Study of Scientific Collaborations in the Intelligence and Security Informatics Research Community by Social Network Analysis

Chi-Bin Cheng[1], Min-Yuh Day[1], Sheng-Pao Shih[1], Weide Chang[2]

[1] Department of Information Management –Tamkang University
151 Yingzhuan Rd., Tamsui Dist., New Taipei City, Taiwan

[2] Department of Computer Science –California State University
Sacramento, USA

{cbcheng,myday,sbao}@mail.tku.edu.tw, changw@ecs.csus.edu

Abstract. *The field of Intelligence and Security Informatics (ISI) is a relatively new research area and co-authorship is the predominant trend in the field of ISI. This study aims to investigate the collaborative characteristics in the ISI research community through social network analysis. The collaboration network of ISI researchers at the domain/disciplinary level based on publications in the field of ISI is built, and three centrality measures, closeness, betweenness, and degree centralities, are used to identify the most prominent researchers in the community. The visualization tool of* pajek *is also used to identify the main actors, clusters and components in the network.*

1. Introduction

Scientific collaborations are important to facilitate and enhance the activities of scholarly research and communications (Ding 2011). The research network is referred to the collaboration relations among researchers, which can be observed in general via the co-authorship shown in the published papers. Two authors are connected in a collaboration network if they have written one or more papers together. The field of Intelligence and Security Informatics (ISI) is a relatively new research area and co-authorship is the predominant trend in the field of ISI. The science of (ISI) focuses on the development and use of advanced information technologies, including methodologies, models and algorithms, infrastructure, systems, and tools, for security related applications through an integrated technological, organizational, and policy based approach. Though there have been studies on the scientific collaboration patterns in Information Systems (IS) research, little attention has been paid to the ISI research community. Social network analysis (SNA), emerging from the interaction among and between human actors, is an interdisciplinary research paradigm that combines information systems, sociology, physics, biology, computer science, and management science. SNA is a useful tool when analyzing scientific collaboration patterns in a particular journal or conference (Fischbach et al., 2011).

The objective of this study is to provide a comprehensive view of the ISI research community to better understand its collaborative characteristics and thus

enhance the collaboration in this community. To this end, we build the collaboration network of ISI researchers at the domain/disciplinary level based on publications in the field of ISI, and investigate the structural characteristics of the collaboration network by SNA. The visualization tool of SNA is employed to identify the main actors, clusters, and components in the collaboration network.

Works on scientific collaboration pattern identification by SNA are reported in literature, e.g. Ding 2011 and Fischbach et al. 2011. There have been studies on the collaboration in the general Information System (IS) research community, e.g. Fischbach et al. (2011). However, to the best of our knowledge, the collaboration in ISI research field is not studied.

2. Methodology

Researchers who are the most prominent in the community are often located in the strategic locations on the co-authorship network, which may allow them: (1) to communicate directly with many other researchers, (2) to be close to many other researchers, or (3) to be as intermediary in the interactions of many other pair of researchers. The concept of centrality in social network analysis is suitable to identify the most "central" researchers in the co-authorship network (Newman 2004). There are several ways to define the centrality of a node in the social network, closeness centrality, betweenness centrality, and degree centrality are used to quantify the prominence of a researcher in the ISI research community.

Closeness centrality and degree centrality can be used to measure a researcher's impact in the research community and his value in scientific collaboration (Yan and Ding 2009). Closeness centrality indicates the position of a researcher in the network and his virtual distances from others. Degree centrality measures both strong and weak ties of a researcher; the researcher who coauthors with multiple researchers would have a higher degree centrality. Researchers involving in interdisciplinary research would have a higher betweenness centrality. Betweeness creates advantages of lowering the risk and increasing the value of collaboration; researchers with high betweenness centrality have more opportunities to broker the flow of information and thus contribute a greater value in scientific collaboration. Closeness centrality, betweenness centrality, and degree centrality are formally defined in the following subsections.

2.1. Centrality definitions

In an undirected network, the distance between two vertices (researchers/authors) is simply the number of edges in the shortest path (also called *geodesic*) that connects the two vertices. The distance from vertex v to vertex w is defined as the length of the geodesic from v to w. The closeness centrality of a vertex is based on the total distance between this vertex and all other vertices. When a vertex is close to all other vertices, this vertex is able to transmit information to others more efficiently, and thus is consider at the center of the network. The closeness centrality of a vertex v is defined as the number of all reachable vertices divided by the sum of all distances between the vertex v and all its reachable vertices (Newman 2004). Let G denote the graph, n is the number of all reachable nodes of v in G, and $d(v,w)$ is the pair-wise geodesic between v and w), the closeness centrality of v, $C(v)$, is defined as:

$$C(v) = \frac{n-1}{\sum_{v,w \in G} d(v,w)}. \quad (1)$$

Researchers with a great $C(v)$ are able to disseminate and access new information faster to and from others (Newman 2004).

