Available online at www.sciencedirect.com
ScienceDirect
Physica A 383 (2007) 687–702
PHYSICA A
www.elsevier.com/locate/physa

Assortativity and act degree distribution of some collaboration networks

Hui Chang[a], Bei-Bei Su[b], Yue-Ping Zhou[a], Da-Ren He[a,*]

[a] *College of Physics Science and Technology, Yangzhou University, Yangzhou 225002, China*
[b] *Wuxi Professional College of Science and Technology, Wuxi 214028, China*

Received 11 February 2007; received in revised form 18 April 2007
Available online 27 April 2007

Abstract

Empirical investigation results on weighted and un-weighted assortativity, act degree distribution, degree distribution and node strength distribution of nine real world collaboration networks have been presented. The investigations propose that act degree distribution, degree distribution and node strength distribution usually show so-called "shifted power law" (SPL) function forms, which can continuously vary from an ideal power law to an ideal exponential decay. Two parameters, α and η, can be used for description of the distribution functions. Another conclusion is that assortativity coefficient and the parameter, α or η, is monotonously dependent on each other. By the collaboration network evolution model introduced in a reference [P. Zhang et al., Physica A 360 (2006) 599], we analytically derived the SPL distributions, which typically appeared in general situations where nodes are selected partially randomly, with a probability p, and partially by linear preferential principle, with the probability $1-p$. The analytic discussion gives an explicit expression on the relationship between the random selection proportion p and the parameters α and η. The numerical simulation results by the model show a monotonic dependence of assortativity on the random selection proportion p. The empirically obtained assortativity versus α or η curve for the four collaboration networks with small maximal act size, T_{max}, shows a good agreement with the model prediction. According to the curves, one can qualitatively judge the random selection proportion of the real world network in its evolution process.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Collaboration network; Assortativity; Act degree

1. Introduction

There has been considerable interest in the study of social collaboration networks including Hollywood actor collaboration network and scientist collaboration network [1–7]. The kind of networks can be described by bipartite graphs. In these graphs one type of vertices is "actors" (such as movie actors or scientists) taking part in some activities, organizations or events. The other type of vertices is the activity, organization or event named "acts" (such as movies or scientific papers). For description of collaboration networks, a projected single-mode (unipartite) network of a bipartite graph can be used, which contains only one type of nodes,

*Corresponding author.
E-mail addresses: darendo10@yahoo.com.cn, darenhe@gmail.com (D.-R. He).

0378-4371/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2007.04.045

actors. In the unipartite graph all actors participating in the same act are connected by links. Thus, each act is represented by an "act complete subgraph (ACSG)". It is worth noting that every complete graph with h nodes can be subdivided into some complete subgraphs with m nodes ($m<h$), therefore, a complete subgraph in the unipartite graph may not be an ACSG. In social network theory, a clique is defined as a maximal complete subgraph of three or more nodes [4]. Most of the ACSGs are cliques, however, we observed empirically that some ACSGs might have only two (even only one) actor-nodes. For example, one scientist may write a scientific paper. Additionally, a research group director may select some people among the previous paper authors to do the next research. The authors of the new paper will be part of the authors of the previous one. In the projected unipartite graph the first paper ACSG subsumes the second smaller ACSG, which obviously is not maximal. Therefore an ACSG may not be a clique. To handle real world collaboration networks, we define ACSGs from a practical standpoint to reflect the collaboration relationships in the application domain. In other words, we define an ACSG as the collection of all nodes that collaborate with each other in the act according to the empirical data.

The number of nodes in an ACSG is addressed as "act size" and denoted by T. The number of ACSGs, in which an actor takes part, is addressed as "act degree" of the actor nodes and denoted by h. One vertex i's degree k_i is defined as the number of its adjacent edges and can be expressed as

$$k_i = \sum_{j \in \Gamma(i)} a_{ij}, \tag{1}$$

where $\Gamma(i)$ represents the set of neighbor nodes of node i, and a_{ij} denotes the element of the adjacency matrix. Obviously, in ACSG j, every node has the degree value $T(j) - 1$ [4,8,9]. In this article we use the standard definition of degree distribution [10]: $P(k)$, the probability of a node with degree k, stands for the number of nodes with degree k in the network. Other distributions are defined similarly.

Newman has proposed a parameter, assortativity, for a description of degree–degree correlation [11]. Assortativity is denoted by r, which lies in the range of $-1 \leqslant r \leqslant 1$. If $r>0$, the network shows positive nearest neighbor degree–degree correlation and is named "Assortative Mixing". If $r<0$, the network shows negative nearest neighbor degree–degree correlation and is named "Disassortative Mixing". The formula, which Newman proposed to calculate degree–degree assortativity coefficient in an undirected network [11], is

$$r = \frac{M^{-1}\sum_{\phi}(\prod_{i\in F(\phi)} k_i) - ((M^{-1}/2)\sum_{\phi}(\sum_{i\in F(\phi)} k_i))^2}{(M^{-1}/2)\sum_{\phi}(\sum_{i\in F(\phi)} k_i^2) - ((M^{-1}/2)\sum_{\phi}(\sum_{i\in F(\phi)} k_i))^2}, \tag{2}$$

where $F(\phi)$ denotes the set of the two vertices connected by the ϕth link and M is the total number of links in the network [11].

In many previous collaboration network investigations the networks were defined such that the edges are either present or not. The elements of adjacency matrix can take either 1 (two nodes are connected) or 0 (disconnected). This may cause the inaccuracy of some statistical properties since "cooperation strength" information was ignored. We may discuss transportation networks as an example. A lot of transportation networks were investigated in two different topological spaces: Space P and Space L [9,12–17]. In Space P, one node represents, for example, one bus stop, and one edge is connected between a pair of stops when at least one bus route provides direct service. In other words, an edge means that passengers can take at least one bus route for a direct travel between the two bus stops. If passengers have to transfer bus routes then the pair of stops is connected by more than one edge. This topology can be defined as a collaboration network. The bus routes can be viewed as collaboration acts, and bus stops can be regarded as actors. All the bus stops serviced by a common bus route and the edges between them form an ACSG. In Space L an edge between two nodes exists if they are consecutive stops on at least one bus route although node definition is the same. This topology cannot be regarded as a collaboration network. Actually, in Space P transportation networks and other collaboration networks the edges are different, characterizing by different capacity or the amount of traffic. In ordinary un-weighted Space P, the edges only express the existence of direct bus routes between the pair of stops. The information about the number of bus routes should be attached on the edges as an edge weight to show the "cooperation strength". This discussion is effective for all the collaboration networks [18–25]. In such weighted collaboration networks the elements of adjacency matrix, w_{ij}, can take integers, which equals the edge weight, and may be greater than 1. Degree of node i should be replaced by s_i, "the node i's strength", in the weighted network, which is defined as

$$s_i = \sum_{j \in \Gamma(i)} w_{ij}. \tag{3}$$

Degree distribution should be replaced by node strength distribution, which describes the topological features with cooperation strength considered.

To the best of our knowledge, Ref. [25] may be the only reference where a definition of weighted assortativity was suggested. This quantity was defined as

$$r^w = \frac{H^{-1}\sum_\phi (w_\phi \prod_{i\in F(\phi)} k_i) - ((H^{-1}/2)\sum_\phi (w_\phi \sum_{i\in F(\phi)} k_i))^2}{(H^{-1}/2)\sum_\phi (w_\phi \sum_{i\in F(\phi)} k_i^2) - ((H^{-1}/2)\sum_\phi (w_\phi \sum_{i\in F(\phi)} k_i))^2}, \tag{4}$$

where w_ϕ denotes the weight of the ϕth link and H is the total weight of all links in the network [25]. Both the two weighted quantities, s and r^w, reduce to the corresponding un-weighted definitions if all the w_{ij} equals 1 [18–25]. We also note that there has been effort for unifying assortative and disassortative weighted network characterization [26,27].

Degree distribution often attracts attentions since they significantly describe the topological structure of networks and may imply the evolution characteristics of networks. For example, it is widely acknowledged that many real world networks showed power-law degree distribution induced by linear preferential evolution mechanism [6]. Assortativity also reveals some aspects of network topology. Therefore it is important to search the relationship between degree distribution (or act degree distribution that may be more important in collaboration networks) and assortativity.

