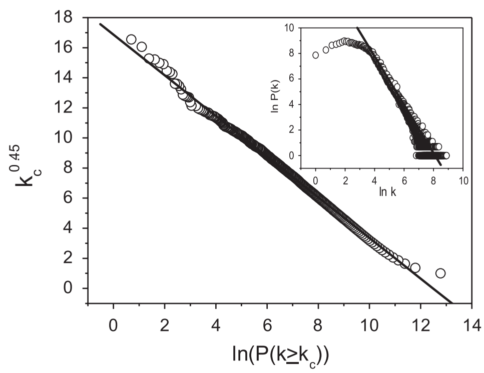

Fig. 10. Accumulative degree distribution of the Hollywood actors collaboration network. The solid line represents the least square fitting of the data (part of the data in the inset).

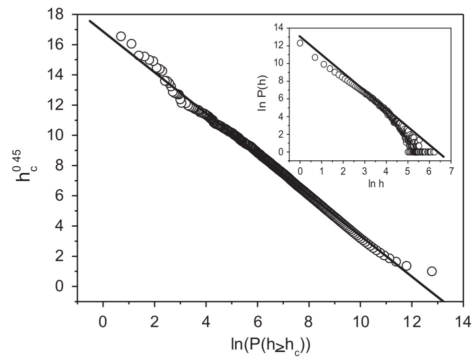

Fig. 11. Accumulative act-degree distribution of Hollywood actors collaboration network. The solid line represents the least square fitting of the data (part of the data in the inset).

trying. Thus, naively the preferential mechanism should be dominating. However, the data span more than one hundred years and no one can have so long an artistic life. So there is no pivot (or hub) controlling the overall situation, and there will not be a preferential mechanism for the whole network. In addition, even if the artistic life of an actor can last for such a long time, there still will not be a globally overwhelming preferential attachment, because of the need for new actors to bring in freshness to the cinema scene. In addition, there are also many special uncorrelated considerations in choosing actors. Thus, the strength of a random mechanism may be equal to the preferential one. This may explain the SED function obtained in the degree distribution. The inset of Fig. 11 shows the act-degree distribution of the Hollywood actors collaboration network in a log–log plot. Again, Fig. 11 shows that SED describes the data better. Comparing Figs. 10 and 11, we find that the degree distribution and the act-degree distribution have the same SED form with the same key parameter value $\mu = 0.45$.

The study of traffic networks is always of great interest. The transportation route (scheduled flight, coach number, bus route, etc.) can also be considered as an act, a station in an act as a vertex (actor), and the edge between two vertices expresses their collaboration in a common route. In this way, a route forms a complete graph as well. All such graphs joined together by the common vertices form the whole bus route network. We report a statistical investigation of the bus route network of Beijing, using the data in the year 2003. Fig. 12 shows the accumulative act-size distribution of the Beijing bus route network in 2003, which is very similar to Figs. 5 and 7. It can also be fitted with a shifted Poisson distribution. Fig. 13 shows the accumulative degree distribution (with multiple edges counted) of a bus route

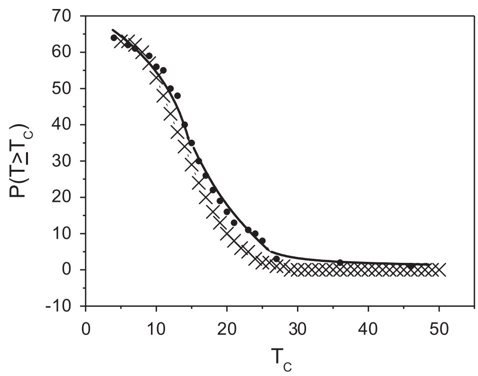

Fig. 12. Accumulative act-size distribution of the bus route network of Beijing in 2003. The solid circles represent the empirical data. The black curves only represent a possible smooth connection of the data. The large crosses represent the fitting by $P(T)$ function as explained in the text. The fitting parameters are $\lambda = 55.9,\ b = 50$.

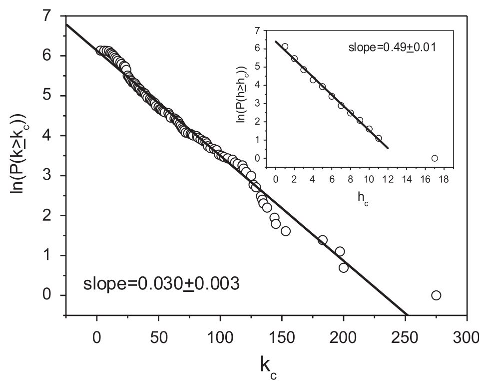

Fig. 13. The main figure shows the empirical results of accumulative degree distribution (multiple edges are counted) of the bus route network of Beijing in 2003. The inset shows the results of an accumulative act-degree distribution. The solid lines represent the least square fitting of the data.

network of Beijing in 2003. It follows an exponential distribution very well; this means that in the evolution of this network there is no dominating node. The edges come into being for their own somewhat uncorrelated special reasons; this leads to a network generated mostly by random attachments. The inset of Fig. 13 shows the accumulative act-degree distribution of the bus route network of Beijing. Again we find that both the distributions show qualitatively the same exponential function, although the slope of the line fitting to the data for the degree distribution is not exactly the square of the slope for act-degree distribution.

