Available online at www.sciencedirect.com
SCIENCE DIRECT®
Information Processing and Management 42 (2006) 1599–1613
INFORMATION PROCESSING & MANAGEMENT
www.elsevier.com/locate/infoproman

Connection and stratification in research collaboration: An analysis of the COLLNET network

Yin Li-chun [a,b,*], Hildrun Kretschmer [a,c], Robert A. Hanneman [d], Liu Ze-yuan [a,b]

[a] *WISE LAB, Dalian University of Technology, Dalian 116023, China*
[b] *Institute of Science Studies and Management of Science and Technology, Dalian University of Technology, Dalian 116023, China*
[c] *Department of Library and Information Science, Humboldt-University Berlin, Dorotheenstrasse 26, D-10099 Berlin, Germany*
[d] *Department of Sociology, University of California – Riverside, USA*

Received 16 March 2006; accepted 16 March 2006

Abstract

Co-authorship among scientists represents a prototype of a social network. By mapping the graph containing all relevant publications of members in an international collaboration network: COLLNET, we infer the structural mechanisms that govern the topology of this social system. The structure of the network affects the information available to individuals, and their opportunities to collaborate. The structure of the network also affects the overall flow of information, and the nature of the scientific community. We present a number of measures of both the macro- (whole-network) and micro- (actor-centered) structure of collaboration, and apply these to COLLNET. We find that this scientific community displays many aspects of a "small-world," and is somewhat vulnerable to disruption should major figures become inactive. We also find inequality in the roles played by individuals in the network. The inequalities, however, do not create a closed and isolated "core" or elite.

Keywords: Social network analysis; Cooperation network; Topological structure; Scale-free

1. Introduction and motivation

In most areas of academic science, collaboration in research and publication is very common. A good deal of evidence suggests that cooperation among researchers is increasing in a wide range of fields from mathematics to neuroscience (http://www.oakland.edu/enp/trivia.html; Braun, Glanzel, & Schubert, 2001). Collaboration may be seen as a process in which knowledge flows among scientists (Calero, van Leeuven, & Tijssen, 2005), and individual scientists gain access to new "capital." As networks of collaboration increase in size, scientists may be gaining access to information both directly (from the individuals with whom they collaborate) and

* Corresponding author. Address: WISE LAB, Dalian University of Technology, Dalian 116023, China. Fax: +86 411 84706082.
E-mail addresses: yinliuxing1290@yahoo.com.cn (L.-c. Yin), kretschmer.h@onlinehome.de (H. Kretschmer), robert.hanneman@ucr.edu (R.A. Hanneman), liuzy@vip.163.com (Z.-y. Liu).

0306-4573/$ - see front matter
doi:10.1016/j.ipm.2006.03.021

indirectly (through the collaborators of their collaborators). The structure of the larger network may affect the work done by an individual scientist in ways that are not apparent to them. The structure of the whole network of collaboration may also affect scientific productivity. Some network structures may promote diverse and creative work; other network structures may create separation and retard creativity. Otte and Rousseau (2002) showed that social network analysis (SNA) can be used successfully in the information sciences, as well as in studies of collaboration in science.

In this paper we will examine the structure of a network of collaboration among a number of researchers from diverse fields of study who shares an interest in the problem of scientific collaboration (the COLLNET network). We will examine this network as a "social network." Social networks are the people or groups connected by some relations. The people or groups are called "nodes" (also called vertices or actors according to different disciplines), and the relations (or connections) "links" (edges, or ties). In this article, we'll use the terms node, actor, author, and vertex interchangeably. The terms link, tie, edge, and co-authorship are also used interchangeably. We will examine the social relation of collaboration among the researchers by examining their "affiliation" network, in which actors are "connected" by co-authorship of publications.

Many co-authorship networks have been studied (Batagelj & Mrvar, 2000; Beaver & Rosen, 1978; Egghe, Rousseau, & Van Hooydonk, 2000; Glänzel & Schubert, 2004; Kretschmer, 2004; Kretschmer & Aguillo, 2005; Price, 1963; Wagner, 2005) to investigate the patterns, motivation, and the structure of scientific collaboration. There are some organizations take the co-authorship research as their goal. Besides the immediate interest for scientometrics and informetrics, co-authorship networks are of interest for understanding the topological structures governing the features of large networks, in general (Barabasi et al., 2002; Newman, 2001a, 2001b, 2001c). Especially in recent years, prompted by two parallel developments of emergence of large databases on the topology of various real networks and increased computing power, scientists have used co-authorship networks to understand the processes generating network topology (e.g. Watts & Strogatz, 1998). Network theory has become one of the most visible pieces of the body of knowledge that can be applied to the description, analysis, and understanding of many complex systems, spanning biological and sociological components.

Our interest in the structure of networks of collaboration among researchers focuses on how the topology (or structure) of networks may promote or inhibit creativity and cumulative knowledge in fields of inquiry that cross disciplinary boundaries. We will conduct our inquiry about the structure of the COLLNET network at two levels: a macro- (whole-network) level that focuses on the extent and robustness of connection; and, a micro- (individual-centered) level that focuses on patterns of inequality and stratification that may divide scientists. Micro- approaches provide many measures to evaluate the varying importance of the different scientists in a network according to one criterion or another. These measures have proved of great value in the analysis and understanding of the roles played by scientists in cooperation networks.

2. Theoretical perspectives

A network, also called graph, is a pair $G = (V, E)$ consisting of two sets: a set of nodes $V = \{1, 2, \ldots, N\}$, and a set of lines $E = \{e_1, e_2, \ldots, e_L\}$ between pairs of nodes. If the line between two nodes is non-directional, then the network is called undirected; otherwise, the network is called directed. A network is usually represented by a graph, where nodes are drawn as small points, undirected lines are drawn as edges and directed lines as arcs connecting the corresponding two nodes. Cooperation networks are undirected networks, where nodes represent the scientists and the links represent the cooperation relationship of collaboration on joint publication.

