Co-Authorship Network and the Correlation with Academic Performance

Qianwen Xu[1, 2], Victor Chang[1]*

1. *School of Computing, Engineering and Digital Technologies, Teesside University, Middlesbrough, UK*
2. *IBSS, Xi'an Jiaotong-Liverpool University, Suzhou, China*

iamarielxu@163.com, victorchang.research@gmail.com

Keywords: Social network analysis, graph cluster analysis, co-authorship network, Spearman correlation test, link analysis; Internet of Things (IoT); Big Data Analytics.

Abstract: This paper aims to study the internal structure of the co-authorship network and the relationship between the network and the authors' academic performance in the network. In order to conduct this research, bibliographic data of 166 authors from three top higher education institutions of Shanghai was collected and the method of social network analysis (SNA) was performed to analyze the data. In the link analysis, the centrality, egocentric network efficiency, authorities, and hubs were analyzed. In the graph cluster analysis, this paper employs clustering algorithms based on betweenness. Lastly, the Spearman correlation test was performed to analyze the relationship between academic performance and SNA metrics. This paper found that and betweenness centrality, eigenvector centrality, authority and hub position, and efficiency were significant to g-index. The research provided a glimpse of the co-authorship network's internal structure in China. Additionally, the SNA method of identifying productive scholars can also be applied to other areas, such as the network of equipment in the Industry 5.0 to help companies identify the strong and weak links in the producing process.

1 INTRODUCTION

Today, an organization or a person's performance is often evaluated for management purposes. There are several purposes for doing the performance evaluation, including motivating the individuals or the organizations to produce with higher quality and quantity, guiding the individuals to achieve the objectives of personal development as well as of organizations, comparing an individual or an organization's performance with others, as well as providing evidence of non-discriminatory promotion (Heathfield, 2019). In the academic world, researchers' performance is evaluated by academic performance, such as teaching evaluation, research results, and other academic indicators. Evaluating the scholars' academic performance is crucial because the evaluation outcomes are employed to recruit and allocate funds and because colleges and universities can gain a great reputation for having highly productive scholars or researchers (Abbasi et al., 2012). Nevertheless, identifying, clustering and configuring productive researchers to optimize research synergies is not an easy job.

When the Internet of Things (IoT) merges with social networks, every building and every car can link to each other or interact with other people and things. The social network definition is no longer just the socialization of people but also the social network of people, people, people and things, and objects and things in a wider range. Scholars or papers in different areas or from various institutions can be connected as well. With the support of the emerging technologies, the scholars or researchers can be linked more easily with each other through papers and their co-authorship networks can be established easily and not limited to the number of literature databases.

In this research, the method of social network analysis (SNA) will be conducted to study the co-authorship network's internal structure and the correlation between the network and the authors' academic performance in the network. There are two reasons for this paper to study the co-authorship with the SNA method. On the one hand, this is a complementary SNA study to the area of co-authorship. Although the study on SNA is mature comparatively, in the area of co-authorship network, most papers mainly concentrated on the country or province level, instead of the level of cities or universities. There are few papers on the micro co-author networks. Therefore, this research will try to outline the internal structure of the co-author networks and took the case of Shanghai, China, as an example. On the one hand, this paper attempts to show the higher education institutions in detail how to use simple techniques or indicators of SNA to determine, cluster and configure productive and efficient scholars. It is important for higher education institutions to identify a scholar with excellent academic performance, but the accessibility and simplicity of the techniques are also important. Therefore, this paper will use some simple SNA indicators and algorithms and interpret them in detail through the case of the Chinese co-authorship network to make it possible and easier for higher education institutions to use SNA methods. The above reasons are also the main differences between our research and other similar studies. In addition, from the perspective of methods, this paper refers to the research of Abbasi et al.'s (2011, 2012) in terms of the link analysis methods and the Spearman correlation test. However, we will identify the authority and hub and include them as the Spearman correlation test variables.

In this study, the academic performance of the scholars or researchers will be measured by the g index. Then the link analysis and graph cluster analysis will be employed to study the co-author network. At last, to analyze the relationship between the academic performance of the authors and their co-authorship networks, such as the centrality, efficiency of the egocentric network, authorities and hubs, a Spearman correlation test will be used in this research. The scholars or researchers in China will be selected to be the research target of this paper in that although there are a number of existing papers that related to China's co-authorship network, they mainly concentrate on a province or a nation's level instead of a city or a university's and the literature that relates to the internal structure of the network are quite little. (Andersson et al., 2014). Additionally, bibliographic data of the targeted authors from Shanghai's three top higher education institutions, including Shanghai Jiaotong University, Fudan University and Tongji University, have been collected and analyzed in the research. The main reason is that they are the members of the 985UNIs and the top three universities in one city, Shanghai. The members of 985UNIs stand at the peak of the pyramid in the higher education system of China (Wang et al., 2014). The collaboration between the members is quite close and the data is adequate for the research.

This paper has been organized as follows. Section 2 briefly introduces the literature related to the social network analysis and the co-author network as well as the methods this paper intends to use. Section 3 provides an introduction to the data and the results. Section 4 provides an analysis of the results and Section 5 concludes the paper.

2 LITERATURE REVIEW AND PROPOSED METHODS

2.1 Literature Review

A social network is a group of nodes or participants linked to each other through individual relationships, such as relatives, friendship, cooperation between companies, etc. (Chung et al., 2005). In general, social networks contain two types, namely, socio-centric networks and egocentric networks.
In recent years, with the development of the Internet, more and more devices are connected to the Internet, and new related services and applications are emerging as well. The increasing amount of data has promoted the development of 5th generation mobile networks (5G) and further promoted the development of emerging information technologies, such as Big Data Analytics and IoT. With the support of these technologies, many different things, such as industrial equipment, mobile phones, laptops, etc., are connected to each other on the Internet. A large amount of various data and information are also generated. Recent estimates indicate that one in three people, on average, has over two devices linked to the Internet (Amadeo et al., 2016). The data transmitted through the Internet is about 40 exabytes Every day (Stephens et al., 2015). Moreover, big data analysis techniques can be used to collect, organize, and analyze these data to extract valuable information behind these data.