The authors of Refs. [8,9] presented models, which led to a conclusion that, when collaboration networks evolved by linear preferential principle, act degree distribution and degree distribution showed same power-law function form. The model presented in Ref. [9] also discussed the situation when collaboration networks evolved by random selection. This evolution mechanism induces exponential act degree distribution and degree distribution. The general cases, where collaboration networks evolved partially by random selection and partially by linear preferential principle, were discussed in Ref. [9]; however, the model for the cases was not analytically solved. Instead, the authors (cooperators and we) only made a guess that act degree distribution and degree distribution should show similar stretched exponential distribution (SED), which interpolates between the power-law distribution and the exponential distribution, as was proposed by Laherrete and Sornette in 1998 [28]. Some empirical proofs for the conclusions have been reported in Ref. [9]. The proofs were obtained in some real world networks: the bus route networks (BRN) of Beijing (urban) and Yangzhou in 2003, the Travel Route Network of China (TRNC), Huai-Yang recipes of Chinese cooked food (HYRCCF), and the Collaboration Network of Hollywood Actors (CNHA). In Ref. [8] assortativity of the proposed model was discussed but only in the case where un-weighted collaboration networks evolved by linear preferential principle. As mentioned, more general evolving situations of collaboration networks were considered in Ref. [9], however neither un-weighted nor weighted assortativity was discussed.

In this paper, we shall present empirical investigation results of some new real world networks, and some new property results of the networks, which were already discussed in Ref. [9]. We shall show that in these real world collaboration networks, in general cases, act degree distribution, degree distribution and node strength distribution show so-called "shifted power law (SPL) distributions", which can be analytically deduced by the model presented in Ref. [9]. Empirical investigation and model simulation results on the weighted and un-weighted assortativity will also be presented. We shall discuss the relationship between assortativity and the parameters, which describe the SPL distributions, and compare the model prediction about this relationship with the empirical data of some real world networks.

The other parts of the text will be as follows. In Section 2 we shall present empirical investigation results about degree and node strength distribution, act degree distributions and act-size distributions. In Section 3 we shall report empirical investigations on un-weighted and weighted assortativity. In Section 4 we shall analytically and numerically discuss the model introduced in Ref. [9] and try to explain the empirical results. In the last section, Section 5, we shall summarize the text and make some discussions.

2. Empirical study on degree, act degree, act-size and node strength distributions

2.1. Traditional Chinese herb prescription formulation network

We shall, firstly, present empirical investigation on traditional chinese herb prescription formulation network (TCHPFN).

Chinese herbology is a result from the simple dialectical materialism philosophy of the traditional Chinese. It emphasizes the inner cause of any illness and states that a healthy body should be able to maintain a dynamic balance against the world outside of the body through self-adjustment. Therefore, any illness is due to unbalance between inner self and outside world. However, the balance is affected by many different and complicated changing factors, and unbalance cannot always be treated by a single herb, which is able to attack just a few problems. In addition, any herb has side effects. It may be effective in dealing with one factor but it would cause other unbalance. So we need to combine appropriate herbs to make an effective prescription where different herbs work together in a complementary manner in order to cure an illness and minimize the side effects. In every Chinese herb prescription, there are usually four general components. They are "Jun", i.e., major herbs attacking the illness—the main treatment; "chen", i.e., the auxiliary herbs which boost the effect of the main treatment; "zuo", i.e., booster herbs that catalyze the major and auxiliary herb effects; "shi", i.e., herbs that guide and orchestrate the power of the whole medicine to the point of illness. We symbolize every herb as a node in the graph, and draw a link between herbs included in a prescription to represent their interactions, i.e., their collaboration relationship at the process of curing an illness. A prescription is then represented as an ACSG. Some herbs are shared among different prescriptions and can be seen as bridges between different ACSGs. The set of ACSGs then is now a network of prescriptions.

We have included 1536 prescriptions (ACSGs) and 681 herbs (actors) from Refs. [29,30] for our study. These herbal prescriptions are the results of long-term experiments conducted by the Chinese people and serve as the representative samples of the prescription population. According to our research, a herbal prescription (ACSG) contains at most 15 different herbs (two such cases) and at least one herb (168 such cases). It is most probable that a prescription would contain three different herbs (306 such cases), and becomes less possible when the number of herbs in a prescription increases or decreases.

In the study, we find out that liquorice has the maximum degree in the Chinese herb prescription network, i.e., it is the most popular herb. Following liquorice are tuckahoe, ginseng, angelica, Milkvetch root, largehead atractylodes rhizome in a decreasing order. These are either mild auxiliary herbs or boosters for the auxiliaries. They are widely used in different prescriptions becoming the major bridges among different ACSGs in the network. On the other hand, herbs of small degree are rarely used which are highly specialized major or auxiliary herbs such as hedgehog skin, ginkgo seed, white haricot, Stemonaceae, silk, etc. The average act degree of nodes in TCHPFN is 9.21, meaning that each herb, on average, takes part in little more than nine prescriptions.

Inset of Fig. 1 shows the act-size distribution, $P(T)$, of the TCHPFN. Fig. 1 shows the accumulative act-size distribution, $P(T \geqslant T_c)$, of the TCHPFN, which can be described with a shifted Poisson distribution, $P(T) = (\lambda^{T+b}/A(T+b)!)\mathrm{e}^{-\lambda}$. In it 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. Fig. 2 shows the accumulative act degree distribution, $P(h \geqslant h_c)$, of the TCHPFN. Fig. 3 shows the accumulative un-weighted degree distribution, $P(k \geqslant k_c)$, of the TCHPFN. Fig. 4 shows the accumulative node strength (weighted degree) distribution, $P(s \geqslant s_c)$, of the TCHPFN. The three distributions shown in Figs. 2–4 can be well described with SPL functions. Take $P(k)$ as an example, SPL function can be expressed as

$$P(k) \propto (k+\alpha)^{-\eta}, \tag{5}$$

where η and α are constants. The function can be shown by a linear line with a slope value η on the $\ln P(k) - \ln(k+\alpha)$ plane. For $\alpha = 0$, one finds that

$$P(k) \propto k^{-\eta}, \tag{6}$$

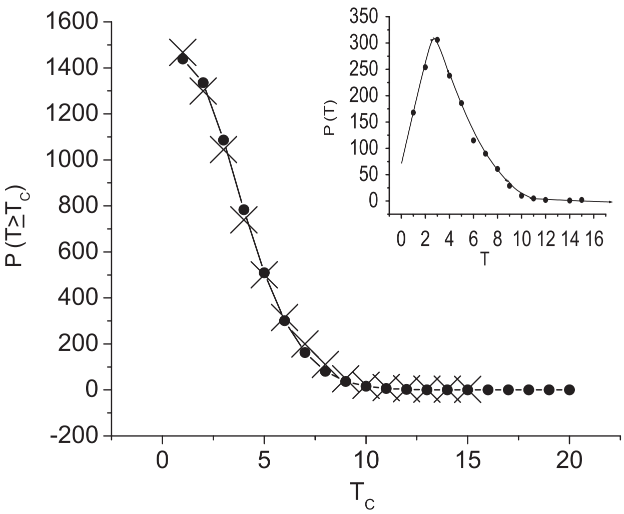

Fig. 1. Accumulative act-size distribution of TCHPFN. The inset shows the corresponding act-size distribution. The solid circles represent the empirical data. The black curves only represent possible smooth connections of the data. The large crosses represent the fitting by $P(T)$ function, which is presented in the text. The fitting parameters are: $\lambda = 6.5$ and $b = 6$.

