Analysis of computer science paper collaboration network among cities in China

Yang Xin

School of Economics, Guangxi University, Nanning, 530004, China

Abstract: Each city will cooperate in scientific research activities in order to improve its own level of scientific and technological innovation. This paper uses the computer science cooperation paper data between cities in my country from 1982 to 2014, and uses complex network analysis methods to find out: (1) my country's city paper cooperation network There are obvious core-peripheral characteristics in the degree centrality and betweenness centrality of , but its degree centrality always retains core-peripheral characteristics.

Keywords: collaborative paper; complex network analysis; network centrality

1. Introduction

Under the general trend of economic globalization, economic and trade exchanges between various regions are becoming increasingly frequent, and production factors such as finance, talents and knowledge are also flowing large-scale and intensively between different regions along with people's economic activities. If these resources can be obtained and utilized well, a region can be at the forefront of the wave of economic globalization, significantly improve its own economic strength, and further promote the overall improvement of regional social security, scientific and technological innovation culture, and business environment, thus Enhance regional core competitiveness. In the process of competing with other regions, it relies on its own development advantages to attract a large number of production factors, promote the development of its own comprehensive strength, and form a virtuous cycle.

From the perspective of domestic economic development, it generally manifests as economic development competition between different provinces and regions and the competition between different cities for talents, technology, capital and other resources. The stronger the innovation capability of a region, the higher the technological progress or total factor productivity of the region, which will lead to a higher economic growth rate. Under fully mobile factor market conditions, resources will spontaneously converge to areas with higher innovation capabilities to maximize factor production efficiency. Therefore, for a city to improve its level of scientific and technological innovation, it can also attract more production resources, thereby gradually improving the city's economic development level under reasonable allocation and utilization, enhancing the city's comprehensive competitiveness, and further improving the city's competitiveness. Stand out from the competition with other urban areas, absorb more elements, and achieve sustainable development of the urban economy. Among them, the city's level of scientific and technological innovation plays a vital role as a decisive factor. The achievements of scientific and technological innovation generally include papers and monographs, product prototypes, patents, etc. The richer the city's scientific and technological innovation achievements, the better it can transform scientific and technological knowledge into actual economic benefits. Generally speaking, regions with higher levels of technological innovation have richer technological innovation achievements. Therefore, people often use the number of technological innovation achievements in a region to reflect its technological innovation level.

With the popularization and application of the Internet, the cost of communication between people has been greatly reduced. Crossing long geographical distances is no longer a necessary condition for people to communicate. The proportion of scientific and technological innovation results produced by cooperation in various fields is also increasing. Taking the Nobel Prize, which represents the highest contribution to mankind in various fields, as an example, more than two-thirds of the Nobel Prize-winning projects were awarded in collaboration between two or more people. This shows that cooperative behavior has a significant role in scientific and technological innovation activities. vital
Among the many scientific and technological innovation achievements, scientific research papers are more active in scientific and technological exchanges and cooperation activities due to their relatively wide range of creative groups, weak geographical dependence of the research process, and online results delivery methods. In order to improve their own scientific and technological innovation capabilities and enhance their comprehensive competitiveness, each city will also strengthen scientific research cooperation with other cities. As the team of scientific researchers continues to expand and the level of academic research continues to improve, the frequency of paper collaboration between cities has gradually increased, and an intricate paper collaboration network has been formed between cities. At different time points, the number of paper collaborations between cities also reflects the speed of development of urban scientific and technological innovation levels. Studying the inter-city paper cooperation network will help to understand the development process of scientific and technological innovation in cities, understand the regional distribution characteristics of scientific and technological exchange activities, and provide objective and effective basis for the formulation of future regional economic and scientific and technological innovation policies.

Therefore, this article uses the complex network analysis method to analyze the structural characteristics and time evolution trends of the complex network of paper collaboration between cities in my country based on the paper cooperation data in the field of computer science in 221 cities in China from 1982 to 2014, providing a basis for future scientific and technological innovation and regional development. Provide objective and effective basis for the formulation of economic policies.