Sharing information has been vital to human survival all the time. People often choose to share their first-hand information or experiences of common interest with others on social networks, which has resulted in an in-depth exploration in the field of human social behavior. Nowadays, the Internet connection has been ubiquitous in our lives, making it easier for people to understand the world around them anytime, anywhere (Jameel at el., 2018). Social network analysis (SNA) originated in the 1930s, and its exceptional method of analysis progressively established in studying the theory of the social capital, strong and weak relationship and structure hole (Hou et al., 2020). With the rapid development of the Internet, the method of SNA has been employed in various areas, such as information propagation, geography, management, and economics.

Costa (2020) provides a framework for the collaboration between the scholars in the area of SNA and the intelligence and law enforcement that against trading wildlife so that the synergies of their work can be leveraged. Costa's work suggests that with the support of SNA, wildlife trafficking can be investigated in several ways to analyze data sets to gain various valuable intelligence products, using SNA to construct an intelligence network so that investigators can apply intelligence systematically and so on.
Mora et al.'s (2018) provides a way to obtain data on popular sports areas and manage this knowledge to assist in urban planning decisions. Their study found that social networks of sports provide useful information on how urban infrastructures are used by the citizens, what actions need to be taken and where to take action. Through all these networks, valuable knowledge can be obtained from the users.

Kim and Hastak (2018) studied the application of social networks in disaster management. They investigated the patterns created by the interactions between Facebook users during disasters, which suggested a significant part of the social media in disseminating emergency information. Results show that individuals, organizations, and emergency agencies of social networks play a vital role in disaster management. Graham et al. (2015) investigated over 300 local government officials in the United States. They found that the relationship between the degree of using social media and the evaluation of the ability of local cities in crisis management is significant.

In 1999, SNA, as a method of collecting and analyzing interpersonal connection patterns among groups, was first introduced to the knowledge management field by Morten T. Hansen of Harvard University. The study results found that weak interunit ties between departments are conducive to discovering the existence of useful knowledge in other departments, but are not conducive to the transfer of complex knowledge between departments. Only the existence of strong connections can achieve the transfer of complex knowledge.

Co-author network analysis is a method of visually presenting the influence of research and can reveal researchers' collaboration patterns and actions at the level of individual, institution, or geography. In Liu et al. (2006) research, they analyzed the co-author network of the past joint ACM/IEEE, IEEE, and the ACM digital library conference to examine the digital library domain’s state. Meanwhile, Newman (2004) studied the pattern of scientific cooperation through a co-authorship network. Moreover, the research of Morel et al. (2009) indicated that the method of the co-author analysis is essential in supporting the strategic planning for neglected disease research. Although the history of the social network analysis has been very long, with the development of the emerging information technologies in recent years, including the big data, the Internet of Things, etc., the speed of the SNA's development has been accelerated.

The analysis of Fonseca et al. (2016) identified participants and their connections to reveal the social network structure. They reviewed the applications of co-author network analysis, the fundamental steps, and the common network concepts of co-authorship networks in the area of health. Reznik-Zellen et al. (2020) explored different aspects of the Journal of the Medical Library Association (JMLA) co-authorship from 2006 to 2017. Their research found that among the 1,351 authors of JMLA, 69% are not the members of the Medical Library Association (MLA) and the co-authorship between the MLA members are closer.

2.2 Proposed Methods

In this paper, the methods of SNA, in terms of the link analysis and graph cluster analysis, will be employed to construct and analyze the scholars’ co-author network.

Among a number of techniques or measures of SNA, this paper selects the link analysis and graph cluster analysis for the following reasons. Firstly, through the link analysis, we can spot the most productive scholars and understand their roles in the co-author network. Link analysis is used to identify relationships between different parties, such as linking authors to other authors or linking authors to papers. Link analysis is not only able to determine relationships but is also able to provide information on how to use other attributes to illustrate the type of link and its strength. In particular, link analysis is essential for learning the influence between the different entities. Secondly, the centrality indicators this paper selects, including, are the most fundamental in SNA so that it makes it easier or low-budget for the relevant educational institutions to adopt. Finally, the graph cluster analysis is adopted because this paper tries to learn about the cooperation pattern between authors. The cluster analysis can classify the authors that share the similarity into one group and present the differences between groups. Additionally, the results can be directly presented by graphs, which is intuitive.

After the network construction and two parts of SNA analysis, to obtain an understanding of the correlation between academic performance and the SNA measures, the Spearman correlation test will then be carried out. The most productive researcher can be determined. The link analysis methods and the Spearman correlation test this study plans to adopt draws on Abbasi et al.'s (2011, 2012) research. Nevertheless, this research is an extension of their research in identifying the authority and hub (Kleinberg, 1998) by employing the HITS algorithm. Additionally, two kinds of betweenness algorithms will be used in the graph cluster analysis. This paper is also an extension of our previous work (Xu and Chang, 2019) and includes the authority and hub as the Spearman correlation test variables. All the different types of analyses and algorithms this paper uses can better understand the microscopic of the co-author network.

In the link analysis, four centrality metrics will be measured, consisting of the centrality of the degree, closeness, betweenness and eigenvector. According to Freeman(1978), a participant or a node's centrality may significantly impact his or her satisfaction, leadership and efficiency. Moreover, a participant's performance may be influenced by his or her degree centrality and betweenness centrality in particular. According to Scott (1991), the degree centrality of a participant or a node refers to the number of his or her adjacent participants and is used to evaluate its local centrality. Betweenness measures the centrality by measuring the degree to which a given node stands on the shortest path between other nodes in the graph (Borgatti,1995). Closeness is another kind of centrality, used to calculate the distance from it to other nodes (Freeman, 1980). In addition, eigenvector centrality is an indicator used to calculate the centrality according to the idea that a node's centrality depends on the number of the adjacent nodes and relates to these adjacent nodes' centralities (Bonacich, 1972).