The pattern of cooperation that is represented in a graph can be examined from a macro- (network-centered) or from a micro- (actor-centered) perspective. The macro-structure of a graph informs us about the likely performance of the social structure that arises out of the physics of its connections; the actors embedded in the network may well be completely unaware of this structure. For example, networks in which most actors have connections at short distances to all others are likely to display rapid diffusion – even if the actors embedded in it are no more likely than those in a less dense network to pass along information. The micro-structure of a graph informs us about the differential constraints and opportunities facing individual actors that shape their social behavior. For example, an actor who has many more connections than another may be more influential and have higher social status.

2.1. Macro-level structure

Many aspects of the macro-behavior of networks are thought to follow from their structure. The ability to respond quickly to stimuli, the rate and completeness of diffusion, the ability to identify and construct novel solutions to some new problems, and the institutionalization of cooperation among actors are all affected by the patterns of connections among the actors. There are many ways of characterizing the patterns of connections among nodes in a network. We will focus four key elements of networks: "small-worlds," "scale-free" degree distributions, clustering (or transitivity), and robustness.

2.1.1. Mean geodesic distance ("small-worlds")

"Small-worlds" are the extent to which most pairs of nodes in most networks are connected by short paths. As defined formally by Watts and Strogatz (1998), and informally by Milgram (1967), many social networks may display structures where most individuals are at very few "degrees of distance" from one another. Short distances between actors allow for the rapid diffusion of new ideas and influences among members of the community, and might be regarded as a desirable property of networks of scientific collaboration.

Consider an undirected network, and let us define l to be the mean geodesic (i.e., shortest) distance between nodes pairs in a network:

$$l = \frac{1}{\frac{1}{2}N(N+1)} \sum_{i \geqslant j} d_{ij} \tag{1}$$

where d_{ij} is the geodesic distance from node i to node j; N is the total number of nodes in the connected component.

Where distances are great, it may take a long time for information to diffuse across a population. And, networks with short path distances may also be less subject to disruption, and hence more stable and reliable.

2.1.2. "Scale-free" degree distribution

The desirability of short path lengths would suggest that every actor should be connected directly to every other. In large networks, however, such a pattern would be highly impractical; actors can maintain only so many ties (here, they can only produce some finite number of papers). It is possible to have, on the average, short paths in large networks if some actors are "hubs" who connect groups or clusters. These "hubs" have many more ties than other actors, and function to connect or "broker" connections among nodes that would have otherwise been unable to reach one another. As networks increase in size, there may develop multiple levels of "hubs," and hence an exponential (or "scale-free") distribution of actor degree; a very small number of actors have very many ties, larger numbers of actors have smaller numbers of ties.

The mean nodal degree is the average degree of all nodes in that network.

$$\bar{k} = \frac{\sum_{i=1}^{N} k_i}{N} \tag{2}$$

k_i is the degree of node i, and N the total number of nodes in a network. We define $p(k)$ to be the fraction of nodes in the network that have degree k. Equivalently, $p(k)$ is the probability that a random chosen node in a network has degree k. Rapoport (1957) may be the first theorists who stress the importance of the degree distribution in the networks. A host of real networks (Albert, Hawoong, & Barabasi, 1999; Liljeros, Edling, Amaral, Stanley, & Aberg, 2001; Price, 1965; Redner, 1998) which have been studied show an important characteristic: power-law distribution which is related to their degree distribution.

$$p(k) \sim k^{-\gamma} \tag{3}$$

Networks with power-law distributions are often referred to as "scale-free" networks. The term "scale-free" refers to any functional form $f(x)$ that remains unchanged to within a multiplicative factor under a rescaling of the independent variable. In the function (3), independent variable is k, $p(ak) \sim a^{-\gamma}k^{-\gamma} \sim k^{-\gamma}$ and hence we call it as scale-free network (Newman, 2003).

2.1.3. Clustering (transitivity)

Short distances among actors in a collaboration network provide fast access to diverse resources. However, successful collaboration also depends on sharing deep specialist knowledge. Peer review and critical feedback from others who are intellectually "close" to collaborators is also needed. That is, successful collaboration in science must be embedded in tightly connected and self-conscious intellectual communities – at the same time that the collaborators reach out to new and diverse parts of the scientific community.

In scientific collaboration networks then, we should observe tendency for actors to form local clusters with people with whom they share common interests. In the language of social networks, a colleague of a colleague is likely to be regarded as a colleague. The collegial relation is "transitive." To the extent that the relations among scientists display a tendency towards transitivity, collaboration networks are likely to be characterized by local clusters of individuals who are tied to most of the others. This tendency to cluster is quantified by the "clustering coefficient" (Albert & Barabasi, 2002) which may also be interpreted as the average density of the ego-neighborhoods of actors in a network. The extent which each actor in a network is "embedded" in a local cluster may be indexed as:

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \tag{4}$$

C_i is the ratio between the number E_i of edges that actually exist between the neighbors of a selected node i, and the total number $k_i(k_i - 1)/2$ of possible edges between these neighbors. The clustering coefficient of the graph, which is a measure of the network's potential modularity, is the average over all vertices:

$$C = \frac{1}{N}\sum_{i-1}^{N} C_i \tag{5}$$

The range of C is $0 \leqslant C \leqslant 1$.

2.1.4. Robustness

Small clusters of collaboration may be quite enduring, if most actors have ties to most other actors. As networks grow, and as they develop levels of "hubs," however, sustained "community" may become more difficult to maintain. The loss of a central "hub" may substantially disrupt the closeness of actors in scale-free networks; it is particularly true in purely "hierarchical" networks.

To sustain collaboration in larger networks with exponential degree distributions, it is important that there be some redundancy of ties between local clusters and more central hubs. Should one "leader" and "broker" become unavailable, a robust network of collaboration requires that local clusters be able to remain connected to distant actors by way of ties to other hubs.