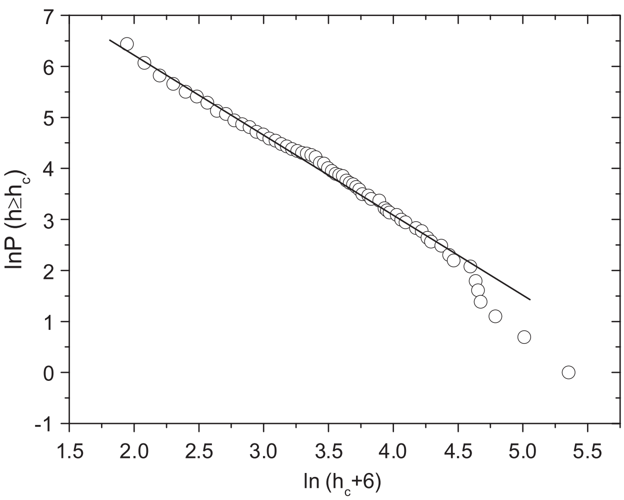

Fig. 2. The empirical results on the accumulative act-degree distribution of TCHPFN.

which indicates a power-law $P(k)$ distribution with the scaling exponent η. For $\alpha \to \infty$, it is easy to show that $P(k)$ tends to an exponential distribution:

$$\ln P(k) \propto (-k). \tag{7}$$

So a distribution for $0<\alpha<\infty$ interpolates between the power-law distribution and the exponential distribution. When the parameter α continuously changes from 0 to ∞, the distribution continuously varies from a power-law distribution to an exponential distribution. Actually, when α takes a value larger than 100, the SPL shows a rather good linear line on a single-logarithmic plane indicating approximately an exponential distribution; and it shows a rather good linear line on a double-logarithmic plane, which indicates approximately a power-law distribution, when α takes a value smaller than 1. Therefore, typical SPL functions can be shown only with α values between 1 and 100. As can be seen in Figs. 2–4, all the three distributions: act degree distribution, node strength distribution and degree distribution, show typical SPL functions. It is well known that in the accumulative counterpoint of a power-law distribution the scaling parameter changes from

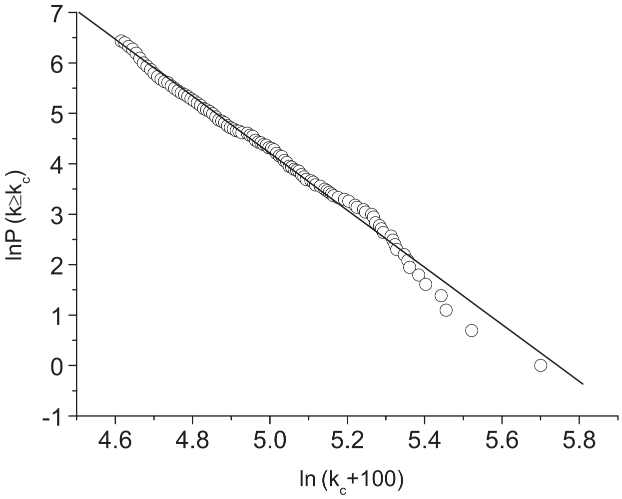

Fig. 3. The empirical results on the accumulative degree distribution of TCHPFN.

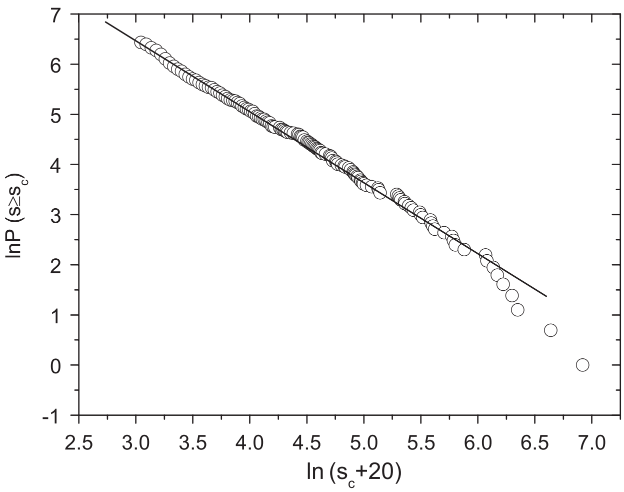

Fig. 4. The empirical results on the accumulative node strength distribution of TCHPFN. The solid line denotes the least square fittings of the data.

η to $\eta + 1$. The slope of a distribution data fitting line on a single-logarithmic plane, which shows that the distribution is exponential, remains unchanged in its accumulative counterpoint. It seems easy to deduce the conclusions that in the accumulative counterpoint of a SPL distribution α remains unchanged and η changes to $\eta + 1$. We use accumulative distributions in Figs. 2–4 since this can greatly decrease the statistical fluctuations. The values of the key parameters, η and α, of the three accumulative distributions are listed in Table 2.

2.2. Bus route networks of five Chinese cities

The study of traffic networks is always of great interest. As aforementioned we are interested only in Space P where the transportation route (scheduled flight, coach number, bus route, etc.) can be considered as an act, a station as a node (actor), and the edge between two nodes expresses their collaboration in a common route. In this way, a route (including all its stations and the relationship edges) forms an ACSG. All such ACSGs joined together by the common nodes form, for the example, the whole bus route network (BRN). In Ref. [9], our cooperators and we presented empirical investigations on BRNs of Beijing and Yangzhou. For BRN in

Beijing, only 65 urban circulating bus routes and 460 bus stops (in 2003) were considered. The data of the BRN in Yangzhou in 2003 (including the bus routes by way of suburbs) only include 26 bus routes and 269 bus stops. Now we present investigation results on BRNs in four large Chinese cities, Beijing, Shanghai, Nanjing and Hangzhou, in 2006 including the bus routes by way of suburbs. We also present the BRN in Yangzhou in 2005 for a comparison. The data were downloaded from Ref. [31]. In Table 1 we present network size data of the BRN systems.

The accumulative act-size distributions of BRNs in four Chinese cities shown in Fig. 5 indicate the similar unimodal act-size distribution functions as can be seen in Fig. 1. Beijing, the capital and the city with the

Table 1
The network size data of the BRN systems

	Yangzhou	Hangzhou	Nanjing	Beijing	Shanghai
A	148.00	683.00	975.82	16,807.80	6340.50
I	0.48	1.29	2.66	14.23	10.47
M	36	150	252	572	968
N	352	827	1764	4199	4374

These data can be downloaded from Ref. [31]. In the table, A represents city area (km^2); I population (million); M total number of bus routes; N total number of bus stops.