2. Literature review

Due to the limited nature of resources, labor and capital cannot grow indefinitely. Economic growth driven by the increase in the number of production factors will eventually encounter bottlenecks due to resource limitations. Solow (1957) proposed that in the process of economic growth, total factor productivity plays an important role, and it promotes economic growth by improving resource allocation efficiency and technological progress. Existing research mainly explores how to improve total factor productivity, transform industrial structure, and transform economic growth mode from the aspects of improving resource allocation efficiency and promoting technological progress (Liu W. and Zhang H., 2008; Tang W. et al., 2014; Chen B. et al., 2015; Wen D., 2019).

It is generally believed that the continuous deepening of scientific research can promote the sustainable development of science and technology. To measure the level of scientific research, relevant scientific research results are usually used to reflect it. When there is a huge breakthrough discovery in a field or a landmark new concept is proposed, it is also reflected. It will urge relevant scientific researchers to produce quite abundant scientific research results. Specifically, the forms of scientific research results can be divided into: papers and monographs, patents, product prototypes, etc.

As scientific research develops to a certain stage, the cost of scientific research increases and the complexity of scientific research activities increases, which puts forward higher requirements for the cross-integration of multiple disciplines. Limited scientific research resources cannot meet the needs of scientific research, prompting conflicts between various scientific researchers and organizations. Scientific research cooperation (Zhao R. and Wen F., 2011). Scientific research cooperation can improve the scientific research output of scientific researchers, and can also bring greater social reputation to scientific researchers and increase academic influence (Zhao J. and Liao J., 2013; Qiu J. and Wen F., 2011). At the same time, it also brings about the reallocation of scientific research and production factor resources in different spaces (Xie C. and Liu Z., 2006). The inter-regional flow of R&D factors will produce spatial spillover effects and play a significant role in promoting economic growth. (Bai J. et al., 2017).

Due to the complex interconnections between partners and organizations in scientific research cooperation, complex network analysis methods are usually used in the research process to describe some overall characteristics of the scientific research cooperation system. Complex network analysis methods are widely used in statistical physics, mathematics, computer networks, social sciences and other fields to study the topology structure of complex networks, the formation mechanism of scale-free characteristics, and the evolution model of complex networks (Wang L. and Dai G., 2006; Zhou T. et al., 2005). The current research on scientific research cooperation by scholars mainly focuses on the direction of paper cooperation. Some scholars collect and organize the author's paper cooperation situation from the historical published paper data of certain journals, and analyze the structure of their paper cooperation network. Fu Y. et al. (2009) We constructed a paper cooperation
network based on the paper cooperation status of the journal "Scientific Research Management" from 2004 to 2008, and used social network analysis methods to analyze it. We found that the paper cooperation network includes an overall network and a small group network, and both have scale-free network characteristics[14]. Liu J. and Lu J. (2004) analyzed the collaboration network of chaos science papers published in Acta Physica Sinica and Chinese Physics from January 1998 to June 2004, and found that their scale-free network characteristics and small world characteristics. In terms of regional scale, scholars have conducted relevant research on paper collaboration networks from regional, national, and international scales[20]. Jiang K. and Yu T. (2017) used the pairwise paper collaboration data of 16 cities in the Yangtze River Delta urban agglomeration in the Web of Science to study the characteristics of the internal knowledge network among cities in the Yangtze River Delta region, and found that there is an obvious spatial polarization phenomenon in the knowledge cooperation network. Core-periphery structure, uneven degree of cooperation in papers between different disciplines, and the "Matthew Effect"[6]. Li J. et al. (2017) used the internal paper cooperation data of the Yangtze River Delta urban agglomeration from 2000 to 2015 from the VIP Journal Network to construct a scientific knowledge network in the Yangtze River Delta region, and analyzed the knowledge cooperation network in the Yangtze River Delta region from two aspects: the overall network and the local network. According to the distribution characteristics and evolution rules, it is found that the overall network shows an expansion trend and has the development characteristics of a "small world" network; the local network degree distribution is a bell-shaped distribution, evolving from a cooperative network to a research network [3]. Li D. et al. (2015) used information on co-authored papers and joint patent applications in China's biotechnology field from 2000 to 2009 to explore the spatiotemporal complexity of the scientific knowledge network and technical knowledge network structure at the city scale in China[8]. Zhang D. et al. (2008) analyzed the characteristics of China's high-level scientific paper collaboration network at the city scale based on the scientific paper collaboration data in the Web of Science from 1975 to 2007, and explored its core-periphery structure[10]. Hu Y. et al. (2009) used the data of academic papers on supply chain management from 1990 to 2007 in the China Journal Network database and the data of SCI papers in the direction of supply chain management from 1990 to 2007 in the international ISI Web of Knowledge database to construct a scientific cooperation network. Comparing the characteristics of domestic and foreign paper collaboration networks within the same discipline, it was found that both domestic and international paper collaboration networks are scale-free networks, but there are large differences in cooperation methods[13]. Liu C. et al. (2017) characterized the topological structure, spatial pattern and proximity mechanism of the global scientific research cooperation network based on the data of co-authored papers in all subject areas in 2014 included in the Web of Science core collection[9].