According to Burt's structural hole theory, to assess a scholar or a researcher's relationship with other scholars within one group, the efficiency of the nodes will also be measured in this paper (Borgatti,1995). Based on the theory, if a scholar often collaborates with many scholars from the same group, the information that the scholar obtains from these major collaborating scholars is likely to be redundant in that scholars in a group always share the same information. Hence, if the scholar is only closely related to one scholar in the group instead of all the scholars in this group, the scholar's network efficiency will be higher.

After the centrality analysis, the HITS algorithm (Kleinberg, 1998) will be adopted to determine the hub and the authority in the co-authorship network. If a participant or a node points to many other vertices, it is regarded as a hub; if a node or participant has many other nodes linked to the node, it is regarded as an authority.
This research adopted two algorithms that are established on two types of betweenness centrality in the graph cluster analysis. The two algorithms' outcomes will be compared with each other and the difference will be explained.

Finally, the Spearman correlation test will be carried out to evaluate the correlation between the social network metrics and the author's performance to identify the most productive researchers (Abbasi et al., 2011). The Spearman correlation test method is employed to analyze the significance of the correlation between two factors (Gauthier, 2001). The g index of each researcher will be calculated to quantify their academic performance. G index, one of the indicators used to measure the performance of a scholar or a researcher, is widely adopted by the academic database (Egghe, 2006). It is estimated based on the descending order of the number of citations of a researcher's paper. The g index is the maximum number of citations obtained by the top g papers that is not under g2.

Based on the literature above, the spearman correlation will test the following hypothesis:

H1: The degree centrality of a scholar or a researcher influences his or her academic performance;
H2: The betweenness centrality of a scholar or a researcher influences his or her academic performance;

H3: The closeness centrality of a scholar or a researcher influences his or her academic performance;
H4: The eigenvector centrality of a scholar or a researcher influences his or her academic performance;
H5: The efficiency of a scholar or a researcher influences his or her academic performance;
H6: The hub position of a scholar or a researcher influences his or her academic performance;
H7: The authority position of a scholar or a researcher influences his or her academic performance.

3 DATA AND IMPLEMENTATION

3.1 Data Collection and Management

In the research, bibliographic data of the targeted authors from Shanghai's three top higher education institutions, including Shanghai Jiaotong University (SJU), Fudan University (FU) as well as Tongji University (TU) was generated from the database of Scopus. The data generated in the research is in the field of social science from 2014 to 2016. The publication information of 166 scholars was obtained. The attributes included the names of the scholars, their affiliations, the number of citations per paper and the number of publications。

This paper established two data sets according to the public information available to the authors. One is the basic information of the target scholars, consisting of the scholars' initials for privacy concerns, the affiliation, number of publications, total citations of other authors, and their g index. The other data set is about the links between the scholars and indicated whether there is a co-author relationship between the scholars. Additionally, the number of collaborations has been assigned to the attribute "weight" because the entire data set cannot be fully displayed in the paper. Hence, Table 1 and Table 2 displayed a part of them.

Table 1: Authors

Author	Affiliation	Country	No. of Publication	Cited by	G-I
Wang L	Fudan University	China	15	60	7
Li C	Fudan University	China	3	10	3
Zhang Y-FD	Fudan University	China	4	28	4
Xu C	Shanghai Jiao Tong University	China	3	29	1
Zhou T	Shanghai Jiao Tong University	China	1	12	1
Ding Y	Shanghai Jiao Tong University	China	1	12	1
Jin M	Shanghai Jiao Tong University	China	1	12	1
Sun Y	Shanghai Jiao Tong University	China	1	12	1
Huang S	Shanghai Jiao Tong University	China	1	12	1
Xie Q	Shanghai Jiao Tong University	China	1	12	1

Table 2: Co-authorships

Author 1	Author 2	Weight
Chu N	Li D	2
Chu N	Gober H -J	2
Chu N	Qiu X	2
Harder M K	Burford G	2
Harder M K	Hoover E	2
Ye X -Y	Zhang H	2
Yu J -M	Wang J -W	2
Yu J -M	Yuan Z -P	2
Yuan Z -P	Wang J -W	2

3.2 Implementation

After the data collection and management for SNA, in order to perform the subsequent analysis, Rstudio and Ucinet (Borgatti et al., 2002) were adopted in this part as tools to visualize the co-author network and measure the network metrics.

First of all, the scholars' co-author network was established and presented as a graph. The nodes in the graph are the scholars. The links between the scholars represent the collaborative relationships between the scholars and each link's width stands for the weight of the link. The weight refers to the number of times that two scholars have collaborated to publish papers.

The following figure is the co-authorship network built in this research (Figure 1).