For comparison, we also investigate the bus route network of a small city in China, Yangzhou, in the year 2003. Similarly, stations in each route are defined as vertices. The stations appearing in the same route are joined with edges, which also form a complete graph. Fig. 14 shows the accumulative act-size distribution of the Yangzhou bus route network in 2003, which is also very similar to Figs. 5 and 7. Fig. 15 shows the accumulative degree distribution of the bus route network of Yangzhou in 2003, which is close to a power law. This is one of the few networks we investigated that approximately shows power-law degree distributions. Why do the bus route networks in Beijing and Yangzhou exhibit distinct degree distributions? Here is a probable answer. Among the 65 bus routes in Beijing, there are 17 routes going through a common station (TianAn Gate), which is the station with the largest number of routes going through it. But the second largest number is only 11 (at the station called Qian Gate). In the 460 stations, most have an act-degree smaller than 4. These are the usual characteristics of the bus routes in a large city. There are many routes and no station can control the human flow of the whole city and overwhelm all other stations. In fact, considering the convenience of public traffic and safety, such stations connecting many routes should be avoided. Thus, large cities such as Beijing do not have pivots (or hubs) in their bus route systems, especially those great pivots which can control the overall traffic. However, in the 26 bus routes in Yangzhou, the largest number of routes through a common station is 13, and the

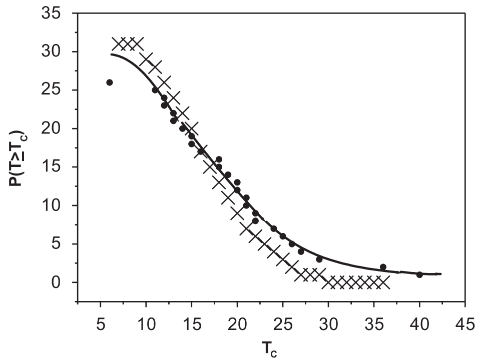

Fig. 14. Accumulative act-size distribution of bus route network of Yangzhou in 2003. The solid circles represent the empirical data. The black curves only represent a possible smooth connection of the data. The large crosses represent the fitting by $P(T)$ function as explained in the text. The fitting parameters are $\lambda = 60.5$, $b = 58$.

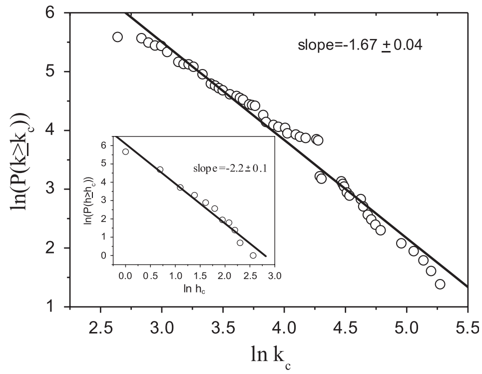

Fig. 15. The main figure shows the empirical results of the accumulative degree distribution (multiple edges are counted) of the bus route network of Yangzhou in 2003. The inset shows the results of an accumulative act-degree distribution. The solid lines represent the least square fitting of the data.

number decreases from 10, 9, 8, to 7 gradually. Among the 269 stations, most have an act-degree smaller than 2. These are the characteristics of the bus route system of a small city. The number of routes is small in such cities, and there are pivots (or hubs) on different levels controlling the overall traffic, including the centers that control most of the traffic. The inset of Fig. 15 shows the accumulative act-degree distribution of the bus route network of Yangzhou. We find that both distributions show qualitatively the same approximate power function, although the scaling exponents are not exactly the same.

4. Conclusion

We propose to study network systems consisting of overlapping complete graphs of actors collaborating in an act. In these systems, only the collaboration between actors in a common act is important and other relations such as their competition, confrontation, etc. can be neglected. Such systems are not limited to social networks. There are many other types of systems that can be modeled using these collaborating networks. Because of their common topological characters, their statistical description and evolution are similar too. First of all, participating in a common collaborating unit is the main mechanism for the development of such networks. Thus, we believe that act-degree distributions characterize the essential features of these networks. We demonstrated in our model and empirical study that the degree distribution is closely related to the act-degree distribution. As the analysis of the act-degree distribution is usually easier than the degree distribution, therefore, we can classify these networks easily with act-degree distributions.