Current research on paper collaboration networks has found that most paper collaboration networks exhibit scale-free network characteristics, have certain "small world" network characteristics, and overall present a core-periphery structure. However, the characteristics of paper collaboration networks have certain subject specificity. Statistical analysis of paper collaboration networks often uses a large number of complex network analysis tools. Common ones include Ucinet, Netdraw, Pajek, Gephi, VOSViewer, etc. Visual tools such as ArcGIS are also used to present the real-world spatial characteristics of paper collaboration networks. The current research has relatively complete descriptions of the characteristics of regional paper collaboration networks at different scales, but there are still several shortcomings. First of all, although there are certain commonalities in the structure of paper collaboration networks between different disciplines, there are also certain differences. Current research focuses more on the description of the structural characteristics and evolution trends of scientific research collaboration networks, and the analysis of network structures between different disciplines. Differences are poorly explained. Secondly, the structure of the paper collaboration network in different time spans should be related to the degree of development of the subject direction and the relevant policy incentive system. In the evolution analysis of the paper collaboration network, the combination of such factors is still insufficient. Therefore, this article uses the paper collaboration data in the field of computer science in 290 cities in China from 1982 to 2014, and uses complex network analysis methods to explore the structure and evolutionary characteristics of the paper collaboration network in the field of computer science in China, in order to make useful supplements.

3. Data sources and analysis methods

3.1 Data source

The data in this article comes from the search results of the researcher's social network search and
mining system ArnetMiner, which mainly uses social network analysis methods. ArnetMiner is based on massive international academic paper data, extracts and integrates the semantic information of researchers, establishes a social relationship network between them, and provides multi-granularity and multi-dimensional semantic search and mining services, including personal information search, research interest analysis, and academic capabilities. Evaluation, expert discovery, reviewer recommendation, expert relationship search, etc. This article uses the ArnetMiner database as the basis to search for inter-city computer science paper cooperation data in China, and eliminates cities without paper cooperation. From 1982 to 2014, there are a total of 15,103 sets of inter-city computer science paper cooperation paper data between 221 cities in China. For each paper with n>1 authors, 1 is added up in the city where the author of each paper is located to construct a paper cooperation matrix between cities. Table 1 is a paper cooperation matrix between 221 cities.

Table 1: Paper collaboration matrix of 221 cities in China from 1982 to 2014

<table>
<thead>
<tr>
<th>city</th>
<th>ankang</th>
<th>anqing</th>
<th>anshan</th>
<th>anyang</th>
<th>baicheng</th>
<th>...</th>
<th>zunyi</th>
</tr>
</thead>
<tbody>
<tr>
<td>ankang</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>anqing</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>anshan</td>
<td>0</td>
<td>0</td>
<td>333</td>
<td>41</td>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>anyang</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>353</td>
<td>1</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>baicheng</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>zunyi</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>4</td>
</tr>
</tbody>
</table>