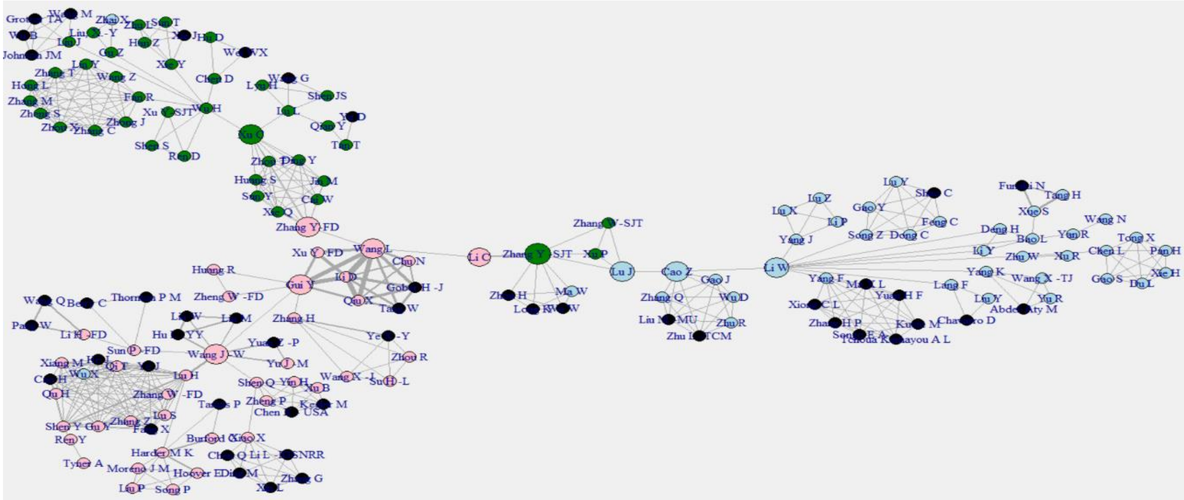

Figure 1: Co-authorship network

As is shown in Figure 1, the green nodes of the network were assigned to the scholars of SJU, the light blue nodes were assigned to the scholars of TU, the pink nodes were assigned to the scholars of FU and black nodes represented the authors of other universities.

Figure 1 indicated that the co-authorship network could be separated into three categories based on the university, namely, the category on the upper left, the category on the lower left, and the right category. Firstly, the lower left (pink group) category representing the FU seems to have more cooperative relations with the external organizations. Moreover, larger nodes seem to play a more significant role in forming the network.

Secondly, after building the co-author network, the link analysis was carried out. Four metrics that measure the centrality of the nodes were calculated, consisting of normalized centrality of the degree, betweenness, closeness and eigenvector. Moreover, the efficiency of each node in the structural holes was measured as well. The results of the measures were sorted out and Table 3 shows some of them as follows:

Table 3: MEASURES

Author	Betweenness	Closeness	Degree	Eigenvector	Efficiency
Sun T	0	0.0914	0.0242	2.22E-07	0.25
Xie J	0	0.0914	0.0242	2.22E-07	0.25
Zhu L	0	0.0914	0.0242	2.22E-07	0.25
Han Z	0	0.0914	0.0242	2.22E-07	0.25
Wei WX	0	0.0912	0.0121	2.01E-07	0.5
Chen D	0.02410	0.1002	0.0182	4.59E-06	0.778
Xie Y	0.04760	0.1004	0.0303	4.61E-06	0.52

Furthermore, in order to determine the hub and authority position of each scholar in the co-authorship network, the HITS algorithm was employed. The algorithm returned two vector columns with the value that indicated their hub and authority position. As the attribute, hub and authority are always connected to each other, this study separated them so that the data can be explained clearer and Table 4 shows some of them as follows.

Table 4: Authority and Hub

Author	Authority	Author	Hub
Sun T	1.75E-09	Sun T	1.75E-09
Xie J	1.75E-09	Xie J	1.75E-09
Zhu L	1.75E-09	Zhu L	1.75E-09
Han Z	1.75E-09	Han Z	1.75E-09
Wei WX	1.44E-09	Wei WX	1.44E-09
Chen D	1.72E-08	Chen D	1.72E-08
Xie Y	1.76E-08	Xie Y	1.76E-08
Hu D	1.44E-09	Hu D	1.44E-09

Finally, this paper conducted graph cluster analysis. The algorithm used in this part for clustering is established on the betweenness centrality. In the co-author network, the effective dissemination of knowledge or academic information is of great importance. Therefore, it is crucial to identify the nodes or links that act as brokers in the network. Thus, the betweenness algorithm was chosen as it helps identify the critical nodes or edges. The clustering algorithm can employ two kinds of betweenness centrality, vertex betweenness and edge betweenness. In the graph cluster analysis, both types of betweenness centrality were used and there are slight differences between the results.

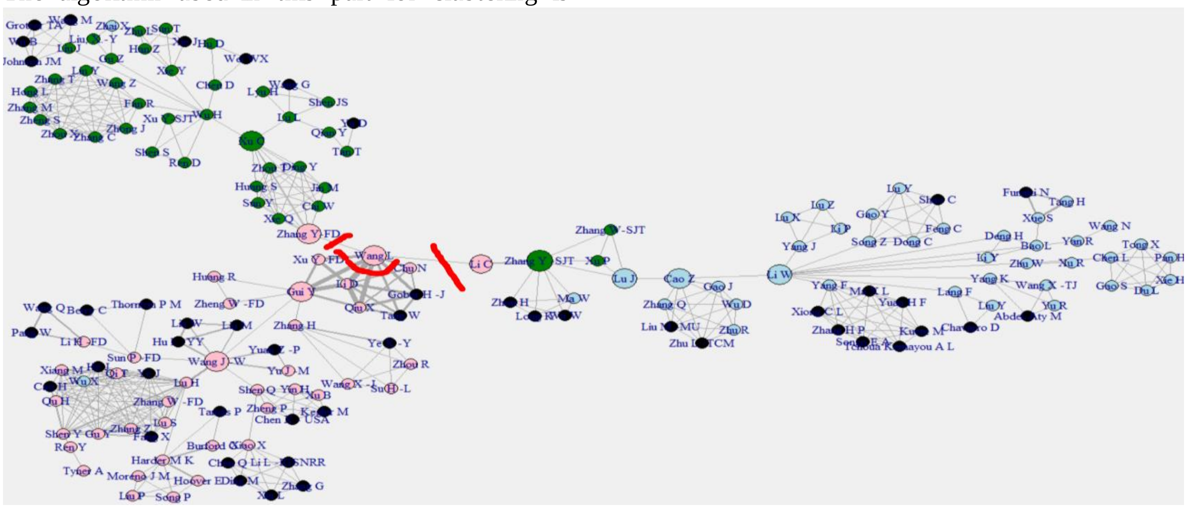

Figure 2: Co-authorship network

4 ANALYSIS OF RESULTS AND DISCUSSION

Results of the social network measures in the link analysis and the clustering result in the graph cluster analysis will be analyzed in this part. This paper will then test the significant relationship between the seven SNA metrics and the g-index that indicates the scholars' academic performance by carrying out a Spearman rank correlation test.