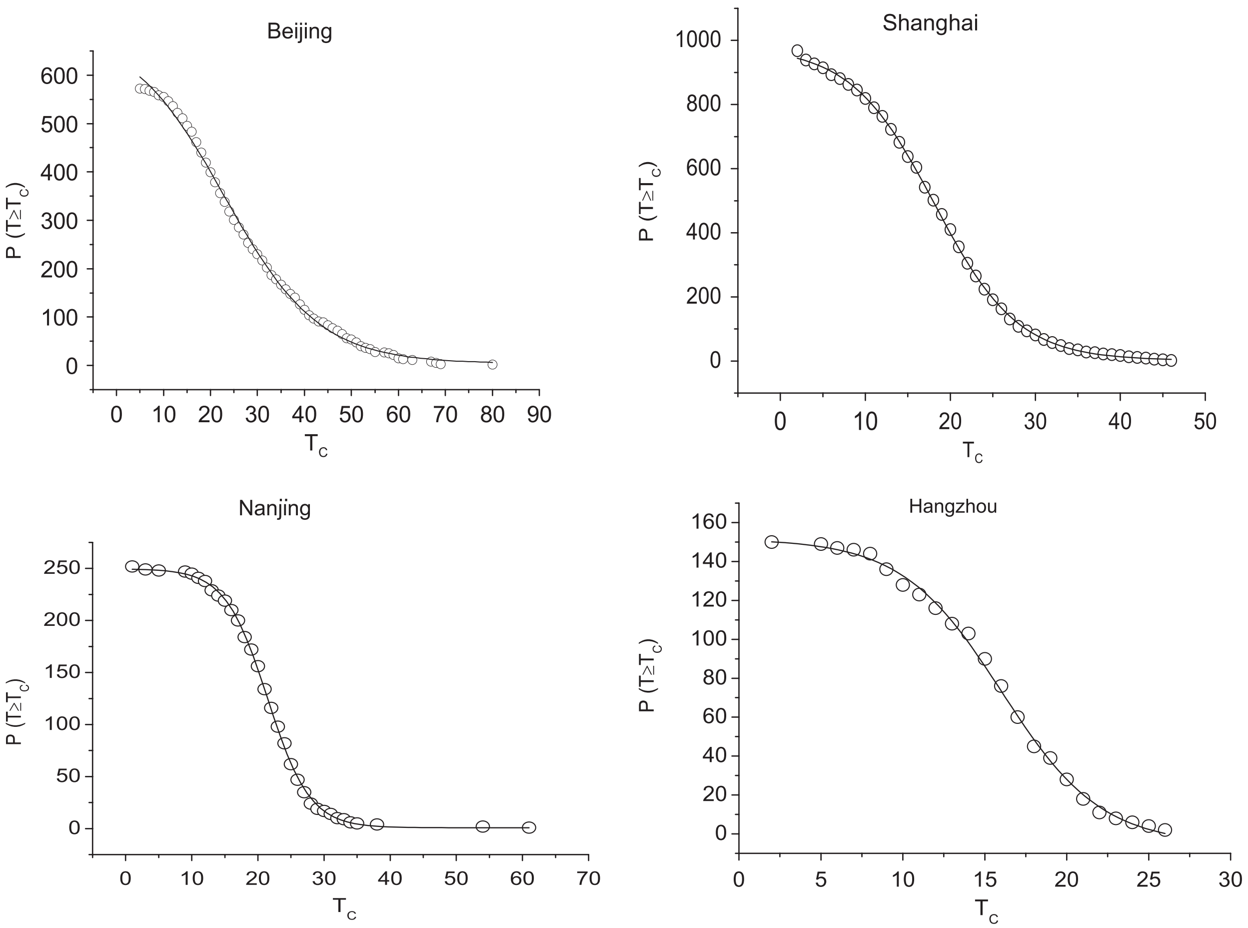

Fig. 5. The empirical results on the accumulative act-size distributions of BRNs in four Chinese cities.

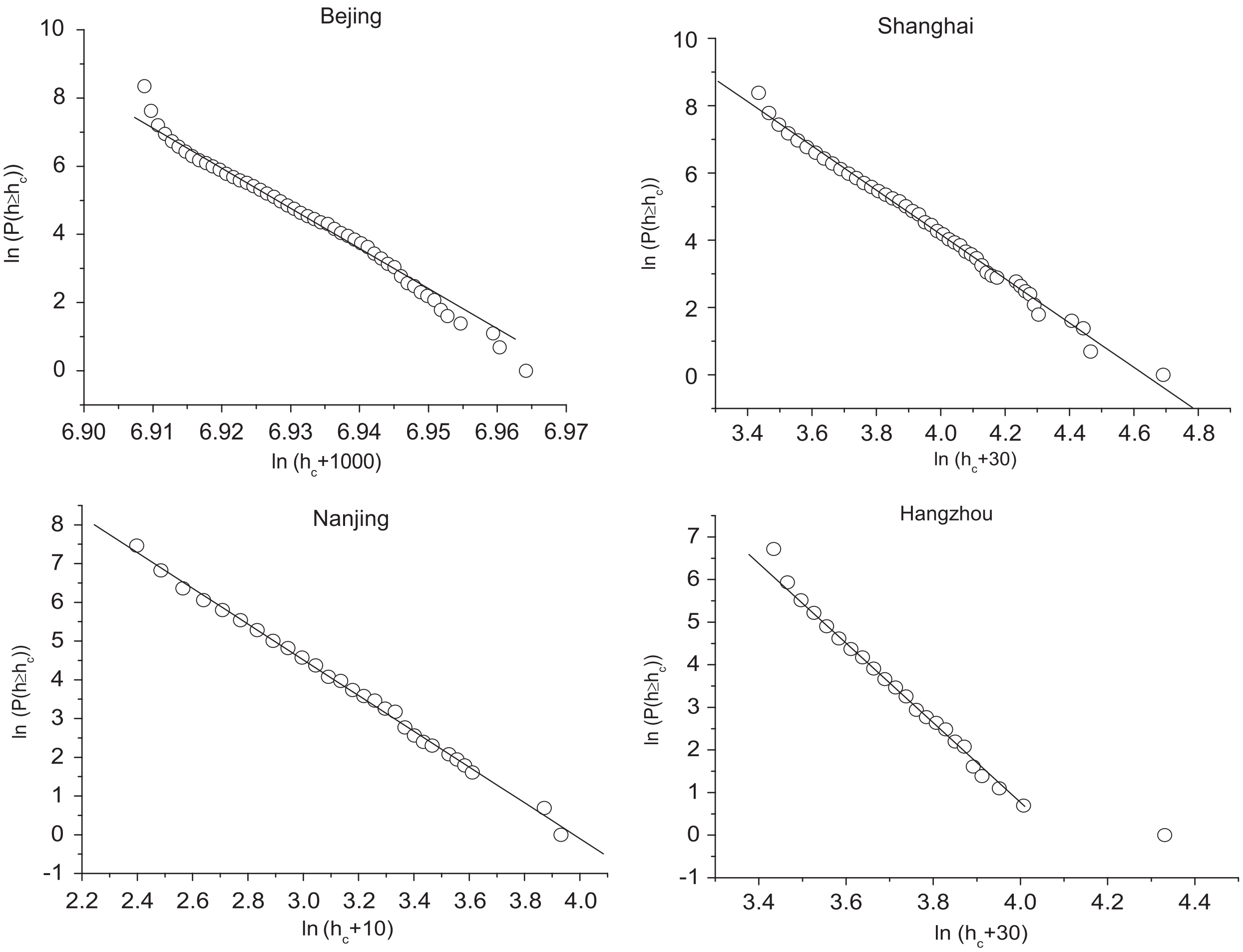

Fig. 6. The empirical results on the accumulative act-degree distributions.

largest values of population and city area of China, shows an approximate exponential act degree distribution of the BRN in Fig. 6; the other three city BRNs show typical SPL functions. All the BRNs show approximate exponential degree distributions and node strength distributions as can be seen in Figs. 7, 8. Each of the BRNs shows that degree distribution and node strength distribution function forms are exactly the same. The values of the key parameters, η and α, of the three distributions are also listed in Table 2.

Some previous empirical investigations on BRN systems reported exponential distributions of degree and act degree in Space P [12,15,16]. We emphasize that, in Figs. 2–4, 6–8, we can put the data on single-logarithmic planes to get tolerably good linear fittings. Of course, the fitting is increasingly being better when the key parameter, α, in the SPL distributions becomes larger. To show this conclusion, we draw the distributions of act degree and node strength of Yangzhou 2005 BRN on single-logarithmic planes as can be seen in Fig. 9(b) and (d). The key parameter, α, can be regarded as infinitely large values since the fittings of data to the linear lines are nice. The act-size distribution shows a unimodal function as can be seen in Fig. 9(a). The degree distribution is shown in Fig. 9(c), the fitting looks even better on the SPL plane. The values of key parameter, η, of the three distributions shown in Fig. 9(b)–(d) are also listed in Table 2.

3. Empirical investigation on un-weighted and weighted assortativity

In addition to the aforementioned η and α values, Table 2 lists the empirical investigation results of un-weighted and weighted assortativity (r and r^w) in some real world networks. All the systems can be regarded as

