

# Research on Social Network Analysis Method in Cooperative Innovation Performance

Sai Zhao

International Business School  
Yunnan University of Finance and Economics  
Kunming, China  
739457672@qq.com

Dongping Yu\*

International Business School  
Yunnan University of Finance and Economics  
Kunming, China  
247096657@qq.com

Jingfei Wang

International Business School  
Yunnan University of Finance and Economics  
Kunming, China  
864639017@qq.com

Zhengyang Meng

International Business School  
Yunnan University of Finance and Economics  
Kunming, China  
386818100@qq.com

**Abstract**—This study analyzes the development status and trends of innovation network, introduces the social network analysis methods commonly used in research of innovation network, and studies the network structure in cooperative network with the research results of cooperation and innovation, enriches the related research in the field of intelligence. Using the method of social network analysis, from the perspectives of centrality analysis, coherent subgroup analysis and core-edge structure analysis, the 2016 China Pharmaceutical Industry Patent Cooperation Data was selected as the target, and a cooperative network was constructed for empirical analysis. Through empirical research, it evaluates the status of innovation entities in innovation network in innovation output, and finds cooperative groups in the cooperative network and identifies the core enterprises and the number of innovation networks.

**Keywords:** *social network analysis, cooperative innovation performance, cooperation network*

## I. INTRODUCTION

Social network analysis is a research method of sociology. Sociological theory holds that society is composed of the relationship between network society and network nodes. The social network analysis method is to analyze the network structure and attribute characteristics by analyzing the relationship within the network. The attribute characteristics of the network include individual attributes and overall attributes in the network. The network individual attribute analysis includes, point centrality, betweenness centrality and closeness centrality, the overall attribute analysis of the network includes small world effects, small group studies, and cohesive subgroups. Zhu Qinghua (2008) argues that 'social network' refers to a collection of connections between multiple points (social actors) and points (the relationship between actors), That is, 'social network' are a collection of social actors and their relationship choices.[1]. 'Points' in social network refer to various social actors, while 'edges' in social networks refer to

various social relationships between actors. The social relations between actors can take many forms, such as friendship between people, relationship between superiors and subordinates, scientific research cooperation, exchanges between members of the organization, and trade relations between countries. Social network analysis is a specific tool for quantitatively studying the relationship between participants in social network. Its purpose is mainly to make social network theory more specific.

In social network analysis, UCINET is the most classic and famous software for relational network visualization. UCINET is a powerful social network analysis software, including a large number of network analysis indicators (centrality, two-way relationship cohesion measure, position analysis algorithm, faction analysis, random two-party relationship model and procedures for testing the network, etc.) and How far statistical analysis tools (multidimensional scales, correspondence analysis, factor analysis, cluster analysis, multiple regression for matrix data, etc.). The social network analysis method integrates four application software programs including NetDraw, Mage, Pajek, and Free. Because UCINET software can read text files and read files in KrackPlot, Pajek, Negopy, Vna and other formats, and can handle up to 32767 network nodes, this paper selects UCINET application software program to process data.

Social network analysis is a unique research method that is different from other qualitative and quantitative research methods for researching innovation network. Social network analysis focuses on the relationship data between social actors, not attribute data or conceptual data. Granovetter (1973) argues that the data that traditional statistical theory-based quantitative research faces are attribute data, and most quantitative studies require the independence of sample individuals[2]. Kuhn (1970) argues that qualitative research describes the meaning, motivation, definition, and typology of actor behavior[3]. The focus of his research is on the analysis of actor conceptual data,

---

\*Corresponding author

but such research often exaggerates the subjective initiative of actors and is often too social. Kilduff et al. (2003) summarize three distinct characteristics of social network research and traditional social science research methods, the focus of social network research is the relationship and relationship model, not the attributes or concepts of actors[4]. Social network research can be analyzed from multiple levels to establish a social network research model that is different from the traditional social science research methods. Social network research can integrate quantitative data, qualitative data and graphical data, which not only compensates for the characteristics of overly abstract quantitative data, but also avoids the problem of insufficient expansion caused by overly specific qualitative data. Uzzi (1997) believes that in the social network structure we are in, individual social behavior is influenced by both the social network structure and individual characteristics. The best way to study social action is social network analysis[5].