Our model also illustrates the essential phenomenology of cooperation networks. In a cooperation network, actors typically lack global information (except for in a very small system); they only know the local information well and use it to collaborate effectively with others in limited ranges. In other words, they can join collaborating units based on the influence of other actors within their local regions. As the relevant information disappears outside the local ranges [9], naturally, the choices of the actors for collaboration outside the local range are mostly random. Thus, only in very small networks or in a network where information spreads quickly so that global information is easily accessible, exact power law distribution for the degree or act-degree may appear. In most cases, the degree and act-degree exhibit an intermediate distribution between a power law and an exponential function, and SED is a better description of it. The key parameter μ in SED describes the extent of the system deviating from equilibrium and of the circulation of information.

Acknowledgements

Kan Chen, Bing-Hong Wang and Da-Ren He are supported by the Science Foundation of National University of Singapore under the grant number R-151-000-032-112. Da-Ren He is also supported by National Natural Science Foundation of China under Grant no. 70371071. Bing-Hong Wang is also supported by National Natural Science Foundation of China under Grant no. 70271070, 10472116 and by the Specialized Research Fund for the Doctoral Program of Higher Education (SRFDP No. 20020358009). Tao Zhou is supported by National Natural Science Foundation of China under Grant no. 70471033 and by Graduate Student Creation Foundation of University of Science and Technology of China under Grant no. KD200408. All the authors would like to thank Prof. Hong-An Che at College of Systems Engineering in University of Shanghai for Science and Technology for the very helpful discussions he organized. We also thank Prof. Zhong-Zhi Zhang and Assoc. Prof. Li-LI Rong for sending their very helpful preprint.

References

[1] R. Albert, A.-L. Barabasi, Rev. Mod. Phys. 74 (2002) 47.
[2] M.E.J. Newman, SIAM Rev. 45 (2003) 167.
[3] P.L. Krapivsky, S. Redner, in: R. P-Satorras, M. Rubi, A. D-Guilera (Eds.), Statistical Mechanics of Complex networks, Springer, Berlin, Heidelberg, 2003, p. 4.
[4] A.-L. Barabasi, R. Albert, Science 286 (1999) 509.
[5] T.A. Witten, L.M. Sander, Phys. Rev. Lett. 47 (1981) 1400.
[6] L.A.N. Amaral, et al., Proc. Natl. Acad. Sci. USA 97 (2000) 11149.
[7] Z. Liu, Y.-C. Lai, et al., Phys. Lett. A 303 (2002) 337.
[8] M.E.J. Newman, Phys. Rev. E 64 (2001) 016131.
[9] X. Li, G. Chen, Physica A 328 (2003) 274.
[10] J.J. Ramasco, S.N. Dorogovtsev, R. Pastot-Satorras, Phys. Rev. E 70 (2004) 036106.
[11] S. Wasserman, K. Faust, Social Network Analysis: Methods and Applications, Cambridge University Press, Cambridge, 1994.
[12] M.E.J. Newman, Phys. Rev. E 64 (2001) 016132.
[13] M. Li, J. Wu, D. Wang, T. Zhou, Z. Di, Y. Fan, arXiv: cond-mat/0501655.

[14] J. Laherrere, D. Sornette, Eur. Phys. J. B 2 (1998) 525.
[15] Y. He, P. Zhang, B. Su, D.-R. He, Bull. of APS 49 (1) (2004) 1006;
Y. Zhang, Y. He, D.-R. He, Bull. APS 49 (1) (2004) 1007.
[16] A. Sun, P. Zhang, Y. He, B. Su, D.-R. He, Bull. APS 49 (1) (2004) 1006;
J. Hu, P. Zhang, R. Qiu, J. Tang, D.-R. He, Bull. APS 49 (1) (2004) 1008.
[17] P.M. Gleiser, L. Danon, Adv. Complex Syst. 6 (2003) 565.
[18] V. Latora, M. Marchiori, Physica A 314 (2002) 109.
[19] Y. He, X. Zhu, D.-R. He, Int. J. Mod. Phys. B 18 (17–19) (2004) 2595;
L.-P. Chi, R. Wang, H. Su, X.-P. Xu, J.-S. Zhao, X. Cai, Chin. Phys. Lett. 20 (2003) 1393
W. Li, X. Cai, arXiv: cond-mat/0309236.
[20] T. Xu, J. Chen, Y. He, D.-R. He, Int. J. Mod. Phys. B 18 (17–19) (2004) 2599.
[21] P. Sen, S. dasgupta, A. Chatterjee, P.A. Sreeram, G. Mukherjee, S.S. Manna, Phys. Rev. E 67 (2003) 036106.
[22] M.E.J. Newman, D.J. Watts, S.H. Strogatz, Proc. Natl. Acad. Sci. USA 99 (2002) 2566.