4.1 Link Analysis

The degree is the easiest way to measure the centrality of a node among the four centrality measures. In this network of scholars, the average value of the degree centrality is about 0.03, while the degree centrality of Lu H from FU was the highest of 0.09. The highest value means that they have more active interactions with others than the other scholars or are more popular among scholars.

Closeness is used to evaluate the global centrality of a node by measuring the distance from the node to other nodes. Among the scholars, the average value of the closeness centrality is about 0.08. The closeness centrality of Wang L from FU was the highest of around 0.14, which means that his or her position on the network is closest to all other scholars on average. Besides, this kind of position made him or her the person who can gain information most efficiently.

Betweenness refers to the number of times that a specific node is located on the shortest path between other nodes in the graph. In the co-author network, the average value of the betweenness was about 0.04, while the highest value of betweenness centrality was gained by Wang L from FU, which was around 0.67. Having the highest value in the network means that his or her part as a gatekeeper or a broker is essential. He or she can control the dissemination of information among the scholars most frequently. As shown in Figure 2, if Wang L was removed from the network, the three universities' single co-author networks would disconnect.

Eigenvector centrality is established on the concept that a node's centrality depends on the number of its adjacent nodes as well as relates to the adjacent nodes' centralities. Among the researchers, the average value of the eigenvector is round 0.3, while the highest value was obtained by Wang L from FU, which is 1.0. There were nine scholars adjacent to him or her, and greater than half of the adjacent scholars also have great centrality values.

From the perspective of structural holes, efficiency refers to the ratio of the total number of disjoint groups of a node's main node divided by that node's centrality value. In the co-author network, the average value of efficiency was about 0.4, and the top 10's efficiencies were over 0.9. A higher value indicates that these scholars spend more effort building a close relationship with just one scholar in a group with connected scholars, instead of with all the scholars in this group. Additionally, they are often able to access knowledge or academic information in many different fields. This may help them make innovation and achieve better performance than other companies.

Table5: Efficiency-Top 10

Author	Affiliation	Efficiency
Li C	Fudan University	1
Ren Y	Fudan University	1
Tyner A	Fudan University	1
Zhu W	Tongji University	1
Yan R	Tongji University	1
Xu R	Tongji University	1
Wang N	Tongji University	1
Li W	Tongji University	0.94
Wu H	Shanghai Jiao Tong University	0.926
Wang J -W	Fudan University	0.901

The lowest value of efficiency is about 0.08, which means that they may have a close relationship with several scholars in a group, rather than with the single scholar in this group. Since scholars in the same group will always share knowledge or information with each other, it will be a waste of time to maintain co-authorships with all the scholars in the same group because they may duplicate knowledge or information from these scholars.

According to the hub and authority results, Lu H has obtained the highest value in both perspectives, so he or she was considered the authority and hub. The authority position of Lu H in this co-authorship network means that in the field of social science, he or she was regarded as authoritative and productive. Meanwhile, he or she was in contact with some other scholars who can be regarded as authoritative as well. For this reason, Lu H was considered as a hub at the same time.

While taking all these factors into consideration, a conclusion can be drawn for the part of link analysis. Although Wang Li was not determined to be the co-authorship network's authority, he or she seems to be the most significant scholar comparatively. The values of his or her closeness centrality, betweenness centrality and eigenvector centrality were the greatest. Meanwhile, his or her values of the degree centrality and efficiency were also great. For the degree centrality, he or she gained a value of about 0.06 and the average value was about 0.03; For efficiency, he or she gained a value of about 0.68 and the average value was about 0.4. In general, Wang L played the most significant role in the co-author network to interact with other scholars as well as to obtain and disseminate academic information more effectively and efficiently.

4.2 Graph Cluster Analysis

In this section, two kinds of algorithms established on the betweenness centrality were adopted to carry out the graph cluster analysis, namely, vertex betweenness and edge betweenness. The results of the algorithm established on the vertex betweenness are shown in Figure 3 and the results of the algorithm established on the edge betweenness are shown in Figure 4.

The two algorithms' results separated the co-author network into three groups, similar to each other. The scholars in the same institutions were grouped in the same cluster.
Nevertheless, some differences still existed between the two algorithms' results. According to the vertex betweenness the algorithm result, Wang L did not completely belong to anyone cluster but shared by all

three clusters. While using the algorithm of edge betweenness, he or she was grouped into cluster Green. In addition, instead of sharing Wang L, the link between cluster Green and cluster Blue can be disconnected by cutting the link between Gui Y and Wang JW as is shown in Figure 4. This means that every author connected to Gui Y, except Wang JW, belongs to cluster Green, not cluster Blue in Figure 3.

In general, the clustering analysis was performed according to the betweenness centrality. Betweenness centrality measures the number of times a given node is located on the shortest path between other nodes. The result of the analysis shows that although there were some cases of scholars collaborating with other institutions outside their university, and compared with the co-authorship outside the university, the cooperation within the university is much closer.

In order to determine the hub and authority of each cluster, the HITs algorithm employed in the link analysis was adopted again after carrying out the vertex betweenness algorithm. The result shows that, among the scholars, TU's authority and hub are Yang F, SJU's is Fan R, and FU's is Lu H.