Cooperative innovation performance is a commonly used research indicator in innovation performance, and its connotation is basically subject to innovation performance. Regarding the connotation of innovation performance, domestic and foreign scholars (Griliches, 1979; Romer, 1986; Campbell and Palmer, 1997) have no unified opinions[6][7][8]. It is not difficult to find out the definition of innovation performance in academic circles by combing the existing literature at home and abroad. Most of them start from the process of innovation (Janssen and Van Yperen, 2004) [9]and the multi-factor integration perspective including the innovation process and output results (Han Yi et al., 2007)[10].

Combining the predecessors' point of view, Drucker (1997) considers innovation performance as an evaluation of innovation activities, which measures the innovation output results of an innovation activity, and further points out that the factors affecting innovation performance are diverse rather than A single element or indicator[11]. When some scholars define the concept of innovation performance, they effectively pay attention to the measurement of performance, emphasizing that innovation performance should be the corresponding output brought by "innovation" (Berg, 2006). With the deepening of research on innovation performance, the research has been extended to the two levels of management and economy. The innovation model combining industry and education or production, education and research has attracted the interest of scholars at home and abroad[12]. Productivity (Uwe et al., 2010)[13], innovation capabilities, and market competitiveness (Phelps, 2010)[14], market structure adjustment (Su, Yu-shan, etc., 2009) [15]were investigated and analyzed as factors of innovation performance.

## II. RESEARCH METHODS AND THEORETICAL BASIS

### A. Variable selection

Social network analysis methods are relatively mature, and social network can be constructed from many different perspectives, including central analysis, coherent subgroup analysis, core-edge structure analysis, and structural equivalence analysis. This paper analyzes the constructed cooperative network from the perspectives of centrality

analysis, coherent subgroup analysis and core-edge structure analysis.

Liu Jun (2014) believes that 'centrality' is one of the focuses of social network analysis. The power or central position of individuals or organizations in social network is one of the earliest discussions of social network analysts. Individual centrality measures how much a person is in a social network, and the importance of the network reflects the importance of this in the network. Therefore, social actors and individuals in social networks have as much centrality as individuals[16]. In addition to calculating the centrality of individuals in social networks, social network analysis can also calculate the centrality of the entire network. Liu Jun (2014) believes that the network centrality describes the degree of difference of each point in the whole network. According to different calculation methods, the centrality can be divided into three types: point centrality, betweenness centrality and closeness centrality [16].

#### 1) Point centrality

In social networks, there is a direct link between social actors who are central and have greater 'power' in the network and other actors. Point degree is the most direct metric to describe the centrality of nodes in social network analysis. The degree of centrality can be measured by the number of points that are connected to the network at that point. The higher the degree of node, the higher the centrality of the node, and the greater the importance of the node in the social network structure. The point centrality measures how close a node is to all other nodes in a social network, that is, how many connections are connected to a node in a social network. For a network graph with  $g$  nodes, the point center of node  $i$  is the total number of direct connections between  $i$  and other  $g-1$  nodes, which is expressed as follows,

$$C_D(Ni) = \sum_{j=1}^g (i \neq j) \quad (1)$$

In formula (1),  $CD(Nj)$  represents the point degree of node  $i$ , and  $\sum_{j=1}^g (i \neq j)$  is used to calculate the total number of directly connected lines between node  $i$  and other  $g-1$  nodes  $j(i \neq j)$ . The calculation of  $CD(Nj)$  is for simply centrality the cell values of the row or column corresponding to node  $i$  in the network matrix. The degree of centrality of such measurements depends not only on the relevance of each node to other nodes, but also on the size of the network size ( $g$ ). Stanley Wasserman and Katherine Faust (1994) conducted an in-depth study of the centrality of the point and finally calculated the standardized measurement formula to eliminate the impact of network size changes on the centrality of the point. The formula is as follows,

$$C'_D(Ni) = \frac{C_D(Ni)}{g-1} \quad (2)$$

In the standardized point-to-center measurement formula, the ratio of the point center of node  $i$  is divided by the maximum number of other  $g-1$  nodes that can be connected, and the proportion of network nodes directly related to node  $i$  is obtained. The proportion of network nodes ranges from 0.0 to 1.0. On any node in the social network structure, 0.0 means no connection, and 1.0 means that each node is directly connected. In social network, the centrality of standardized social actors

measures the extent to which actors participate in relationships, with the highest scored actors being the most visible. If the value of the centrality of the standardized social actors is closer to 1.0, then the degree of participation of the actors in the social network will be higher.

