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 © 2006 Elsevier Ltd. All rights reserved.
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.