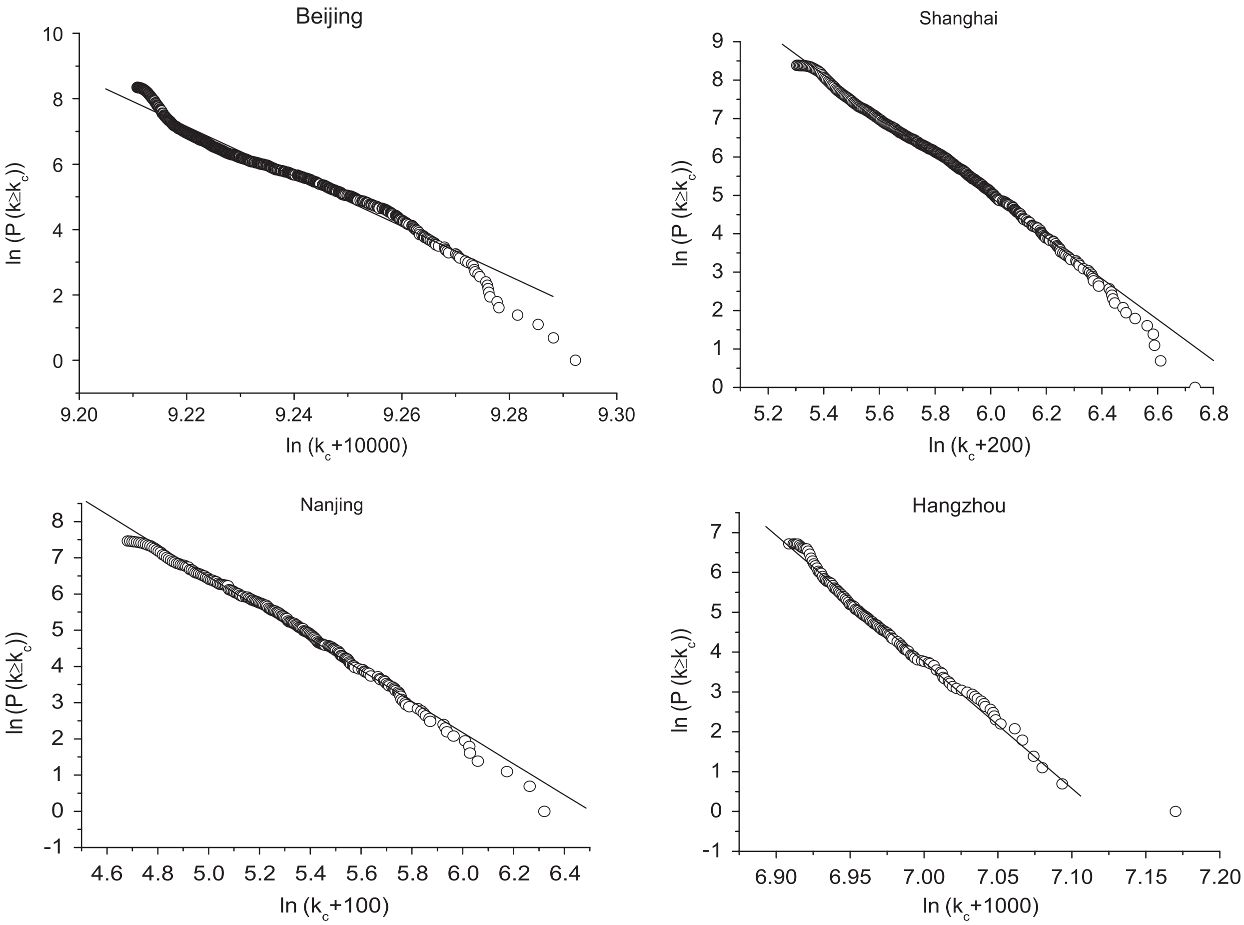

Fig. 7. The empirical results on the accumulative degree distributions.

collaboration networks. The systems include the above-discussed TCHPFN and the BRNs in five Chinese cities, as well as three systems discussed in Ref. [9]: the TRNC, HYRCCF and the CNHA. In TRNC the scenic spots are defined as nodes, and the collaboration between two scenic spots in a travel route as an edge. 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 speciality and also shares the profit. We choose 240 routes; there are total 171 nodes and 719 edges in TRNC network. Huai-Yang denotes two geographical regions located in middle- eastern part of China. Each recipe of Chinese cooked food in the regions is quite famous in China, which can be modeled as an ACSG. In the recipe some kinds of foods, which can be defined as nodes, play their special role in their cooperation to form a delicious dish. We choose 329 recipes of Huai-Yang system; there are total 242 nodes and 1713 edges in the network. CNHA has served as the best example for collaboration networks. The movie actors are defined as nodes, and movies can be regarded as ACSGs. In this study CNHA contains 210,448 nodes (film actors) and 80,000 films (acts). The un-weighted and weighted assortativity coefficients of the three systems, TRNC, HYRCCF and CNHA were not discussed in Ref. [9].

In Table 2 there is a question mark that means unknown value of the weighted assortativity of HYRCCF due to a problem in the data structure.

The results shown in Table 2 strongly recommend that the un-weighted assortativity can be either positive or negative in collaboration networks; but the weighted assortativity is certainly positive although the conclusion (that should be treated as a guess) is obtained only by nine examples. The conclusion may become more convictive when we show the same conclusion, in the next section, by the model presented in Ref. [9].

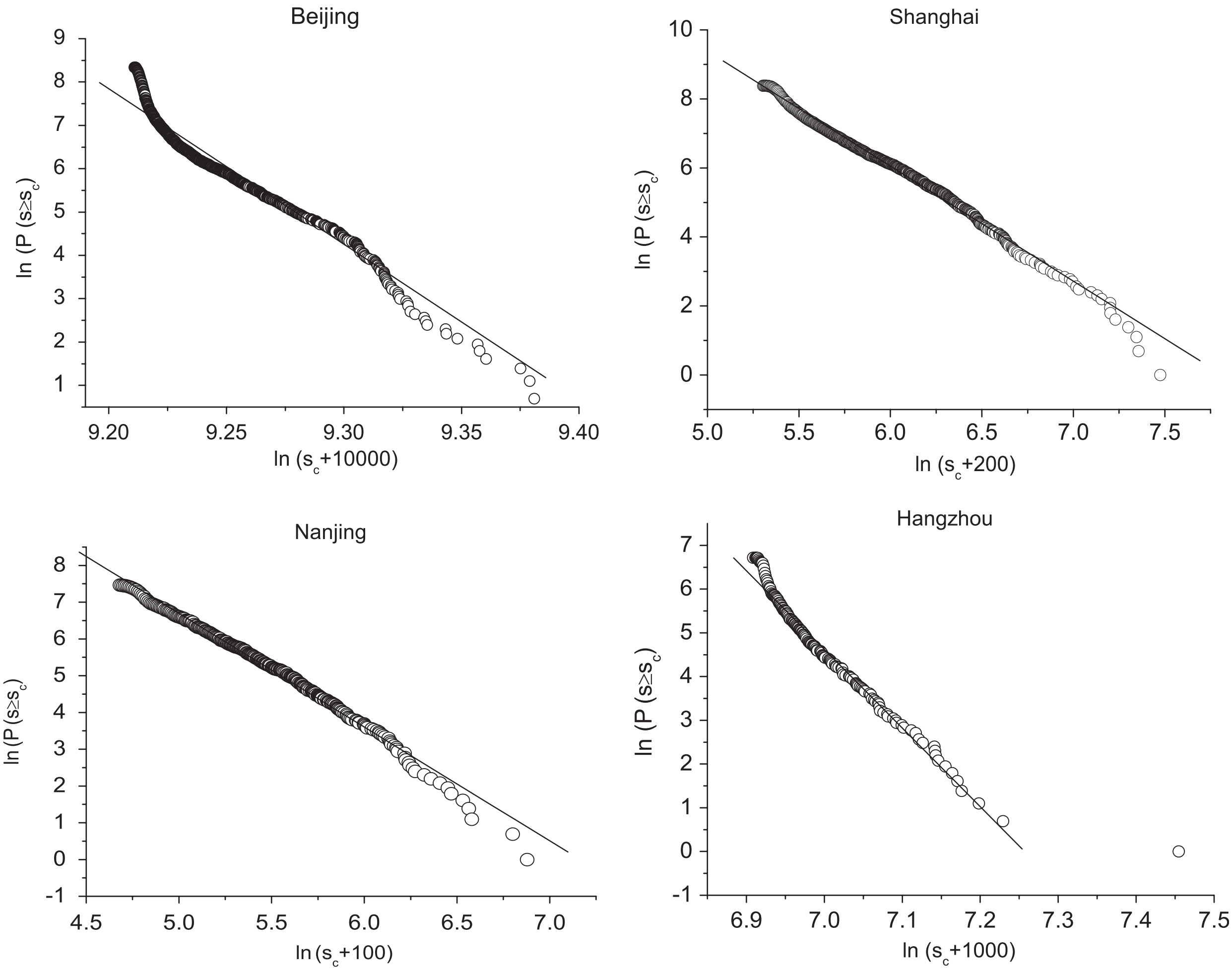

Fig. 8. The empirical results on the accumulative node strength distributions. The solid lines denote the least square fittings of the data.