2) *Betweenness centrality*

In a social network, if a social actor is in a path between multiple points, that is, the social actor has the ability to control the interaction between the other two actors, the social actor is considered to be in an important location. According to the view, the researchers describe the individual centrality of social actors and measure the degree of interaction between social actors at multiple points. Luo Jiade (2005) believes that the more social actors occupy the social network, the higher the centrality will be, the more social actors are connected through the network[17]. We use  $g_{jk}$  to represent the number of direct connections between point  $j$  and point  $k$ . If there is a third point  $i$  in the middle of the line connecting point  $j$  and point  $k$ , the number of shortcuts in the line connected by the three points is represented by  $g_{jk}(i)$ . Use  $b_{jk}(i)$  to represent the degree of control of the ability of the third point  $i$  to interact between point  $j$  and point  $k$ . This degree of control is expressed as the probability of the number of shortcuts on the shortcut, ie  $b_{jk}(i) = g_{jk}(i) / g_{jk}$ .

The point  $i$  is summed in the middle of the corresponding point in the network diagram, and the sum obtained is called the absolute intermediate center of the point (denoted as  $C_{ABi}$ ). The specific measurement formula is as follows,

$$C_{ABi} = \frac{\sum_{j < k} b_{jk}(i)}{g_{jk}} \tag{3}$$

The standardized intermediate center (denoted as  $CR_{Bi}$ ), that is, the relative intermediate degree is measured as,

$$CR_{Bi} = \frac{2C_{ABi}}{(n-1)(n-2)} \tag{4}$$

Where  $CR_{Bi}$  ranges from 0.0 to 1.0, 0.0 means that the point cannot control any actor and is at the edge of the network; 1.0 means that the point is the core of the network and has a lot of rights to 100% control other actors. This value can also be used to compare the betweenness centrality of points in different network diagrams. In social network, the centrality of standardized social actors measures the degree to which social actors control resources in complex relationships. Social actors with high scores are the core controllers in the network. If the value of the standardized intermediate center is closer to 1.0, it means that the social actors have more control over the resources in the relational network.

3) *Closeness centrality*

Closeness centrality is a measure that is not controlled by others and reflects how close a node is to other nodes in the network. The overall centrality of a node depends on the distance between the node and other nodes in the network. If the distance is closer, the overall centrality of the node is higher. For social networks, the higher the closeness centrality, the greater the difference between nodes in the network; otherwise, the difference between nodes in the network is smaller. The shortcut distance between the other nodes connected to the  $i$ -node in the network graph is summed, and

the sum obtained is called the absolute closeness of the point. The specific measurement formula is as follows,

$$C_{APi}^{-1} = \sum_j d_{ij} \tag{5}$$

Where  $d_{ij}$  is the shortcut distance between point  $i$  and point  $j$  (that is, the number of lines included in the shortcut). The ‘relatively close centrality’ of a node is also called a standardized indicator of the closeness centrality. It is obtained by dividing the ‘absolute center’ of the node by the minimum closeness centrality. Its measurement formula is,

$$C_{RPi}^{-1} = \frac{C_{APi}^{-1}}{(n-1)} \tag{6}$$

From closeness centrality we can get, the information resources, power, prestige and influence of the innovative subject are related to the distance between the node and the center point. The farther the distance, the weaker the ability of all aspects of the innovation subject. The larger the value close to the center, the more the point is not the core point of the network.