The vulnerability of a collaboration network to disruption, or its robustness, can be examined empirically, by seeing how much disconnection results from the removal of the hubs. If the removal of a small number of leaders results in the network "falling apart" into disconnected component, the network lacks robustness. The ability to sustain network ties in scientific work is important to the cumulative development of deep specialist knowledge; just as connection to diverse and distant scientists may be important to creativity.

2.2. Micro-level structure: stratification

While each node has its own distinctive contribution to the whole network, it is not an easy thing to measure it effectively and directly. We applied a variety of centrality measures (degree, closeness, betweenness centrality) to investigate different contribution of nodes in this network, which are important to understanding power, stratification, ranking, and inequality in social structures (Wasserman & Faust, 1994). We also examine the extent to which there is a single or multiple "cores" or "elites."

2.2.1. Degree centrality

Generally, nodes with higher degree or more connections, in a sense, are more central to the structure and tend to have a greater capacity to influence others. This capacity may result in status distinction, and "elite" status. Cooperation networks are undirected network. When an author involved in the cooperation, he or she either be the "source" of information, or "sinks" or "receiver" of information. It is usually a measure of how influential (as "source") or prestigious (as "receiver") the node may be.

2.2.2. Closeness centrality

A more sophisticated centrality measure is closeness (Freeman, 1979) which emphasizes the distance of a node to all others in the network by focusing on the geodesic distance from each node to all others. Closeness can be regarded as a measure of how long it will take information to spread from a given node to others in the network. Degree has a shortcoming because it only takes into account the immediate links that a node has, rather than links to all others. One node might be tied to a large number of others, but those others might be rather disconnected from the network as a whole. In a case like this, the node could be quite central, but only in a local neighborhood. Degree centrality, then, identifies actors who are locally influential (and may be globally influential, as well). Closeness centrality focuses on the extensivity of influence over the entire network.

Let CL_i be the closeness of node i, d_{ij} has the same meaning as in function (1)

$$\mathrm{CL}_i = \sum_{j=1}^{N} \frac{1}{d_{ij}} \tag{6}$$

In this case the closeness is the sum of the reciprocated distances so that infinite distances contribute a value of zero (Borgatti, Everett, & Freeman, 2002).

2.2.3. Betweenness centrality

Another important class of centrality measures is the class of betweenness measures. Betweenness is a measure of the extent to which a node lies on the paths between others which can evaluate individual's ability to control the flow of knowledge between most others. The simplest and most widely used betweenness measure is that of Freeman (1977, 1979), which also be referred to as geodesic betweenness. Let B_i be the betweenness of node i; g_{jik} be the all geodesics linking node j and node k which pass through node i; g_{jk} be the all geodesics linking node j and node k .The betweenness of node i is as follows (Borgatti et al., 2002):

$$B_i = \sum_{j,k \neq i} \frac{g_{jik}}{g_{jk}} \tag{7}$$

In collaboration networks, actors with high "betweenness" are the brokers and connectors who bring others together. Being between means that an actor has the ability to control the flow of knowledge between most others. Individuals with high betweenness are the pivots in the network knowledge flowing. The nodes with highest betweenness also result in the largest increase in typical distance between others when they are removed. So, we can see high betweenness nodes are crucial for the formation of small-world effect.

2.2.4. K-core

Individual actors may have high centrality, but if these actors are not closely connected to one another, they do not form a single "elite" or "stratum." Some exponential networks may display multiple factions or elites that are only loosely connected – a "pluralist" structure. Other collaboration networks, though, may display a unitary and tightly bounded "core." The notion of K-core is introduced by Seidman (1983). The K-core of a network is a sub-structure in which each member has ties to at least K other members. By varying the value of K, we can see whether a network displays a single or multiple cores, and how tightly linked the members of cores are. Nodes in the core with larger values of K correspond to vertices that are more tightly linked, and form groups in more central positions in the network's structure. The core of maximum order is also called the main core. There exists an efficient algorithm for determining the cores in software Netdraw (Borgatti, 2002).

We denote the neighborhood of node $v \in V$ by $N(v)$:

$$N(v) = \{u \in V : (v : u) \in E\} \tag{8}$$

And the rooted neighborhood of node $v \in V$ by $N^{+}(v)$:

$$N^{+}(v) = N(v) \cup \{v\} \tag{9}$$

3. The structure of the COLLNET network

COLLNET (abbreviation of collaboration network) is an affiliation network. It is a global interdisciplinary research network, founded in 2000. The focus of this organization is to examine the phenomena of collaboration in science; its effect on production, innovation and quality, and the benefits and outcomes accruing to individuals, collaboration in e-science. This group is made up of 64 members from 20 countries of all continents in 2003 (the date of the data reported here). The members intended to cooperate on both theoretical and applied aspects on the topic "Collaboration in Science and in Technology." Kretschmer et al. collected all possible data about the cooperation of COLLNET members which are the basic data of our present work. There are 223 bibliographic multi-authored publications (including books, articles in peer reviewed journals, contributions in monographs, articles in conference proceedings or manuscripts) what reflect at least two COLLNET members' collaboration relationship. More details of the COLLNET can get in the Webpage: www.collnet.de. For the 64 members of COLLNET, until 2003, there still are 16 members who have cooperated with nobody else in COLLNET. In our study, because we pay more attention to the structure of the cooperation network and knowledge flowing through this network, they are excluded from our analysis. So, in our database, there are 48 members who at least cooperate with one COLLNET member.

Fig. 1 is the cooperation graph of the COLLNET members. There are total 48 nodes and 63 links in the network. By removing six smaller isolate components from the graph G we get the truncated COLLNET collaboration graph G′ in which all nodes can keep touch with every others through one or more paths. G′ contains 32 nodes and 49 edges. A glimpse to Fig. 1, we can get an implicit view that some nodes play an important role in the network, while others are peripheral. But what extend do they affect the network, and what role do the others play in the network is blurred. To make the question clearer is our main goal in this paper.