Table 2
Empirical results on nine real world collaboration networks

System	M	N	r	r^w	$\alpha - s$	$\eta - s$	$\alpha - h$	$\eta - h$	$\alpha - k$	$\eta - k$
BY	36	352	−0.11	0.04	Infinite	0.015	Infinite	0.38	100	5.0
BH	150	827	0.019	0.114	1000	18.5	30	9.9	1000	30.3
BN	252	1764	0.047	0.160	100	3.1	10	4.6	100	4.3
BB	572	4199	0.034	0.171	10,000	37.8	1000	127.4	10,000	77.9
BS	968	4374	0.05	0.202	200	3.3	30	6.6	200	5.3
TC	1536	681	−0.12	0.17	20	1.5	6	1.7	100	5.7
TR	240	171	0.15	0.36	20	2.2	10	3.4	50	6.6
HA	80,000	210,448	0.2653	0.00046	40	2.1	30	5.4	0	2.3
HY	329	242	−0.296	?	10	1.2	5	1.6	20	2.4

In the table the un-weighted and weighted assortativity coefficients are denoted by r and r^w, respectively. Parameters, η and α, of the accumulated act degree distribution, degree distribution and node strength distribution are denoted by $\eta - h$, $\eta - k$, $\eta - s$, $\alpha - h$, $\alpha - k$, $\alpha - s$, respectively. M denotes the total number of acts; N total number of nodes. The system BRN Yangzhou 2005 is denoted by BY; BRN Hangzhou 2006 by BH; BRN Nanjing 2006 by BN; BRN Beijing 2006 by BB; BRN Shanghai 2006 by BS; TCHPFN by TC; TRNC by TR; CNHA by HA; and HYRCCF by HY.

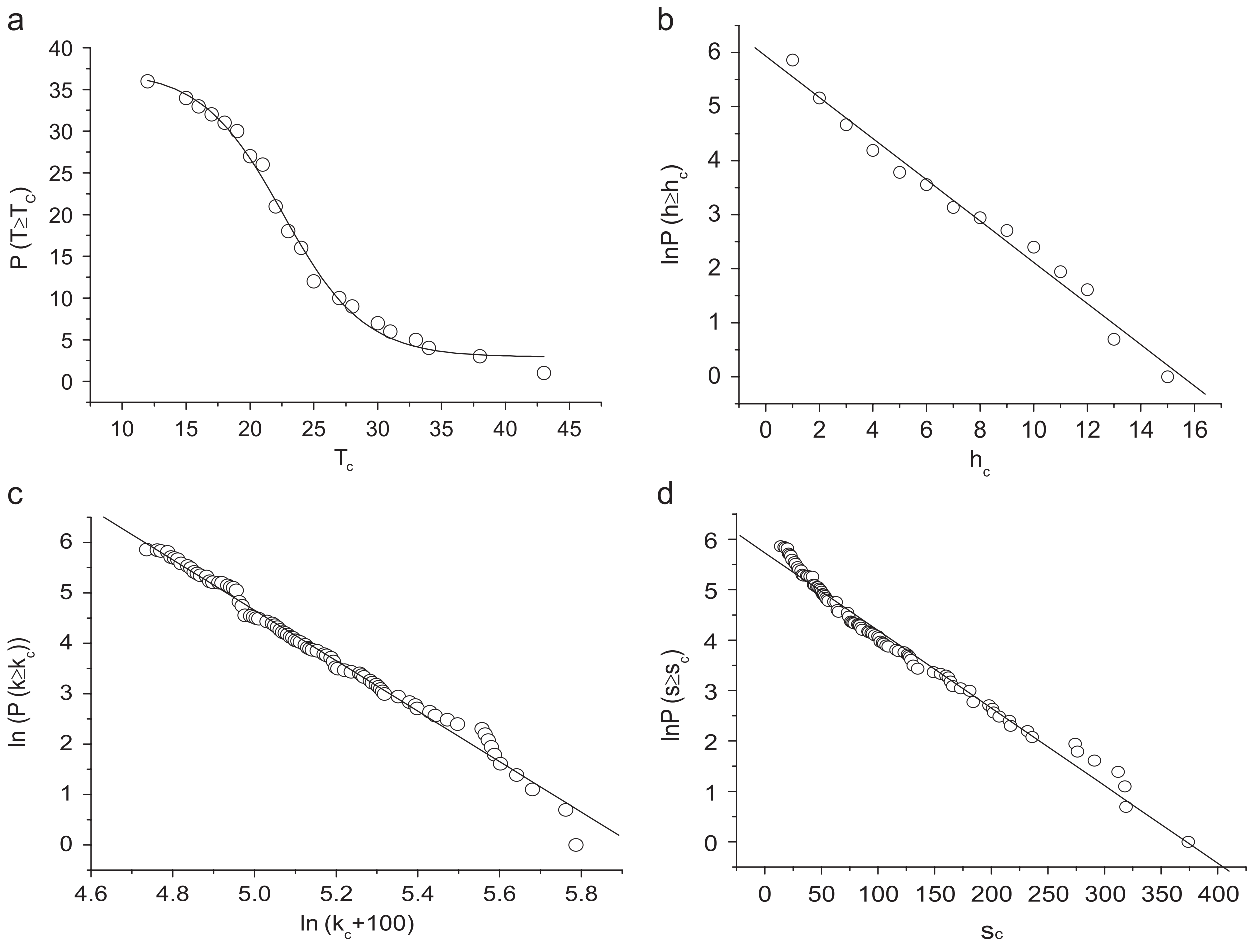

Fig. 9. The empirical results on: (a) accumulative act-size distribution; (b) accumulative act-degree distribution; (c) accumulative degree distribution; and (d) accumulative node strength distribution of Yangzhou BRN in 2005.

In the last section we shall show that the model description should be reasonable at least for a type of collaboration networks, therefore this certainly promote the reliability of the conclusion.

The results about the city BTNs may recommend another conclusion. The weighted assortativity coefficients show clear tendency that they are increasingly being larger when the city BRN becomes bigger; however, un-weighted assortativity coefficients sometimes ruin the rule. It was suggested, after an empirical investigation on 22 Polish city BRNs [12], that small BRNs (node number are smaller than 500) showed negative un-weighted assortativity values, while big BRNs (node number are larger than 500) showed positive un-weighted assortativity values. The results in Table 2 indicate that four large Chinese city BTNs, where the node numbers are larger than 500, show positive un-weighted assortativity values. The small BTNs in Yangzhou (node number are smaller than 500) show negative un-weighted assortativity values. The results are in agreement with the conclusion reported in Ref. [12].

It seems not so easy to make conclusions by the values of η and α listed in Table 2. We shall discuss about them in the last section.

4. A collaboration network evolution model

We have introduced a model [9] for a description of the evolution of collaboration networks. The general situation, where network growth partially prefers random selection and partially prefers linear preferential principle, was considered; however, the model for this case was not analytically solved. We shall analytically solve it in this section. The model can be briefly stated as follows. In the model there are m_0 nodes at $t = 0$, which are connected and form some ACSGs. In each time step a new node is added. It is connected to $T - 1$ old nodes (as a simplification, T is a constant integer) to form a new ACSG of T nodes. As shown in Sections 2 and 3 (and also in Ref. [9]), all the investigated real world collaboration networks showed asymmetric, unimodal distribution of T. A unimodal distribution of a quantity, such as age, salary or stature, means that most values of the quantity are positioned around its averaged value. Therefore one can expect that, qualitatively, it does not influence the simplified model conclusions if one assumes T is a constant instead of considering T as a distribution. The rule of selecting the $T - 1$ old nodes is as follows: randomly selecting old nodes with a probability p, and using a linear preference rule with probability $1 - p$ [32,33]. With the linear preference rule an old node i is selected with the probability $\Pi \propto h_i/\sum_j h_j$, where h_i denotes its act degree, and j denotes another old node. This means that to build a new ACSG, by probability $1 - p$ one selects each possible old node according to how many ACSGs it has taken part in; while by probability p one selects nodes according to many special uncorrelated considerations (practically just like randomly). Referring the famous model proposed in Ref. [6], we have

$$\frac{\partial h_i}{\partial t} = p(T-1)\frac{1}{m_0 + t} + (1-p)\frac{(T-1)h_i}{T(m_0+t)}. \tag{8}$$