The interaction between any two non-directly connected researchers (i.e., who never collaborated before) is likely to be provoked by the intermediate researcher who connects the two researchers through their shortest path(s). Such intermediate researchers play an important role in the network for controlling the flow of interactions. Researchers who lie between most of the shortest paths of the pairs of researchers could be viewed as the central persons in the community. This notion, known as the *betweenness* of a vertex v, $B(v)$, computes the number of geodesics between pairs of vertices passing through v, is formally defined as (Freeman, 1977).

$$B(v) = \sum_{v,w,x \in G} \frac{\sigma(w,u;v)}{\sigma(w,u)}, \quad (2)$$

where $\sigma(w, u)$ is the number of geodesics between w and x, and $\sigma(w, u; v)$ is the number of geodesics between w and u passing through v. The above equation can be interpreted as the sum of all probabilities a shortest path between each pair of nodes w and u passes through node v.

In a simple undirected network, the degree of a vertex is the number of vertices that are adjacent to this vertex, i.e. the number of its neighbors. Vertices with high degrees are more likely to form a dense section of the network. In the co-authorship network, the degree centrality of a vertex v is defined as its degree as presented in Equation (3). We can use the average degree of all vertices to measure the structural cohesion of a network (Nooy et al., 2005).

$$D(v) = \sum_{w} x_{vw}, \quad (3)$$

where x_{vw} equals 1 if there is a link between v and w, 0 otherwise.

2.2. Sample data

The study of scientific collaboration pattern in the ISI research community is based on the co-authorship presented in the international conferences on intelligence and security informatics held by IEEE (IEEE ISI). IEEE ISI is a well-established conference that focuses on research in the ISI research field. To compile an ISI literature list, we used the online data source DBLP Computer Science Bibliography. The data source supplies information of each article, including: title, authors, title of the conference, year of publication, and number of pages.

We develop a web focused crawler program to download literature information of all ISI papers published between 2003 and 2010 found in DBLP. After preprocessing the data, a list of 574 papers by 1204 authors was collected. It is noted that single-authored papers are not considered in the co-authorship network. Network analysis software, *Pajek* (Nooy et al., 2005), are used in this study for social network analysis. A

computer program is developed to convert the list of coauthors to the format of network file to be read by the network analysis software.

As noted in prior research (Xu and Chau, 2006), the most productive authors or leaders in a scientific community are important assets to that discipline. They often introduce new ideas into their research communities. In our sample data, there are 1204 distinct authors between 2003 and 2010 in terms of the amount of published articles. The most prolific authors are Hsinchun Chen (44), follow by Daniel Dajun Zeng (19), Fei-Yue Wang (19), Christopher C. Yang (12), Judee K. Burgoon (12), Mark K. Goldberg (10), Malik Magdon-Ismail (10), Ram Dantu (8), Jay F. Nunamaker (8), and Antonio Badia (7), where figures in the parentheses are the numbers of published articles.

3. Data Analysis and Discussion

The centrality measures discussed in the previous section are used to analyze the sample data and the visualization tool provided by *pajek* is employed to identify the clusters in the network.

3.1. Centrality analysis

Researchers with a high closeness centrality collaborated widely with others. These researchers form a core component in the ISI research community. The top 10 researchers with the greatest closeness centralities are shown in Table 1. An author with high closeness score is likely to receive information more quickly than others since there are fewer intermediaries between the author and others. These authors are closely connected to each other through collaborations.

Table 1. Top 10 researchers with the greatest closeness scores

Rank	Vertex ID	Closeness	Author
1	1	0.0591053	Hsinchun Chen
2	2	0.0515175	Daniel Dajun Zeng
3	43	0.0493501	Guanpi Lai
4	242	0.0476537	Michael Chau
5	26	0.0469205	Marc Sageman
6	37	0.0453844	Catherine A. Larson
7	231	0.0447190	Damien Daspit
8	21	0.0447190	Edna Reid
9	16	0.0447190	Jialun Qin
10	30	0.0445882	Yilu Zhou

Table 2 shows the top 10 authors with the highest betweenness scores. A high betweenness score implies that the author is able to control the flow of information in the ISI network and is thus considered as a leader in the ISI community network. Table 3 shows the top 10 authors with the highest degree centrality scores. The author has a high degree centrality indicating that he has collaborated with many authors.

It is noted that Hsinchun Chen is ranked as the first in all three centrality measures, and he is also the most productive researcher in the ISI research field. Daniel Dajun Zeng is ranked as the second in closeness and degree measures, and the third in betweenness measure; he is also ties with Fei-Yue Wang as the second productive researcher. Though Fei-Yue Wang is the second in the betweenness measure and the third in the degree measure, he is not in the top 10 list of the closeness measure (he is at the 19th place).

3.2. Visualization of social network analysis

Figure 1 shows the visualization of ISI co-authorship network using *pajek*, where the network of ISI research community is not fully connected from the global view. It contains a number of components (subsets) for which there are no paths between authors in one component and authors in another component. The visualization also shows the giant component in the ISI co-authorship network, where Hsinchun Chen and Daniel Dajun Zeng are closely connected to each other through collaborations, which could be viewed as the core component of the ISI community.