3.1. Macro-structure: connection

3.1.1. "Small-worlds"

The definition (1) of the mean geodesic distance l is problematic in networks that have more than one component (i.e. un-connected sub-graphs). In such cases, there exist node pairs that have no connecting path. Conventionally one assigns infinite geodesic distance to such pairs, but then the value of l also becomes infinite. To avoid this problem one usually we define l to be the mean geodesic distance between all pairs that have a connecting path. Pairs that fall in two different components are excluded from the average (Newman, 2003).

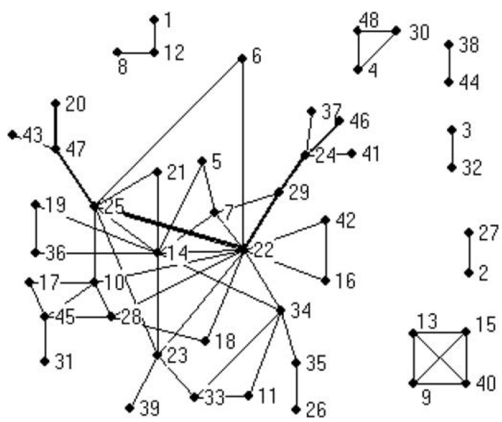

Fig. 1. The cooperation graph of COLLNET network (Kretschmer & Aguillo, 2004) (a series of bold lines is the diameter of the network).

We apply the approach described by Newman to calculate the average geodesic distance to our study. The l of our graph is 3.020. The appropriate phrase for COLLNET, then, is perhaps "three degrees of separation," for the typical distance between pairs of COLLNET members. For some other databases of collaboration that have been studied, such as MEDLINE, Los Alamos e-Print Archive, SPIRES, and NCSTRL where contain more than 10,000 nodes, the average distances are no more than 10 (Newman, 2001a, 2001b). To get another notion of the distance of a network, we might think about diameter. The diameter of a network is the largest geodesic distance in the (connected) network. The diameter of a network tells us how "big" it is. In our network, the reachable biggest component's diameter is 6. One example of this diameter can be seen as the bold line in Fig. 1. The diameter of Medline, Los Alamos e-Print Archive, SPIRES, and NCSTRL is 24, 18 (mean value of the four database), 19, and 31 respectively. Considering the small scale of our network, both of the average geodesic distance and the diameter of the COLLNET show less connection than we might expect (for a network of its size).

There are two possible reasons for the relatively long distances of our network. On the one hand, geographical distance may retard cooperation between COLLNET members. The COLLNET members come from more than 20 countries and regions where they spend large part of their time in. To communicate with each other face by face is rather expensive. On the other hand, the fields of study of many COLLNET participants do not have strong traditions of collaborative production. For biomedicine, high-energy physics, and astrophysics, collaboration among scientists is a necessary condition. Large numbers of persons are involved in gene mapping for the Human Genome Project, or are involved in running modern accelerators for elementary particle detection, or are involved in exploration cosmos abstruseness with complex, delicate facilities. In informetrics or scientometrics, the most common location of COLLNET researchers, solitary authors have long been a common circumstance.

3.1.2. Clustering

A common property of social networks is that cliques form, representing circles of friends or acquaintances in which every member knows every other. This tendency towards "clustering" or "transitivity" of relations tends to create patterns of small but densely connected separate clusters. When local clustering is combined with the integrating effects of "hubs" (apparent from the exponential degree distribution), we have "small-worlds" in which most scientists collaborate primarily in dense local clusters; but these clusters are at very short distances from one another.

For the COLLNET network as a whole, the clustering coefficient is the average of all individual C_i's. The result is: average cluster coefficient is $C = 0.643$. That is to say, on the average, the collaborators of any one node have a high probability of being collaborators with one another. This clustering is higher than most of cooperation networks (Medline: 0.066; Los Alamos e-Print Archive: 0.380; SPIRES: 0.726; NCSTRL: 0.496) (Newman, 2001a, 2001b, 2001c).

Function (4) is the most general method to calculate the clustering coefficient. However, for disconnected graphs, it can be somewhat misleading. For example, in Fig. 1, both group 1 including 9, 13, 15, 40 and group 2 including 14, 22, 23, 25 clustering is very high. According to function (4), the clustering coefficient of these eight nodes is 1. But this does not capture, well, the separation of the two groupings. An alternative approach is shown as function (10).

$$C'_i = \frac{E_i}{\mathrm{MaxDeg}} C_i \tag{10}$$

MaxDeg is the maximum degree of nodes in a network. The value of C' of our network is 0.097 which is lower than C. For a small closed cooperation group, in general, the C is high. In order to grasp more detail of the network, other indicator should be included, such as weight of cooperation.

3.1.3. Average degree and degree distribution

The average degree of the nodes in a network can reflect the compact-ness of the network. The higher the average degree, the tighter the network. The average degree of nodes in our network is 2.625, which means that the cooperation between COLLNET members is sparse. This is the same as Kretschmer's result (Kretschmer, 2004). The average degree also reflects the total number of people with whom a scientist wrote publications. The average number of collaborators of COLLNET members is lower than that of the purely theoretical disciplines (3.87 in high-energy theory, 3.59 in computer science), much less in the wholly or partly experimental ones (18.1 in biomedicine, 15.1 in astrophysics) (Newman, 2001a, 2001b, 2001c). This result is easy to understand; the subject matter of COLLNET is rather narrow, and the network is very young (founded in 2000).

The range and variability of degree can be quite important, because it describes whether the population is homogeneous or heterogeneous in structural positions. One could examine whether the variability is high or low relative to the typical scores by calculating the coefficient of variation (standard deviation divided by mean, times 100) for degree. By the rules of thumb that are often used to evaluate coefficients of variation, the current values: 82, is rather high. Clearly, the population is rather diverse in this regard (Hanneman, 2004).