This equation can be written as

$$\frac{\partial h_i}{\partial \ln(m_0 + t)} = \frac{(1-p)(T-1)}{T}\left[h_i + \frac{TP}{1-p}\right]. \tag{9}$$

This can be solved to give

$$h_i = C_i(t+m_0)^{(T-1)(1-p)/T} - \frac{Tp}{1-p}, \tag{10}$$

where C_i is the integration constant, which can be determined using the condition $h_i(t = t_i) = 1$. t_i is the time when the node i is attached to the network. This gives rise to

$$C_i = \frac{1 + Tp/(1-p)}{(t_i + m_0)^{(T-1)(1-p)/p}}, \tag{11}$$

Thus we have

$$h_i = \left(1 + \frac{Tp}{1-p}\right)\left(\frac{t+m_0}{t_i+m_0}\right)^{(T-1)(1-p)/p} - \frac{Tp}{1-p} \tag{12}$$

or

$$t_i = (t+m_0)\left(\frac{h_i+\alpha}{1+\alpha}\right)^{-\beta} - m_0, \tag{13}$$

where $\alpha = Tp/(1-p)$ and $\beta = T/[(T-1)(1-p)]$. Now we have

$$P(h_i < h) = P(t_i > t_i(h)) = (t_i - t_i(h))/(t+m_0), \tag{14}$$

where

$$t_i(h) = (t+m_0)\left(\frac{h+\alpha}{1+\alpha}\right)^{-\beta} - m_0, \tag{15}$$

Thus

$$P(h_i < h) = 1 - \left(\frac{h+\alpha}{1+\alpha}\right)^{-\beta}. \tag{16}$$

The act degree distribution is then given by

$$P(h) = \frac{\mathrm{d}P(h_i < h)}{\mathrm{d}h} = \frac{\beta}{1+\alpha}\left(\frac{h+\alpha}{1+\alpha}\right)^{-\eta}. \tag{17}$$

This is the SPL function where $\eta = \beta + 1$.

Similarly, we can obtain (considering that $s_i = h_i(T-1)$ when T is a constant)

$$\frac{\partial s_i}{\partial t} = p\frac{(T-1)^2}{t} + (1-p)\frac{(T-1)s_i}{Tt}. \tag{18}$$

This equation can be written as

$$\frac{\partial s_i}{\partial \ln t} = p(T-1)^2 + (1-p)\frac{(T-1)s_i}{T}. \tag{19}$$

This can be solved to give

$$s_i = C_i t^{(T-1)(1-p)/T} - \frac{T(T-1)p}{1-p}. \tag{20}$$

Using the condition $s_i(t = t_i) = T - 1$, we have

$$C_i = \frac{T - 1 + T(T-1)p/(1-p)}{t_i^{(T-1)(1-p)/T}}. \tag{21}$$

Thus we have

$$s_i = [(T-1)(1 + Tp/(1-p))]\left(\frac{t}{t_i}\right)^{(1-p)(T-1)/T} - \frac{T(T-1)p}{1-p}. \tag{22}$$

Let $\alpha = T(T-1)p/(1-p)$ and $\beta = T/[(T-1)(1-p)]$. Now we have

$$P(s_i < s) = P\left(t_i > t\left(\frac{s+\alpha}{T-1+\alpha}\right)^{-\beta}\right) = 1 - \left(\frac{s+\alpha}{T-1+\alpha}\right)^{-\beta}. \tag{23}$$

We obtain

$$P(s) = \frac{\mathrm{d}P(s_i < s)}{\mathrm{d}s} = \frac{\beta}{T-1+\alpha}\left(\frac{s+\alpha}{T-1+\alpha}\right)^{-\eta}, \tag{24}$$

where $\eta = \beta + 1$. The conclusion is that, in general cases, the act degree distribution $P(h)$ and node strength distribution $P(s)$ show SPL function form with the same η value, but the key parameters, α, are different. To check whether the real data follows such a SPL distribution, we need to tune the parameter α and check whether, for example, $\ln P(h)$ vs. $\ln(h+\alpha)$ is a straight line.

In all the practical systems we investigated, the accumulative distribution of act-size, $P(T \geqslant T_C)$, could be well fitted by a shifted Poisson distribution, $P(T) = (\lambda^{T+b}/A(T+b)!)\mathrm{e}^{-\lambda}$. If considering such a distribution of act-size, the numerical simulation by the model shows that the analytic conclusions about act-degree

Table 3
The weighted and un-weighted assortativity coefficients of the model

p	r^w	r
0	0.0008 ± 0.003	-0.27 ± 0.0019
0.2	0.050 ± 0.003	-0.10 ± 0.0019
0.4	0.13 ± 0.004	-0.014 ± 0.0027
0.5	0.18 ± 0.005	0.060 ± 0.003
0.6	0.20 ± 0.005	0.10 ± 0.0035
0.8	0.28 ± 0.005	0.22 ± 0.004
1.0	0.31 ± 0.0045	0.29 ± 0.004

distribution and node strength distribution functions still qualitatively remain unchanged. So, it is basically correct to consider T as a constant. This is in agreement with the empirical investigation reported in Sections 2 and 3 that act-degree distribution and node strength distribution show SPL function.

We numerically calculated the weighted and un-weighted assortativity coefficients by the model. Table 3 shows some of the results. In the numerical simulation, the network grows until 5000 nodes (our simulation with more nodes shows similar results). Table 3 leads to the following conclusions: (1) the model always shows positive weighted assortativity; (2) the model may show either positive or negative un-weighted assortativity; and (3) the assortativity value is increasingly being larger as the random proportion, p, increases. The conclusions (1) and (2) are in agreement with the empirical investigation results. By conclusion (3) we may understand the complicated empirical investigation results listed in Table 2 better. We shall discuss this in the next section.

5. Conclusion and discussion

Empirical investigation results of nine real world collaboration networks have been presented. The systems include five transportation networks, three technical networks (networks of prescriptions, cooked food recipes, travel routes) and a social network (movie actor collaboration). The networks have common topological characteristics. There are two types of basic elements. One of them can be addressed as acts, the other called actors. In the projected actor unipartite graphs the acts are expressed by so-called act-complete subgraphs in which each pair of nodes are linked by an edge denoting their cooperation in the act.

The investigations propose some conclusions. Firstly, act degree distribution, degree distribution and node strength distribution usually show SPL function forms, which can continuously vary from an ideal power law to an ideal exponential decay. A parameter, α, can be used for description of the position between these two extreme cases. When $\alpha = 0$, SPL becomes ideal power law. When $\alpha \to \infty$, SPL tends to an ideal exponential decay. Another parameter, η, shows the scaling exponent when the SPL becomes a power law. It describes how fast the function decays.

Secondly, the weighted assortativity coefficients show positive values in all the networks; while the un-weighted assortativity coefficients may show either positive or negative values.

The third conclusion may be obtained from the investigation results of the five transportation networks. The data show a monotonic dependence of the weighted assortativity values on the network size.

These conclusions are important; however, it seems hard to derive, directly from the empirical data, the relationship between the parameters of SPL functions, α and η, and the assortativity coefficients. The relationship should be very interesting for a lot of scientists.

A simplified model is presented, by which the first conclusion can be analytically derived. The basic idea of the model is that collaboration networks evolve via a process in which acts are organized gradually. The evolution dynamics of the collaboration networks are divided into two tendencies: random selection and linear preferential principle. By this idea the quasi-continuous evolution equation can be solved analytically that leads to SPL distributions of act degree and node strength in general situations where nodes are selected partially randomly, with a probability p, and partially by linear preferential principle, with the probability $1-p$. This analytic treatment gives an explicit expression on the relationship between the random selection proportion p and the SPL function parameters α and η. When $p = 0$, which means a complete dominance of linear preferential principle, one has $\alpha = 0$, then SPL becomes ideal power law. When $p = 1$, which indicates a complete dominance of random selection, one finds $\alpha \to \infty$, SPL tends to an ideal exponential decay.

We cannot get explicit expressions of assortativity coefficients analytically by the model, but it is easy to compute the coefficients numerically. The results confirm the second empirical conclusion, and show a clear monotonic dependence of both the weighted and un-weighted assortativity on the random selection proportion p. The conclusion then is: assortativity coefficient, r or r^w, and SPL function parameters, α and η, are monotonously dependent on the random selection proportion p. This gives rise to the possibility for knowing how the parameters of SPL functions depend on the assortativity coefficients.