Table 2. Top 10 authors with the highest betweeness scores

Rank	Vertex ID	Betweenness	Author
1	1	0.0079069	Hsinchun Chen
2	3	0.0042949	Fei-Yue Wang
3	2	0.0030345	Daniel Dajun Zeng
4	26	0.0026572	Marc Sageman
5	4	0.0024364	Christopher C. Yang
6	43	0.0015718	Guanpi Lai
7	76	0.0013299	Yungchang Ku
8	191	0.0008990	Saurabh Gupta
9	242	0.0007622	Michael Chau
10	53	0.0007303	Justin Zhan

Table 3. Top 10 authors with the highest degree scores

Rank	Vertex ID	Degree	Author
1	1	0.0606816	Hsinchun Chen
2	2	0.0299252	Daniel Dajun Zeng
3	3	0.0274314	Fei-Yue Wang
4	15	0.0207814	Yuval Elovici
5	5	0.0199501	Judee K. Burgoon
6	14	0.0157938	Homa Atabakhsh
7	11	0.0157938	Jennifer Jie Xu
8	37	0.0157938	Catherine A. Larson
9	9	0.0149626	Jay F. Nunamaker

4. Concluding Remarks

This study investigates the collaboration between researchers in the ISI research community by social network analysis. Prominent researchers in the community are identified through analysis based on three centrality measures. The result provides an insight into the structural characteristics of research collaboration networks in ISI. The visualization tool is also informative for identifying the main actors, clusters and components in the network.

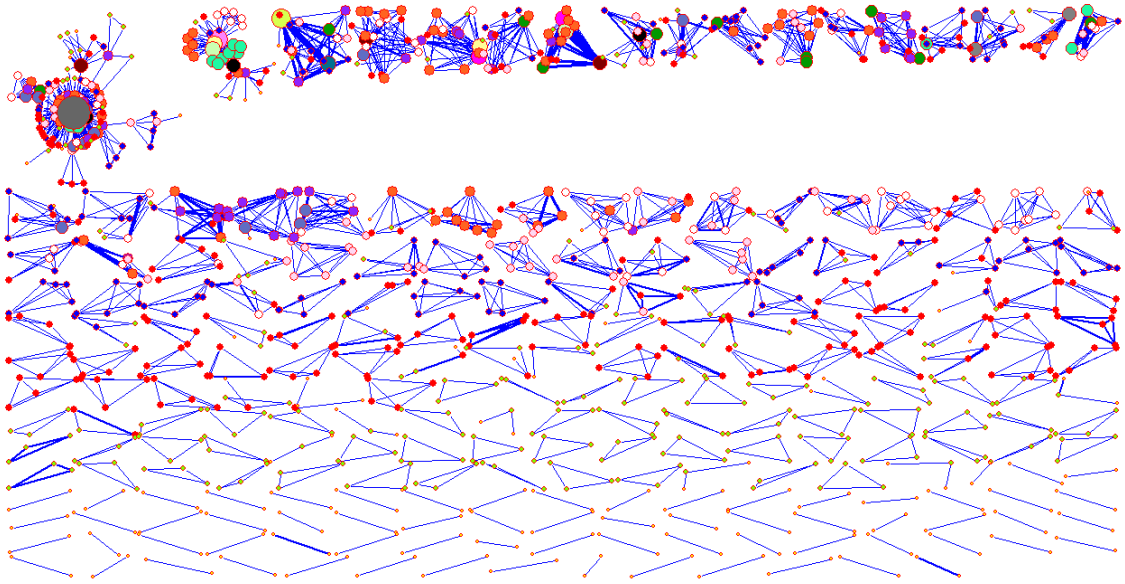

Figure 1. Visualization of the ISI co-authorship network (global view)

References

Ding, Y. (2011) "Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks," *Journal of Informetrics* (5:1), pp 187-203.

Fischbach, K., Putzke, J., and Schoder, D. (2011) "Co-authorship networks in electronic markets research," *Electronic Markets*, pp 1-22.

Freeman, L.C. (1977) "A Set of Measures of Centrality Based on Betweenness," *Sociometry* (40:1), pp 35-41.

Newman, M.E.J. (2004) "Coauthorship networks and patterns of scientific collaboration," in: *Proceedings of the National Academy of Sciences of the United States of America*, pp. 5200-5205.

Nooy, W., Mrvar, A., and Batagelj, V. (2005) *Exploratory social network analysis with Pajek* Cambridge University Press,.

Xu, J. and Chau, M. (2006) "The social identity of IS: analyzing the collaboration network of the ICIS conferences (1980-2005)," *Proceedings of the International Conference on Information Systems*, Milwaukee, pp. 569-589.

Yan, E.J. and Ding, Y. (2009) "Applying Centrality Measures to Impact Analysis: A Coauthorship Network Analysis," *Journal of the American Society for Information Science and Technology* (60:10), pp 2107-2118.