Connection and Stratification in Research Collaboration: An

Analysis of the COLLNET Network

YIN Li-chun [1],
Hildrun Kretschmer [1,3],
Robert A. Hanneman[4],
LIU Ze-yuan [1]_
Wang Bao-lin[4]

[1.] WISE LAB, Dalian University of Technology, Dalian, 116023, China E-mail: yinliuxing1290@yahoo.com.cn; [2.] Department of Library and Information Science, Humboldt-University Berlin, Dorotheenstr. 26, D-10099 Berlin, Germany;[3.] Department of Sociology, University of California - Riverside, USA [4.] The sixth job region, The second oil factory,Daqing,163411, China

Abstract

The co-authorship network of scientists represents a prototype of social networks. 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. Structure of the network effect not only individual's collaboration pattern deeply, but knowledge flowing process profoundly. However, a single view is not enough to grasp the characters of the network in details. We have to either produce global view on the structure of the network as a whole, or make a detailed inspection of some features of individuals. In this paper, some techniques for analyzing complex networks (different approaches to identify "interesting" individuals and groups, analysis of internal structure of the main core and hierarchical clustering) and visualizations of their parts are presented on the case of COLLNET collaboration graph.

Keywords: Social network analysis, Cooperation network, Topological Structure, Evaluate

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.Collaboration may be seen as a process in which knowledge flows among scientists, 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 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.

In this paper we will examine the structure of a network of 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). 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.

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. The Structure of the COLLNET network

There are 64 members in COLLNET until 2003, among them 16 members cooperated with nobody else in COLLNET. In our study, 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 6 smaller isolate components from the graph G we get the truncated COLLNET collaboration graph-G'. G' contains 32 nodes and 49 edges.

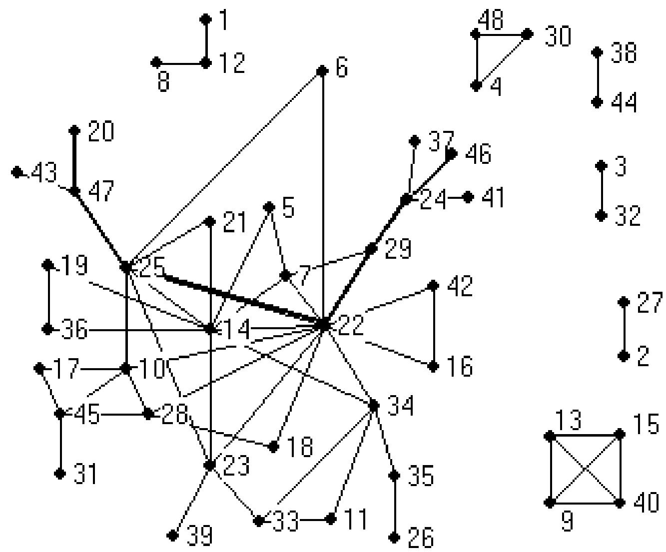

Fig. 1. The cooperation graph of COLLNET network (Kretschmer, 2004) (a series of bold lines is the diameter of the network).

2.1. Macro structure: Connection

2.1.1. "Small-worlds"

The mean geodesic distance 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. To get another notion of the distance of a network, we might think about diameter. In our network, the reachable biggest component's diameter is 6. One example of this diameter can be seen as the bold line in the Figure 1. 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).

2.1.2. Clustering

A common property of social networks is that cliques form, representing circles of friends or acquaintances in which every member knows every other. The extent of the clique can be expressed by clustering coefficient.

For the COLLNET network as a whole, the clustering 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.

2.1.3. 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. The average degree also reflects the total number of people with whom a scientist wrote publications. This result is easy to understand; the subject matter of COLLNET is rather narrow, and the network is very young (founded in 2000).

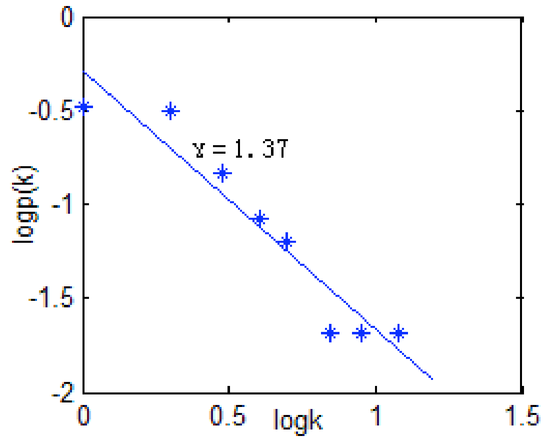

Fig. 2. Degree distribution on COLLNET

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 Figure 2, which displays the degree distribution *p(k)* -- the probability that k nodes connect to a certain node. Figure 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: , where ' =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.

2.2. Individual level analysis: Stratification

2.2.1. 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 been 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 been 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.

2.2.2. 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.

2.2.3. 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.

3. 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.