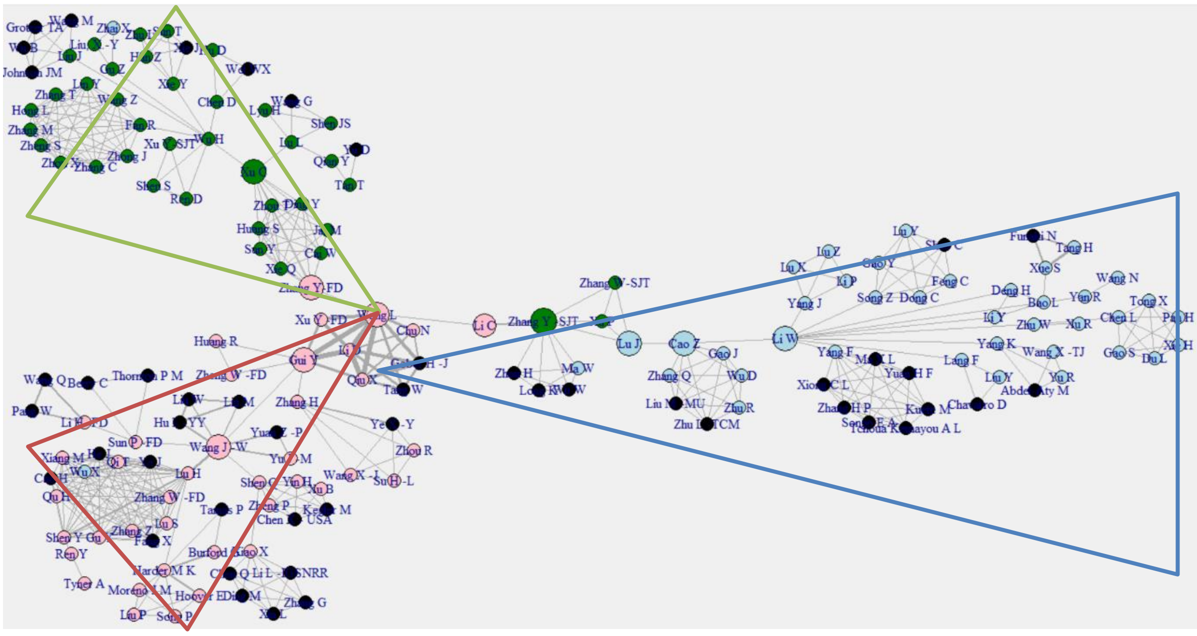

Figure 3: Co-author network – Vertex

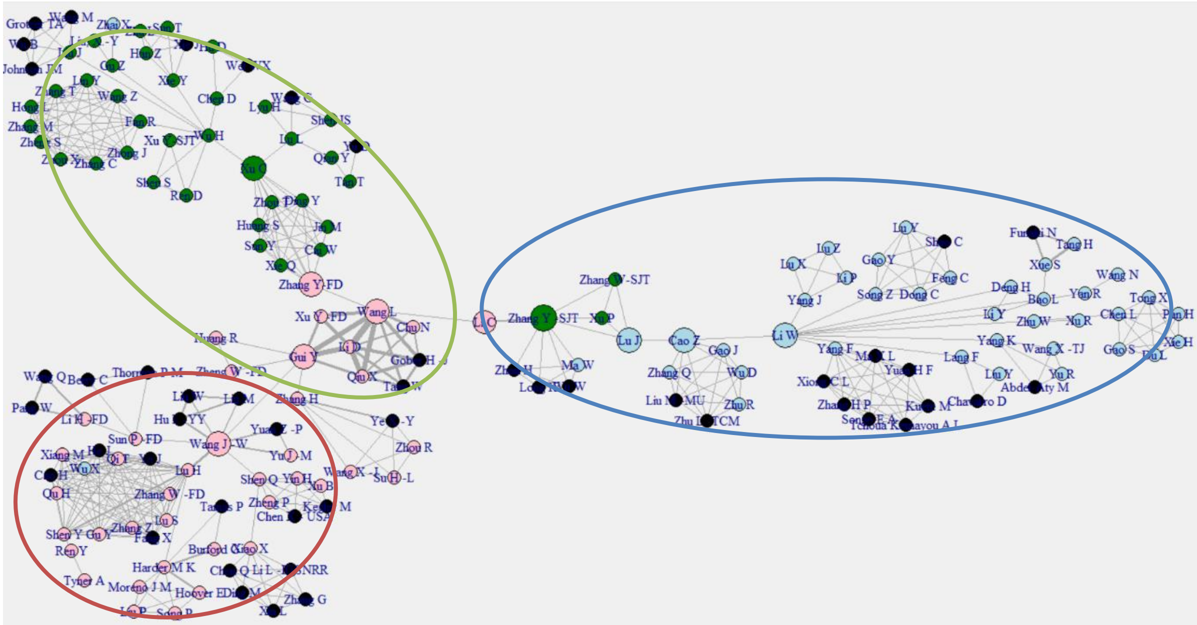

Figure 4: Co-author network – Edge

4.3 Spearman rank correlation test

Finally, the Spearman rank correlation test was carried out to test the seven hypotheses this paper proposed and study the relationships between the social network metrics and academic performance. Table 6 shows the results.

In this table, the variables were tested at the significance level of 0.01 and 0.05. '**' means the correlation is significant at the level of 0.01 and '*' means that the correlation is significant at the level of 0.05. If the value of the significance level is more than 0.01, the assumption is not accepted and the relationship is not of significance; if the value of the significance level is less than 0.01, the assumption is accepted and the relationship is significant (Hochberg & Benjamini, 1990). As shown in Table 6, the correlations between the centrality of betweenness, hub, eigenvector, efficiency, authority and g-index are all significant because the value of their significance level individually is far less than to 0.05. However, the variance of g index cannot be explained by the variance of closeness centrality or degree centrality very well because the values of their significance are much greater than 0.01, about 0.95 and 0.84, respectively. Therefore, among the hypotheses this paper proposed, H1, H3 was rejected, and H2, H4, H5, H6 and H7 were accepted.

The coefficients of the measures indicate that the researchers or scholars having greater centrality of eigenvector and betweenness, higher efficiency, authority and hub are likely to get a greater g index, which means that the productivity of the researchers is higher. A scholar or a researcher with a greater betweenness centrality means that he or she is more resourceful, that is, they collaborate more frequently with other authors, and he or she can control the knowledge transfer between researchers most frequently. A scholar or a researcher with a greater eigenvector centrality means that they have more connections with other scholars, who also have various good connections. A scholar or a researcher with greater efficiency indicates that he or she collaborates more with a variety of scholars or researchers from different groups than with all scholars within the same group. Additionally, if a scholar is identified to be the authority, indicating that he or she is quite authoritative or productive in the co-author network. Moreover, a researcher is considered the hub means that various scholars who are authorities can be contacted through him or her.