Heterogeneity of degree means a few nodes have a large quantity of connections, but a big part of the nodes only have few connections. This is shown in Fig. 2, which displays the degree distribution $p(k)$ – the probability that k nodes connect to a certain node. Fig. 2 is the degree distribution of COLLNET cooperation (in logarithms to the base of 10). Graphing on the double-log scale reveals the extent to which the degree distribution is exponential (or "scale-free"). We note that the fit of the exponential model ($R^2 = 0.9134$) is quite good. We can also see that, although the COLLNET is not a big network, it shows an interesting scale-free characteristic: $p(k) \sim k^{-\gamma}$, where $\gamma = 1.37$. That is, the exponential has a modest slope; while there is an exponential distribution, the "steepness" of the inequalities is not great.

The presence of an exponential degree distribution suggests that the individual scholars in the network are more closely connected than would be expected by a process of random collaboration. Instead, it suggests that there may be "preferential attachment" in which actors who have previously collaborated are more likely to be parts of new collaborations than those who have not previously collaborated. To the degree that this occurs, previously distant actors are brought together into new collaborations by way of central "hub" actors.

3.1.4. Robustness of the network

Some networks display a surprising degree of robustness: although key components regularly malfunction, local failures rarely lead to the loss of the global material-carrying ability of the network. However, error tolerance is not shared by all systems: it is displayed only by a class of inhomogeneously wired scale-free networks (Albert, Jeongand, & Barabasi, 2000). Such networks display an unexpected degree of robustness, the ability of their nodes to communicate being unaffected even by unrealistically high failure rates. However, error tolerance comes at a high price in that these networks are extremely vulnerable to attacks (that is, to the selection and removal of a few nodes that play a vital role in maintaining the network's connectivity). Such error tolerance and attack vulnerability are generic properties of communication network. According to the

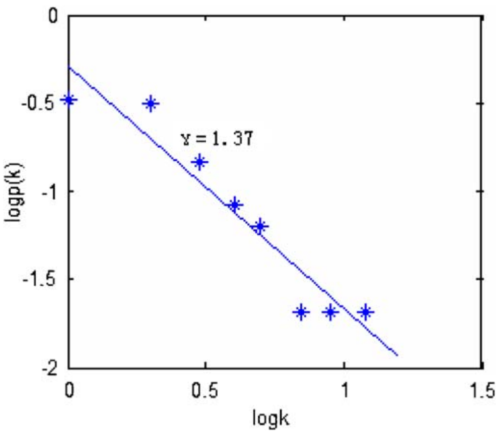

Fig. 2. Degree distribution of COLLNET.

feature of such kind of network, there are also some papers studying the different defensive strategies aimed at various attack strategies (Eubank et al., 2004; Liljeros, Edling, Stanley, Åberg, & Amaral, 2003).

In networks of scientific collaboration, cumulative knowledge and a sense of community identity (and the emergence of new "fields" of study) may depend on the robustness of collaboration. In some "schools," the death or departure of a key researcher may disconnect the flow of cooperation, and result in the collapse of the collaboration network. Other networks (particularly those with less hierarchical structures) are more "robust" in the face of the loss of central "hubs" or leading figures.

Though our network is not large-scale network, there exist some nodes which play important roles in maintaining the structure of the network. To test robustness, we remove some of them in order to see if collaboration breaks down. First, we used KEYPLAY (a program for identifying an optimal set of nodes in a network) to identify an optimal set of key nodes for deletion with the hope of crippling the network. With the selection of the criteria for fragmentation of the network into disconnected components, the result comes out: nodes 14, 22, 25 are the most influential nodes for the integration of the COLLNET network. After removing three nodes and all connection related to them, we draw the remaining network collapses into eight partitions (see Fig. 3).

The loss of three "hubs" in the COLLNET network results in considerable fragmentation; this suggests that the community is somewhat fragile, and dependent on central figures to maintain connection across the entire graph. Removal of additional, less "central" nodes, however, does not result in a complete collapse of collaboration. Fig. 4 shows the structure that remains after random deletion half of all actors.

Under the random removing of nodes, until more than 50% nodes are removed, the graph collapses into six pieces. In other words, a connected cluster of nodes that spans the COLLNET survives even for quite large fractions of crashed nodes. COLLNET thus displays many of the characteristics of a "scale-free" network, and also a pattern of fairly robust connection that maintains connection among distant actors, even in the absence of "key" individual participants.

3.2. Individual level analysis: stratification

Within any structure of collaboration, some individuals are more influential and visible than others as a result of their position in the network. Individuals in "central" positions in the network may also form an "elite" of a sort if they are connected to one another. The presence of "elites" may limit collaboration across

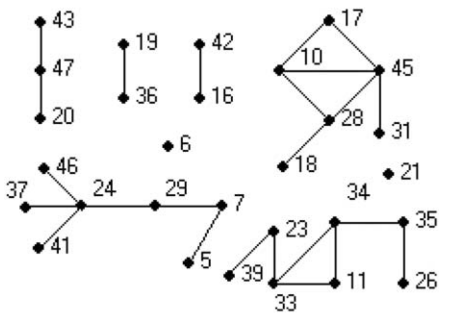

Fig. 3. Main-COLLNET after removing the nodes 22, 14, 25.

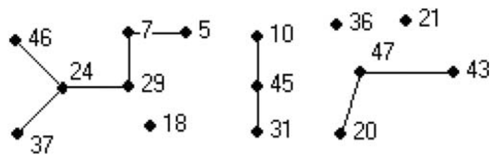

Fig. 4. Graph after random deletion of more than 50% of all nodes.

the network if members of the elite form and "exclusionary" "status community." So, the existence of small worlds and hubs – which may be so useful to efficient communication and creativity – may also generate patterns of inequality that are inconsistent with democratic participation and open communication.

Using measures of centrality, we are able to identify actors in the COLLNET network who occupy more "influential" positions. Table 1 shows these results.