B. *Condensed Subgroup Analysis*

An important task of social network research is to reveal the structure of the network. Condensed subgroup analysis is an important method for social network analysis, which is significant to understand the overall structure of the network. Separating subgroups from the overall network can understand the impact of subgroups on the overall network. For example, when there are two subgroups in the network, the intersection of the two subgroups means that the information can be transmitted quickly on the network. conversely, if there is no intersection between the two group analysis, the information transmission will be slower, and the two sets of features are different. According to different theoretical ideas and calculation methods, there are different types of condensate subgroup definitions and analysis methods, including factions, n-factions, n-sects, k-clusters, k-nuclei, and cohesion subgroup density. The main application of this paper is k-cluster, because it can better reflect the idea of cohesion.

C. *Core-edge structure analysis*

The core edge structure analysis divides the nodes in the network into two regions, a core region and an edge region according to the closeness of the connections between the nodes in the network. Nodes in the core occupy a relatively heavy position in the network. The purpose of core-edge structure analysis is to study which nodes in the social network are at the core and which are in non-core positions. The core-edge structure analysis in the social network analysis method can analyze the "location" structure of the network and distinguish the core and edge of the network. Therefore, the core-edge mechanism analysis method can be used to analyze which innovation entities in the innovation network are at the core position and which are in non-core positions, revealing the status quo of corporate innovation output in China's pharmaceutical industry.

III. ANALYSIS OF RESEARCH RESULTS

A. Analysis of central results

TABLE I. POINT CENTER ANALYSIS RESULTS

	1 Degree	2 NrmDegree	3 Share
广西信业生物技术有限公司	30.000	0.183	0.011
广西达庆生物科技股份有限公司	30.000	0.183	0.011
东莞市达庆医疗器械有限公司	30.000	0.183	0.011
广西中医药大学附设中医学校	20.000	0.122	0.008
广西中医药大学	20.000	0.122	0.008
广西重阳老年公寓	20.000	0.122	0.008
中国人民解放军军事医学科学院毒物药物研究所	20.000	0.122	0.008
华南理工大学	17.000	0.104	0.006
山东格林康宝药业有限公司	16.000	0.098	0.006
广州医科大学附属第一医院	16.000	0.098	0.006

In the above results, the first column (Degree) is the absolute degree of point centrality, and the second column (NrmDegree) is the relative degree of point centrality. From the results, we can see that the highest degree of point is Guangxi Xinye Biotechnology Co., Ltd., Guangxi Daqing Biotechnology Co., Ltd., Dongguan Daqing Medical Devices Co., Ltd., with an absolute center degree of 30. According to the definition of point degree, it indicates that the three companies have direct contact with the other 30 nodes in the network, that is, the three companies have worked with 30 authors to apply for patents. However, point centrality only considers the direct connection between a subject and other subjects, while ignoring his connection with all subjects in the entire network. That is to say, even if a subject has direct contact with many other subjects, the subject directly related to the subject has no connection with other subjects, even if the subject has a high degree of centrality, but this situation only exists on the 'local' network. From the perspective of the entire network, the significance of point centrality is not great.

TABLE II. INTERMEDIATE CENTRALITY ANALYSIS RESULT

	1 Betweenness	2 nBetweenness
181 华南理工大学	323.000	0.027
81 广州医科大学附属第一医院	320.833	0.027
286 广西来宾市香江生物科技有限公司	210.000	0.018
228 贵州省中国科学院天然产物化学重点实验室	202.000	0.017
48 中国人民解放军军事医学科学院毒物药物研究所	198.000	0.017
174 肇庆大华农生物药品有限公司	196.000	0.016
175 华南农业大学	184.000	0.015
287 中国科学院上海硅酸盐研究所	180.000	0.015
17 国家电网公司	178.667	0.015
115 华东理工大学	139.000	0.012

From the results, we can see that South China University of Technology has the highest degree of betweenness centrality, followed by the First Affiliated Hospital of Guangzhou Medical University. The betweenness centrality of these subjects is relatively high, indicating that they have a relatively large amount of research resources. At the same time, the results show that some of the subjects have a near-centrality of

0, and these subjects have little ability to control resources. The betweenness centrality of all subjects was counted, and the betweenness centrality degree of 957 subjects was found to be 0, accounting for 87.6% of the total number of subjects. This ratio is very high. At the same time, there are fewer subjects with higher central centers. This also shows that most subjects do not have the ability to control resources, and only a few subjects have relatively high ability to control resources.