Due to the fact that α, η, r and r^w closely relate to the mode (or average) act size (the mode act size denotes the number of nodes, which an act most probably contains), we list the mode act size, T_{max}, of all the 9 real world networks in Table 4.

Table 4
The mode act size, T_{max}, of all the nine real world networks

	BY	BH	BN	BB	BS	TC	TR	HA	HY
T_{max}	14	16	21	20	15	3	2	4	4

The systems are denoted by the same characters as being used in Table 2.

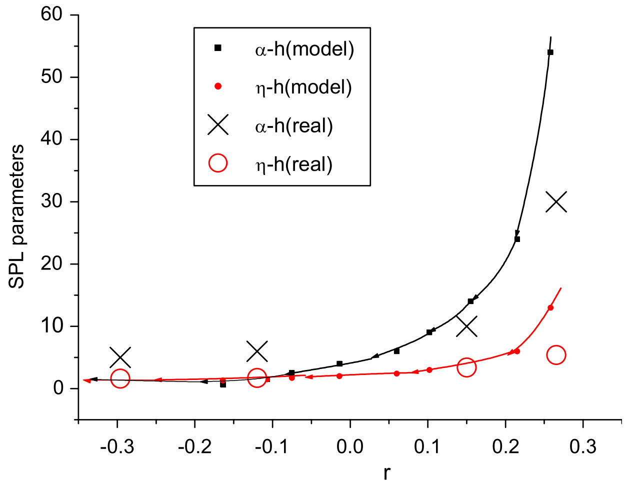

Fig. 10. (Color online) The upper solid curve, which is black online, shows the model relationship between the un-weighted assortativity, r, and the SPL parameter $\alpha - h$ of the accumulated act-degree distributions. The lower solid curve, which is red online, shows the model relationship between r and $\eta - h$. The four forks (black online) represent data of $\alpha - h$ versus r obtained from the four real world networks with small T_{max}. From the left to the right, the forks denote the data obtained from HYRCCF, TCHPFN, TRNC and CNHA in turn. The four void circles represent data of $\eta - h$ versus r obtained from the same four systems.

According to the table, the networks can be divided into two groups. The first group contains networks with relatively larger T_{max} (the BRN networks). The second group contains networks with much smaller T_{max}. For the group with small T_{max}, the data show that assortativity coefficient, r or r^w, and SPL function parameters, α and η, are monotonously dependent on each other. In order to compare the empirical data with the model results quantitatively, we draw, as an example, the un-weighted assortativity, r, and α and η values of the $P(h)$ distributions (denoted by $\alpha - h$ and $\eta - h$, respectively) in Fig. 10.

The Fig. 10 shows a good agreement between model predictions and the empirical results although we can provide only four empirical data. This probably means that the model can be used for an explanation on the relationship between two topological properties, assortativity and SPL parameters, for the real networks with small T_{max}. Also, the evolution mechanism of the collaboration networks expressed in the model may be basically acceptable. The conclusion is that in such networks assortativity coefficients and SPL function parameters are monotonously dependent on the random selection proportion p, which is an important parameter but very hard to be empirically measured. After knowing the empirical and model simulation results we can conclude that, in the evolution processes of the four systems, HYRCCF shows the smallest random selection proportion (that means the largest number of network hub existence); TCHPFN runs the second; the next one is TRNC; CNHA has the largest random selection proportion (that means the smallest number of network hub existence).

Of course we can draw similar figures to show the relationship between r (or r^w) and the SPL parameters, α and η, for the model (with a much larger T) and the five real world networks with larger T_{max}. However, the figures show relatively worse agreement between model simulations and empirical data although the basic tendencies still remain consistent. This may indicate that, for the description of the collaboration networks with relatively larger T_{max}, especially the transportation collaboration networks, the model still needs revision.

Acknowledgments

The research is supported by the Chinese National Natural Science Foundation under the Grant nos. 10635040 and 70371071. We are deeply grateful for the very valuable help and cooperation of Prof. Kan Chen in both the research ideas and analytical calculations. Thanks should be given to Dr. Y-Z. Chen, Miss Y. He, P.-P. Zhang, X.-F. Han, J.-Y. Tang, R. Qiu and Q.-J. Chen for their very important contributions in searching and processing the empirical data.

References

[1] R. Albert, A.-L. Barabasi, Rev. Mod. Phys. 74 (2002) 47.
[2] M.E.J. Newman, SIAM Rev. 45 (2003) 167.
[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Phys. Rep. 424 (2006) 175.
[4] S. Wasserman, K. Faust, Social Network Analysis: Methods and Applications, Cambridge University Press, Cambridge, 1994.
[5] D.J. Watts, S.H. Strogatz, Nature 393 (1998) 440.
[6] A.-L. Barabasi, R. Albert, Science 286 (1999) 509.
[7] M.E.J. Newman, Phys. Rev. E 64 (2001) 016131, 016132.
[8] J.J. Ramasco, S.N. Dorogovtsev, R. Pastot-Satorras, Phys. Rev. E 70 (2004) 036106.
[9] P. Zhang, K. Chen, et al., Physica A 360 (2006) 599.
[10] 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.
[11] M.E.J. Newman, Phys. Rev. Lett. 89 (2002) 208701;
M.E.J. Newman, Phys. Rev. E 67 (2003) 026126.
[12] J. Sienkiewicz, J.A. Holyst, Phys. Rev. E 72 (2005) 046127.
[13] M. Kurant, P. Thiran, Phys. Rev. E 74 (2006) 036114;
M. Kurant, P. Thiran, Phys. Rev. Lett. 96 (2006) 138701.
[14] C.V. Ferber, T. Holovatch, Yu. Holovatch, V. Palchykov. arXiv: physics/0608125.
[15] X. Xu, J. Hu, F. Liu, L. Liu, Physica A 374 (2007) 441.
[16] Y.-Z. Chen, N. Li, D.-R. He, Physica A 376 (2007) 747.
[17] B.-B. Su, H. Chang, Y.-Z. Chen, D.-R. He, Physica A 379 (2007) 291.
[18] W. Li, X. Cai, Phy. Rev. E. 69 (2004) 046106.
[19] M.E.J. Newman, Phys. Rev. E 70 (2004) 056131.
[20] Z. Pan, X. Li, X. Wang, Phys. Rev. E 73 (2006) 056109.
[21] M. Barthélemy, A. Barrat, R. Pastor-Satorras, A. Vespignani, Physica A 346 (2005) 34.
[22] M. Li, Y. Fan, J. Chen, L. Gao, Z. Di, J. Wu, Physica A 350 (2005) 643.
[23] W.-X. Wang, B.-H. Wang, B. Hu, G. Yan, Q. Ou, Phys. Rev. Lett. 94 (2005) 188702.
[24] A. De Montis, M Barthélemy, A. Chessa, A. Vespignani, arXiv: physics/0507106.
[25] C.C. Leung, H.F. Chau, Physica A 378 (2007) 591.
[26] W.-X. Wang, B. Hu, B.-H. Wang, G. Yan, Phys. Rev. E 73 (2006) 016133.
[27] W.-X. Wang, B. Hu, T. Zhou, B.-H. Wang, Y.-B. Xie, Phys. Rev. E 72 (2005) 046140.
[28] J. Laherrere, D. Sornette, Eur. Phys. J. B 2 (1998) 525.
[29] Y.-X. Zhu, Chinese Herb Prescription Manual, second ed., Jindun Publishing House, Beijing, 1996 (in Chinese).
[30] D.-L. Liu, et al., Most Frequently Used Chinese Medication Handbook, People's Military Medication Publisher, Beijing, 1996 (in Chinese).
[31] Hangzhou: ⟨http://www.hzbus.com.cn/⟩; Nanjing: ⟨http://www.njkgc.cn/⟩; Beijing: ⟨http://www.bjbus.com/⟩; Shanghai: ⟨http://www.jt.sh.cn/⟩; Yangzhou: ⟨http://www.yzren.com/life-busstation.asp⟩.
[32] Z. Liu, Y.-C. Lai, et al., Phys. Lett. A 303 (2002) 337.
[33] X. Li, G. Chen, Physica A 328 (2003) 274.