3.2.1. Degree centrality

We can see that node 22 has links to one third of all nodes in graph G′. Nodes 22, 14, and 25 are similar in cooperating with more than one fifth of the members in G′; nodes 43, 20, 46, 41, 37, 31, 26, 39 (can be seen in Fig. 1) cooperate with only one other author. Nodes in the first set have a higher potential to exert influence or be influenced (though it might matter to whom they connected, this measure does not take that into account); nodes in the latter set have lower potential to exert influence or be influenced; nodes in the "middle" will be influenced or influence others if they are connected to the "right" other nodes, otherwise, they might have very little influence or difficult to influence others.

3.2.2. Closeness centrality

For closeness centrality, node 22 is the closest or most central actor according to length of all geodesic distance from it to all others (the disconnected distances are excluded). Two other nodes (14, 25), however, are nearly as close. Through the results of closeness centrality we can see the distance from a node to the whole graph will be shortened greatly if it has the "right" neighbors. Nodes 45, 24, 7, and 28 have the same degree 4, but their closeness is quite different just because nodes 7, 28 have more "central" neighbors than that of the other two nodes.

3.2.3. Betweenness centrality

Betweenness centrality indexes if a node has a favored position by falling on the geodesic paths between other pairs of actors in the network. That means the more nodes depend on a given node to make the shortest connections with other nodes, the more power this node has. In our network, still node 22 appears to have more power than others by this measure. Clearly, there is a structural advantage for node 22 to perceive that it is different from others. And this result is not surprising, because author 22 is the chairman of this organization. Still we can see the ordering of nodes according to betweenness centrality changed from that according to the degrees. All aspects of power are not distributed in the same way.

We can see from Table 1 that node 22 is the genuine dominator (or "star") in our network. On each centrality dimension, this actor occupies the first place. If actor 22 and all the connections to it are deleted, we found nearly all the indicators of G′ changed sharply. Average Geodesic distance changed from 3.079–3.642; cluster coefficient from 0.643 to 0.566; diameter from 6 to 8. That means the knowledge flowing was set back deeply since the removing the node 22.

Table 1
Centrality of major nodes

Node ID	Degree centrality	Closeness centrality (rank)	Betweenness centrality (rank)
22	12	57 (1)	256.867 (1)
14	9	67 (2)	102.867 (4)
25	7	67 (2)	109.967 (2)
23	5	73 (3)	46.700 (9)
10	5	75 (5)	62.900 (7)
34	5	74 (4)	91.300 (5)
45	4	100 (17)	30.833 (10)
24	4	101 (18)	87.000 (6)
7	4	76 (6)	18.467 (13)
28	4	80 (8)	21.767 (12)

Another thing we can find is that only one indicator cannot reflect the individual's position in the structure completely. Nodes with only some links can vary more in their behavior, depending on to whom they are connected.

3.2.4. K-core

As we have seen, some actors are more central in COLLNET than others. We have also seen that the actors who are central in one way are not necessarily central in another. To what extent, then, do the central actors form a "core" of the network? Is there a single "elite" or multiple "elites"? Examining the K-core structure of the graph provides some insight. In Fig. 5, the "cores" of the network are shown, with progressively wider circles showing the increasing inclusion of actors as the "tightness" of connection within the group is relaxed.

We can see that the network displays two "core" sets (40, 9, 13, 15 and 23, 25, 14, 22). The former appears to be an isolated "faction," while the latter constitutes an "inner circle" of the main component of the graph.

In G, the main core is of order 3 (that is, actors have collaborations with at least three other members of the core). Table 2 shows the distribution of number of authors in K-cores in G, and the distribution of number of co-authors in cores for selected members of the main core and the secondary core, are given.

Table 2 shows clearly the isolation of one core (i.e. the one which includes actor 9). It also reveals that the four "elite" members of the other core are very active in maintaining ties outside their group; the main "core" is not a "closed" community.

The authors belonging to the main core and some of their characteristics are presented in Table 3, where $\overline{\text{deg}}$ is the average degree of all co-authors; $\overline{\text{core}}$ is the average core of all co-authors.

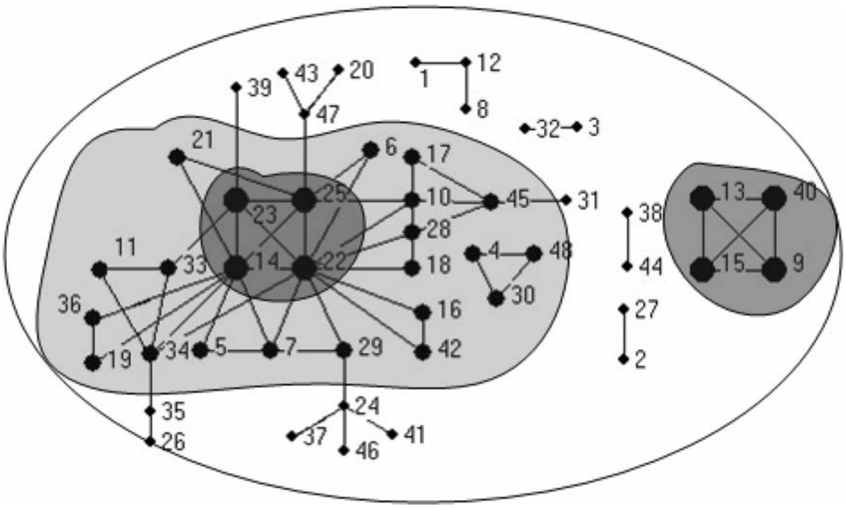

Fig. 5. K-core (the size of nodes depending on the core of each nodes; corresponding to white, darker, and darkest background are the 1, 2, and 3 core).