TABLE III. INTERMEDIATE CENTRAL PRINCIPAL COMPONENT ANALYSIS RESULTS

Component	Nodes	Proportion
1	30	0.027
2	2	0.0028
3	3	0.003
4	8	0.007
5	6	0.005
6	6	0.005
7	2	0.0028
8	5	0.005
9	18	0.016
10	3	0.003

When calculating closeness centrality, the network must be a fully connected graph, ie each node is associated with all other nodes in the network. Otherwise, there is no connection between some nodes and other nodes, it can't reach any other point, there is no distance at all. The more isolated a point is, the smaller the sum of its distance from other points. According to the calculation method close to the centrality, its degree is rather large. Obviously, it violates the original intention of defining closeness centrality. The social network constructed in this paper is not a completely connected graph, so it is not possible to directly calculate the near-centrality. The component analysis must be performed firstly because the component is the largest connected subgraph in the network. Fractional distribution through a collaborative network analysis found that the largest component of the network contained 392 subjects.

B. Condensed subgroup analysis

The following is the result of k-cluster analysis of the cooperative network constructed in this paper. The value of k is 2, and a total of 758 2-plexes are found. The results are as follows,

WARNING: Valued graph. All values > 0 treated as 1  
 WARNING: Directed graph. Direction of arcs ignored.  
 758 k-plexes found.

According to the concept of k-cluster, if the size of a 2-cluster is n, the members in the plex are directly related to at least n- 2 other members. If the first plex contains three members, then each of them has a direct relationship with at least one other member. From the analysis results, the size of

the 2-cluster is relatively small, and most 2-clusters are 3 in size, which means that most of the subjects are directly related to at least one other subject in the cluster to which they belong. This shows that the scope of cooperation between the subjects in the cooperative network is not very wide, and most of the subjects have only cooperative and innovative relations with a few subjects. There are also some parts that are more than 3 in size, indicating that their partnerships are mature, and through these subjects, different 2-from can establish contact. These subjects play an important role in the network.

### C. Core-edge structure analysis

The result of the core-edge structure analysis places the 136-bit body at the core and the 957-bit body at the edge. Among the core subject groups, the subjects with the largest number of patent applications applied for 15 patents, and at least only one patent was applied. In the core-edge structure analysis results of the social network analysis method, there are 58 subjects in the core subject group with more than 4 published documents. That is, there are 31 subjects whose patents are more than 4, which are not placed at the core. Judging from the number of patents, these 41 subjects are at the core, but from the perspective of cooperation with others, these subjects are not at the core.

## IV. CONCLUSION AND INSUFFICIENT RESEARCH

China is the largest emerging pharmaceutical market in the world. The pharmaceutical industry has considerable cooperative patent output. The 2016 China Pharmaceutical Cooperation Patent Data is used to build a cooperative network and the analysis results clearly show the network cooperation relationship among the innovation subjects from a macro perspective, and reveal microscopically which innovation subjects are at the core position and edge position. At the same time, the regions where core and non-core enterprises are located imply the imbalance of development in China's pharmaceutical industry, which indicates the direction that the entire Chinese pharmaceutical industry should work hard. For a single innovation subject, a clear cooperation network structure also provides a reference for the next step of cooperation object selection.

This paper uses social network analysis methods to conduct collaborative network analysis from the perspectives of centrality, cohesion subgroup and core-edge structure, providing a new perspective for collaborative network analysis. The results of the central analysis show that the degree of centrality of the point has a great similarity with the top 20 subjects of the number of patent applications, indicating that the subjects with more patents are more frequently cooperating with other subjects. At the same time, there are few subjects with relatively high centrality, indicating that most subjects do not have the ability to control resources, and only a few subjects have relatively high ability to control resources.