Based on the analysis of the Spearman correlation test results as above, this research can conclude that the academic performance of the scholars or researchers, who collaborate with different scholars or groups and or other scholars who also have good connections, is better than those who do not. Furthermore, scholars or researchers who interact with only one scholar of a group closely rather than all scholars within the same group have better performance than the scholars who do not have. Meanwhile, the scholars who are considered as the authorities or hubs have greater academic performance as well.

Table 6: Spearman rank correlation test

		Bet.	Clo.	Hub	Deg.	Eig.	Eff.	Aut.	G-I
Bet.	Correlation Coefficient	1.000	0.123	.302**	.311**	.348**	.387**	.302**	.456**
	Significance value		.115	.000	.000	.000	.000	.000	.000
Clo.	Correlation Coefficient	.123	1.000	-.192*	0.109	0.083	0.120	-.191*	0.005
	Significance value	.115		.013	.161	.290	.123	.014	.952
Hub	Correlation Coefficient	.302**	-.192*	1.000	.252**	.869**	-.116	1.000**	.190*
	Significance value	.000	.013	.	.001	.000	.137	.000	.014
Deg.	Correlation Coefficient	0.311**	0.109	.252**	1.000	.249**	-.632**	.252**	0.015
	Significance value	.000	.161	.001	.	.001	.000	.001	.843
Eig.	Correlation Coefficient	.348**	.083	.869**	.249**	1.000	-0.019	.869**	.284**
	Significance value	.000	.290	.000	.001	.	.808	.000	.000
Eff.	Correlation Coefficient	0.387**	0.120	-0.116	-.632**	-0.019	1.000	-0.116	.374**
	Significance value	.000	.123	.137	.000	.808	.	.137	.000
Aut.	Correlation Coefficient	0.302**	-.191*	1.000**	.252**	.869**	-0.116	1.000	.190*
	Significance value	.000	.014	.000	.001	.000	.137	.	.014
G-I	Correlation Coefficient	0.456**	0.005	.190*	0.015	.284**	.374**	.190*	1.000
	Significance value	.000	.952	.014	.843	.000	.000	.014	.

**. Correlation is significant if the significance value is less than 0.01.
*. Correlation is significant if the significance value is less than 0.05.

5 CONCLUSION AND CONTRIBUTION

5.1 Conclusion

In this research, through the use of link analysis and the graph cluster analysis, a co-author network of 166 scholars was established and analyzed. The majority of the scholars are from three top higher education in Shanghai, China. In the link analysis, five SNA metrics were calculated, including the centrality of degree, betweenness, closeness, eigenvector and efficiency. In addition, the HITS and betweenness clustering algorithms were also carried out. The analysis results show that the most significant scholar in the co-author network is Wang Li. Finally, this paper used a Spearman correlation test to assess the relationship between the academic performance of the scholars and SNA metrics so that productive researchers can be identified. These test results show that betweenness centrality, eigenvector centrality, authority and hub position, and efficiency are significant to g-index.

5.2 Implications

The analysis results of this research provide a reference for relevant organizations and researchers. According to this research, scholars' co-authorship networks are closely related to their academic performance for the relevant organizations, consisting of the Ministry of Education or higher education institutions. Moreover, it may help them determine, cluster, and configure productive and efficient scholars by constructing and evaluating the co-authorship networks so that the research synergy can be optimized. For the scholars or the researcher, this research suggests that they can establish closer collaboration relationships with different scholars, rather than cooperate with only one scholar. Additionally, they should try to avoid developing co-authorships with a number of different scholars within the same group, which may result in inefficiency and collaborate with resourceful scholars.

With the rapid development of emerging information technologies like the Internet of Things, everything can be linked to each other to form various types of networks. Therefore, SNA can be used to study the relationship between people and multiple fields. Just as this research helps higher education institutions gain research synergies by identifying the most influential and most productive scholars through studying co-author relationships, such research methods can also be applied to other fields.

For example, in Industry 5.0, SNA can be used to study the interaction of various equipment types to identify the strong and weak links in the production network of a company, to help companies improve production efficiency and reduce costs. As manufacturers employ more and more IoT devices, utilizing and managing them in an efficient and optimized way has become the main challenge. Therefore, it is critical for them to identify an effective method. By extending the typical notions and modes of the social network to the networks of Things, the strong and weak links in the production networks can be easily identified. Use some of the SNA indicators in this paper to illustrate, devices with a high degree centrality means they are the central devices in their production networks. Devices with high closeness centrality refer to they can receive production information from others most efficiently; devices with high betweenness centrality mean the devices play a broker in their network. If they are removed, the production networks would be disconnected and the operations would stop functioning. There are a number of SNA measures or algorithms that can be applied to the network of Things. By integrating the method of SNA into IoT, the applications of IoT will be more powerful.

5.3 Contribution and future work

Among the existing literature, most papers that studied the Chinese co-author networks mainly concentrated on the country or province-level instead of cities or universities. There are few papers on the micro co-author networks. Therefore, this research outlines the internal structure of Chinese co-author networks. Although the co-author network established in this research may not be large, it provides directions for the possible research in the future. Our social network approach could provide an alternative but complementary solution to IoT. By identifying the strengths and influences in the research networks, we could strengthen the impacts and collaboration to maximize the efficiency and outcomes. In our future research, we can include data from all the higher education institutions of Shanghai or other cities in China or the world to construct a complete co-author network on the city level and understand the research impacts locally and internationally. Furthermore, more metrics can be included to evaluate their correlations with the academic performance of scholars.

Acknowledgment

This work is partly supported by VC Research (VCR 0000085).