Table 2
Distribution of number of co-authors in cores

Core	No of authors	22	23	14	25	9
3	8	3	3	3	3	3
2	20	9	1	6	3	0
1	20	0	1	0	1	0
Sum	48	12	5	9	7	3
Average core	1.75	2.25	2.4	2.33	2.28	3

Table 3
Authors and number of their co-authors in the main core, total number of co-authors, average core and average degree of all their co-authors, and their collaborativenesss or emergingtiveness

Author	Co-authors	All co-authors	$\overline{\text{core}}$	$\overline{\text{deg}}$	coll
22	3	12	2.25	4.17	1.33
23	3	5	2.4	6.4	1.25
25	3	7	2.28	5.43	1.316
14	3	9	2.33	4.56	1.288
9	3	3	3	3	1
26	0	1	1	1	1
31	0	1	2	4	0.5
39	1	1	3	1	0.33

$$\overline{\deg}(v) = \begin{cases} 0, & N(v) = \phi \\ \frac{1}{|N(v)|} \sum\limits_{u \in N(v)} \deg(u), & \text{otherwise} \end{cases} \tag{11}$$

$$\overline{\text{core}}(v) = \begin{cases} 0, & N(v) = \phi \\ \frac{1}{|N(v)|} \sum\limits_{u \in N(v)} \text{core}(u), & \text{otherwise} \end{cases} \tag{12}$$

We can explore a scientist's cooperation tendency when we think about the $\overline{\deg}$ and the $\overline{\text{core}}$, high values imply that scientist is tend to collaborating with "central" authors. Therefore, measures of "collaborativeness" (Batagelj & Mrvar, 2000) and "mergingtiveness" would be very important to evaluate authors' connective tendency. Quantity:

$$\text{coll}(v) = \frac{\text{core}(v)}{\overline{\text{core}}} \tag{13}$$

The collaborativeness and "mergingtiveness" are the two terms based on the same approach coll(v), measure the openness of author v towards other authors. If core(v) = 0, also coll(v) = 0.

When an author is in the main core, high value of coll stands for high tendency of that author v towards "peripheral" authors, which can measure the "collaborativeness" of that author. When an author is in the peripheral core, low value of coll stands for high tendency of that author v towards "central" authors, which can measure the "mergingtiveness" of that author. The ordering of most collaborative authors in G is nodes 22, 25, 14, 23, then 9 (13, 15, and 40). Node 22 lies on the top too. Further we study the "mergingtiveness" of "peripheral" authors whose degree all are 1. Only nodes 39 and 31's situation are different from others nodes (we list node 26 on behalf of these uniform nodes) because their neighbor's degree is larger than 1. The highest "mergingtiveness" belongs to the node 39 what means it has more opportunity to merge into the main core according to clustering characteristics of the social networks.

4. Conclusions and future applications

In this paper, we presented some possible approaches to analysis of networks and applied them to an affiliation network-COLLNET. Although COLLNET is not a large-scale network, the results are useful to provide an illustration of the main approaches to measuring structural characters, and identify nodes who are the most important hub nodes and who have the most influence in the network.

It is interesting that the COLLNET is close to a scale-free network, and displays the clustering aspect of a "small-world" as well. Structure affects function. The COLLNET network suffers the vulnerability in that it depends too much on few nodes who are crucial to the whole network for knowledge production or spreading. Such as, nodes 22, 14, 25 are of the highest degree in network. At the same time they are the champions in most other measures too. At the same time, there is a degree of robustness and institutionalization apparent in this relatively young network – which may lead, in time, to defining a new field of study. The leading figures do not form a "closed" elite, and the network as a whole is fairly robust against the loss of central figures.

Another thing we want to point out is that in COLLNET network few nodes always locate at the top of many measures of influential position. The network displays a "star" and an "elite" of actors in favored locations. The same forces that give "small-worlds" utility in collaboration networks (short path lengths and clustering) also, necessarily, generate inequality or stratification. Such stratification, however, may either promote or retard creativity and a sense of shared identity. Where highly active nodes form closed and isolated groupings (as in one case in COLLNET) they promote factionalization and specialization. Where the highly active "hub" nodes also connect to wider circles of participants, as in the main "core" of COLLNET, elites may promote diversity, communication, and integration of the community.

Acknowledgements

We gratefully acknowledge the inspiring and professionally motivating atmosphere in the WISELAB of Dalian Technology of University. Thanks to Wang Bing and Vladimir Batagelj for their discussion and direction on this topic.

Appendix A

Explanation for Figs. 1, 3, 4 and 5

1	Aguillo Isidro.
2	Ahrweiler Petra
3	Ambuja R.
4	Bassecoulard Elise
5	Basu Aparna
6	Beaver Donald deB.
7	Bhattacharya Sujit
8	Bordons Maria
9	Brandt Martina
10	Davis Mari
11	Egghe Leo
12	Gomez Isabel
13	Grosse Ulla
14	Gupta Brij Mohan
15	Hartmann Frank
16	Havemann Frank.
17	Hood William W.
18	Jansz Margriet
19	Karisiddappa
20	Katz Sylvan
21	Kharbanda Ved Prakash
22	Kretschmer Hildrun
23	Kundra Ramesh
24	Leydesdorff Loet
25	Liang Liming
26	Liberman Sofía
27	Liu Zeyuan
28	Markusova Valentina
29	Meyer Martin
30	Okubo Yoshiko
31	Osareh Farideh
32	Raghavan Koti S.

(continued on next page)

Appendix A (*continued*)

33	Rao Ravichandra
34	Rousseau Ronald
35	Russell Jane
36	Sangam Shivappa
37	Scharnhorst Andrea
38	Schulze Annedore
39	Tomov Dimiter
40	Voss Rainer
41	Wagner Caroline
42	Wagner-Döbler Roland
43	Wang Yan
44	Wenzel Vera
45	Wilson Concepcion S.
46	Wouters Paul
47	Wu Yishan
48	Zitt Michel

References

Albert, R., & Barabasi, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics, 74*, 47–97.