The results of the component analysis before the near-central analysis showed that the largest component of the cooperative network contained 392 actors, indicating that the links between the cooperative network nodes are relatively close, and many subjects have no other subjects. Establish a

direct contact path. The results of the condensed subgroup analysis show that the scale of the coherent subgroup is relatively small, indicating that the scope of cooperation between the subjects in the cooperative network is not very wide, and most of the subjects only have cooperation with a few enterprises or units. The degree of faction in the cooperative network is relatively large, and there are few exchanges between subgroups in information sharing and scientific research cooperation, which is not conducive to the development of this field. The results of the condensed subgroup analysis also show that the cooperation phenomenon mainly occurs in the same regional unit, and the cooperation phenomenon between different regional units is relatively small, which is not conducive to information exchange and sharing in the field of information science to a certain extent.

The social network analysis method is an effective tool for collaborative analysis. However, this paper only uses a relatively simple binary network to represent the cooperative network, ignoring some information of the actual data. In the subsequent research, we should consider the more accurate representation of the actual data. Assignment networks for collaborative analysis to arrive at more reliable and convincing results.

## ACKNOWLEDGMENT

This work was supported by regional project of national natural science foundation of China (the project number is 71764033), supported by applied basic research project of Yunnan (the project number is 2017FB101), supported by the graduate innovation fund project of Yunnan University of Finance and Economics (the project number is 2020YUFEYC014). Thank for all the sponsors.

## REFERENCES

- [1] Q. H .Zhu, Social Network Analysis Method and Its Application in Information Science[J]. Theory and Exploration, 2008, 31(2): 179-183.
- [2] M. Granovetter, The Strength of Weak Ties [J]. American Journal of Sociology, 1973(78): 1287-1303.
- [3] T .Kuhn, The Structure of Scientific Revolution [M]. Chicago: Chicago university Press, 1970.
- [4] M .Kilduff, W .Tsai, Social Networks and Organizations [M]. London: Sage Publications, 2003.
- [5] B. Uzzi, Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness[J]. Administrative Science Quarterly, 1997, 42(1):35-67.
- [6] Z. Griliches, Issues in Assessing the Contribution of Research and Development to Productivity Growth[J]. The Bell Journal of Economics, 1979, 10(1):92-116.
- [7] P. M .Romer, Increasing Returns and Long-Run Growth[J]. Journal of Political Economy, 1986, 94(5):1002-1037.
- [8] P. K .Campbell, Environmental regulation and innovation: a panel data study[J]. The Review of Economics and Statistics, 1997, 136(4):610-619.
- [9] O. Janssen, N .W .Van Yperen, Employees' goal orientations, equality of leader-member exchange, and the of job performance and job satisfaction[J]. Academy of Management, 2004, 47(3): 368-384.
- [10] Y. Han, J. Q. Liao, L. R. Long, Construction and Empirical Research on Employee Job Performance Structure Model[J]. Journal of Management Sciences, 2007(05): 62-77.
- [11] P. F. Drucker, The discipline of innovation [J]. Harvard Business Review, 1998, 76 (6):149-157.

- [12] C. Berg, Orchestrating Innovation Networks[J]. *Academy of Management Review*, 2006, 24(3):15-19.
- [13] C. Uwe, C. Elisa, M. Andreas, Networks and Innovation: The Role of Social Assets in Explaining Firms' Innovative Capacity[J]. *European Planning Studies*, 2010, 18(12).
- [14] C. Phelps, A Longitudinal Study of the Influence of Alliance Network Structure and Composition on Firm Exploratory Innovation[J]. *Academy of Management Journal*, 2010, 105(16).
- [15] Y. S. Su, E. W. Tsang, M. W. Peng, How do internal capabilities and external partnerships affect innovativeness? [J]. *Asia Pacific Journal of Management*, 2009, 105(30): 312-318.
- [16] J. Liu, Overall Network Analysis Lecture: UCINET Software Practical Guide [M]. Shanghai Gezhi Publishing House, Shanghai People's Publishing House, 2009.
- [17] J. D. Luo, Social Network Analysis Lecture [M]. Beijing: Social Sciences Academic Press, 2005, 96, 98, 156.