REFERENCES

Abbasi, A., Altmann, J., & Hossain, L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics*, *5*(4), 594–607. https://doi.org/10.1016/j.joi.2011.05.007

Abbasi, A., Chung, K. S. K., & Hossain, L. (2012). Egocentric analysis of co-authorship network structure, position and performance. *Information Processing & Management*, *48*(4), 671–679. https://doi.org/10.1016/j.ipm.2011.09.001

Amadeo, M., Campolo, C., Quevedo, J., Corujo, D., Molinaro, A., Iera, A., Aguiar, R. L., Vasilakos, A. V. (2016). *Information-centric networking for the Internet of things: challenges and opportunities. IEEE Network, 30(2), 92–100.* doi:10.1109/mnet.2016.7437030

Andersson, D. E., Gunessee, S., Matthiessen, C. W., & Find, S. (2014). *The Geography of Chinese Science*. 22.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, *2*(1), 113–120.

Borgatti, S. P. (1995). Centrality and AIDS. *Connections*, *18*(1), 112–114.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet for Windows: Software for social network analysis*.

Chung, K. K. S., Hossain, L., & Davis, J. (2005). *Exploring Sociocentric and Egocentric Approaches for Social Network Analysis*. 9.

Costa, J. (2020). Social network analysis in the fight against illegal wildlife trade. Available at: https://www.baselgovernance.org/publications/social-network-analysis-fight-against-illegal-wildlife-trade. (Accessed July 13rd 2020)

Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, *69*(1), 131–152. https://doi.org/10.1007/s11192-006-0144-7

Fonseca, B. de P. F. e, Sampaio, R. B., Fonseca, M. V. de A., & Zicker, F. (2016). *Co-authorship network analysis in health research: method and potential use. Health Research Policy and Systems, 14(1)*. doi:10.1186/s12961-016-0104-5

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. https://doi.org/10.1016/0378-8733(78)90021-7

Freeman, L. C. (1980). The gatekeeper, pair-dependency and structural centrality. *Quality and Quantity*, *14*(4), 585–592.

Gauthier, T. (2001). Detecting Trends Using Spearman's Rank Correlation Coefficient. *Environmental Forensics*, *2*(4), 359–362. https://doi.org/10.1006/enfo.2001.0061

Graham, M. W., Avery, E. J., & Park, S. (2015). The role of social media in local government crisis communications. Public Relations Review, 41(3), 386–394. http://dx.doi.org/10.1016/j.pubrev.2015.02.001.

Heathfield, S. M. (2020). The 5 Goals of Employee Performance Evaluation. available from: https://www.thebalancecareers.com/employee-performance-evaluation-goals-1918866 (Accessed July 13rd 2020)

Hochberg, Y., & Benjamini, Y. (1990). More powerful procedures for multiple significance testing. *Statistics in Medicine*, *9*(7), 811–818. https://doi.org/10.1002/sim.4780090710

Hou, M. L., Sun, G. J., Dong, Z. J. (2020) A summary of the application of social network analysis. Industry and Technology Forum. 2020 (5), p.1.

Jameel, F., Wyne, S., Jayakody, D.N.K., Kaddoum, G., O'Kennedy, R. (2018). Wireless Social Networks: A Survey of Recent Advances, Applications and Challenges. IEEE Access 2018,6, 59589–59617. doi: 10.1109/ACCESS.2018.2872686.

Kim, J., & Hastak, M. (2018). *Social network analysis: Characteristics of online social networks after a disaster. International Journal of Information Management, 38(1), 86–96*. doi:10.1016/j.ijinfomgt.2017.08.003

Kleinberg, J. M. (1998). Authoritative sources in a hyperlinked environment. *In Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*.

Liu, X. , Bollen, J. , Nelson, M. , & Van, d. S. H. . (2005). Co-authorship networks in the digital library research community. *Information Processing & Management*, 41(6), p. 1462-1480.

Martínez, A., Dimitriadis, Y., Rubia, B., Gómez, E., & de la Fuente, P. (2003). Combining qualitative evaluation and social network analysis for the study of classroom social interactions. *Computers & Education*, *41*(4), 353–368. https://doi.org/10.1016/j.compedu.2003.06.001

Mora, H., Pérez-delHoyo, R., Paredes-Pérez, J., & Mollá-Sirvent, R. (2018). *Analysis of Social Networking Service Data for Smart Urban Planning. Sustainability, 10(12), 4732.* doi:10.3390/su10124732

Morel, C. M., Serruya, S. J., Penna, G. O., & Guimarães, R. (2009). Co-authorship Network Analysis: A Powerful Tool for Strategic Planning of Research, Development and Capacity Building Programs on Neglected Diseases. *PLOS Neglected Tropical Diseases*, *3*(8), e501. https://doi.org/10.1371/journal.pntd.0000501

Newman, M. E. J. (2004). Co-authorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, *101*(Supplement 1), 5200–5205. https://doi.org/10.1073/pnas.0307545100

Prell, C., Hubacek, K., & Reed, M. (2009). Stakeholder Analysis and Social Network Analysis in Natural Resource Management. *Society & Natural Resources*, *22*(6), 501–518. https://doi.org/10.1080/08941920802199202

Reznik-Zellen, Rebecca. (2020). Building visualization skills through investigating the Journal of the Medical Library Association co-authorship network from 2006–2017. J Med Libr Assoc. 108(2), pp. 229–241. doi: 10.5195/jmla.2020.775

Scott, J. (1991). *Social Network Analysis: A handbook Sage Publications London UK*.

Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., Ravishankar K. Iyer, Schatz, M.C., Sinha, S., Robinson, G. E. (2015). *Big Data: Astronomical or Genomical? PLOS Biology, 13(7), e1002195.* doi:10.1371/journal.pbio.1002195

Ter Wal, A. L. J., & Boschma, R. A. (2009). Applying social network analysis in economic geography: Framing some key analytic issues. *The Annals of Regional Science*, *43*(3), 739–756. https://doi.org/10.1007/s00168-008-0258-3

Wang, W., Wu, Y., & Pan, Y. (2014). *An investigation of collaborations between top Chinese universities: A new quantitative approach*. 11.