Albert, R., Hawoong, J., & Barabasi, A. L. (1999). Diameter of the World-Wide Web. *Nature, 401*, 130–131.

Albert, R., Jeongand, H., & Barabasi, A. L. (2000). Error and attack tolerance of complex networks. *Nature, 406*, 378–381.

Barabasi, AL., Jeong, H., Ravasz, E., Neda, Z., Schuberts, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A, 311*, 590–614.

Batagelj, V., & Mrvar, A. (2000). Some analyses of Erdos collaboration graph. *Social Networks, 22*, 173–186.

Beaver, D. D., & Rosen, R. (1978). Studies in scientific collaboration. Part III. Professionalization and the natural history of modern scientific co-authorship. *Scientometrics, 3*, 231–245.

Borgatti, S. P. (2002). *NetDraw: Graph visualization software*. Harvard: Analytic Technologies.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet 6 for Windows*. Harvard: Analytic Technologies.

Braun, T., Glanzel, W., & Schubert, A. (2001). Publication and cooperation patterns of the authors of neuroscience journals. *Scientometrics, 51*(3), 499–510.

Calero, C., van Leeuven, T. N., & Tijssen, R. J. W. (2005). Research networks of pharmaceutical firms: geographical patterns of research collaboration within and between firms. In P. Ingwersen & B. Larsen (Eds.). *Proceedings of the 10th ISSI international conference on scientometrics and informetrics, July 24–28, 2005, Stockholm, Sweden* (1, pp. 310–315). Stockholm: Karolinska University Press.

Egghe, L., Rousseau, R., & Van Hooydonk, G. (2000). Methods for accrediting publications to authors or countries: Consequences for evaluation studies. *Journal of the American Society for Information Science, 51*(2), 145–157.

Eubank, S., Guclu, H., Anil Kumar, V. S., Marathe, M. V., Srinivasan, A., Toroczkai, Z., et al. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature, 429*(13), 180–184.

Freeman, L. C. (1977). A set of measures of centrality based upon betweenness. *Sociometry, 40*, 35–41.

Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks, 1*, 215–239.

Glänzel, W., & Schubert, A. (2004). Analyzing scientific networks through co-authorship. In H. F. Moed et al. (Eds.), *Handbook of quantitative science and technology research* (pp. 257–276). The Netherlands: Kluwer Academic Publishers.

Hanneman, R. A. (2004). Introduction to social network methods. As e-print available at http://www.analytictech.com/networks.pdf on May 25, 2004.

Kretschmer, H. (2004). Author productivity and geodesic distance in bibliographic co-authorship networks, and visibility on the web. *Scientometric, 60*(3), 409–420.

Kretschmer, H., & Aguillo, I. F. (2004). Visibility of collaboration on the web. *Scientometrics, 61*(3), 405–426.

Kretschmer, H., & Aguillo, I. F. (2005). New indicators for gender studies in web networks. *Information Processing & Management, 41*(6), 1481–1494.

Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E., & Aberg, Y. (2001). The web of human sexual contacts. *Nature, 411*, 907–908.

Liljeros, F., Edling, C. R., Stanley, H. E., Åberg, Y., & Amaral, L. A. N. (2003). Sexual contacts and epidemic thresholds. *Nature, 423*(5), 605–606.

Milgram, S. (1967). The small world problem. *Psychology Today, 2*, 60–67.

Newman, M. E. J. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E, 64*. Art. no. 016131.

Newman, M. E. J. (2001b). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E, 64*. Art. no. 016132.

Newman, M. E. J. (2001c). The structure of scientific collaboration networks. *PNAS, 98*(2), 404–409.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review, 45*(2), 167–256.

Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science, 28*, 443–455.

Price, D. J. de S. (1963). *Little science, big science*. New York: Columbia University Press.

Price, D. J. de S. (1965). Networks of scientific papers. *Science, 149*, 510–515.

Rapoport, A. (1957). Contribution to the theory of random and biased nets. *Bulletin of Mathematical Biophysics, 19*, 257–277.

Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B, 4*, 131–134.

Seidman, S. B. (1983). Network structure and minimum degree. *Social Networks, 5*, 269–287.

Wagner, C. S. (2005). Six case studies of international collaboration in science. *Scientometrics, 62*(1), 3–26.

Wasserman, S., & Faust, K. (1994). *Social network analysis*. Cambridge: Cambridge University Press.

Watts, DJ., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440–442.

Yin Li-chun is a doctor degree student in Dalian University of Technology, Dalian, China. Her major is Science Study and Management of Science and Technology. Her work deals with bibliometrics from the viewpoint of network. This one is her first international paper. She has published 6 articles in national journals.

Hildrun Kretschmer is associate researcher in the Department of Library and Information Science at the Humboldt-University of Berlin, Germany, Visiting Professor at the Dalian University of Technology, Dalian, China, and Honorary Professor, Xinxiang, China. Her work deals with collaboration processes in science from the viewpoints of bibliometrics and socio-psychology, modelling of collaboration systems and theory, development and change of dynamic social structures in science. She is appointed as Associate Editor of both the international journal *Scientometrics* and the *Journal of Information Management and Scientometrics.* Currently, she is co-ordinating the Global Interdisciplinary Research Network COLLNET: "Collaboration in Science and in Technology", Web site: http://www.collnet.de. She is editor of 10 conference proceedings and she has published more than 70 peer reviewed articles or contributions in national and international journals or monographs, several non-peer reviewed publications. She was working as organizer or program chair of nine international conferences or workshops in the area of webometrics, scientometrics and informetrics, collaboration in science and in technology.

Robert A. Hanneman is a professor of sociology at the University of California, Riverside. He works mostly in statistical and mathematical sociology, social networks, simulation and theory construction, economic, political, and military sociology.

Liu Ze-yuan is a professor and tutor of doctor degree student in Science Study and Management of Science and Technology at Dalian University of Technology, Dalian, China. He works mostly in science and technology strategy, science study theory.