The number of papers published over the years under the subject category of computer software and computer applications retrieved based on the keyword “computer” on CNKI is shown in Figure 1. CNKI is an important platform for Chinese scholars to publish papers for academic information acquisition and exchange. The disciplines on its platform The trend of publication in different years can reflect the development of the discipline to a certain extent. According to the annual trend of publication of papers in the computer discipline on CNKI, this article divides the period from 1982 to 2014 into 1982-1993, 1994-2008, and 2009-2014 There are three intervals in the year. Among them, the number of papers published in the computer science discipline in the 1982-1993 interval is relatively small, all at or below 1,000, which is in the initial stage of scientific research output; between 1994 and 2008, the number of published papers in the computer science discipline has increased significantly. It increased, reaching a peak of 16,304 articles in 2008, which was a stage of rapid growth; from 2009 to 2014, the number of published articles was relatively stable and maintained at a high level. By characterizing the characteristics of paper collaboration networks in different intervals, we examine the evolution patterns of computer science paper collaboration networks among Chinese cities in different stages of computer science development.

![Figure 1: The publication trend of papers in the computer discipline of CNKI over the years](image-url)
3.2 Research methods

Network centrality model

1) Degree centrality, $C_D$ refers to the number of direct connections to a single node in the network, indicating the density of connections between a single node and other nodes in the network. In the national computer science research paper cooperation network, the degree centrality of a city represents the number of other cities that have paper cooperation relationships with the city:

$$C_D(i) = \sum_{j=1}^{N} a_{ij}$$

In the formula, $a_{ij}$ represents the urban scientific research cooperation adjacency matrix. If there is scientific research cooperation, it is assigned a value of 1, and if there is no scientific research cooperation, it is assigned a value of 0.

2) Closeness centrality, $C_C$ means the reciprocal of the sum of the shortest paths from a node to all other nodes in the network multiplied by the number of other nodes, reflecting the closeness between a node and other nodes in the network. Closeness centrality the larger the value, the greater the value of the node in the network and the more central it is.

$$C_C(i) = \frac{N - 1}{\sum_{j=1, j \neq i}^{N} d_{ij}}$$

In the formula, $d_{ij}$ represents the number of shortest paths between nodes, and is the number of nodes.

3) Betweenness centrality, $C_B$ is used to measure the proportion of the number of all shortest paths in the network that pass through the node. The greater the betweenness centrality of a node, the stronger the node's ability to control the network. In the scientific research paper collaboration network, indicating the city’s ability to serve as a “transfer station” for academic knowledge:

$$C_B(i) = \sum_{i=1, j=1, i \neq j}^{N} \frac{N_{jk}(i)}{N_{jk}}$$

In the formula, $N_{jk}$ represents the number of shortest paths between nodes and $N_{jk}(i)$ represents the number of shortest paths between node j and node k passing through node I.

4. Spatial structure of the computer science paper collaboration network

4.1 Overall paper collaboration network space structure

By importing the paper cooperation data of 221 cities from 1982 to 2014 into the pajek software, we can obtain the inter-city computer science paper cooperation network diagram (Figure 2). There are 233 nodes in the paper cooperation network, and the number of interconnections between them reaches 3210. The average network density is 0.12708585, and the average centrality is 28.27149321.

![Figure 2: Paper collaboration network diagram of 221 cities from 1982 to 2014](image-url)
1) Degree centrality

The degree centrality of each city node is calculated in the pajek software, and the network is drawn based on the point-degree centrality of different nodes as shown in Figure 3. The point centrality distribution of different cities is quite different, with a minimum value of 2, a maximum value of 162, an average value of 28.2715, and a standard deviation of 28.2543. The degree centrality of most cities is small and they are distributed on the periphery of the network. Most of the cities distributed in the central core area are municipalities or provincial capitals. The top 5\% cities with centrality are: Beijing, Shanghai, Changsha, Wuxi, Nanchang, Guangzhou, Xianning, Tianjin, Dalian, and Chengdu. Due to their rich economic and educational resources, these cities often have a large number of colleges and universities or scientific research units, and their scientific research activities are more frequent than those of ordinary prefecture-level cities. In particular, the centrality of Beijing and Shanghai is 162 and 159 respectively, far higher than the average network degree of 28.2715, which is also much higher than the centrality level of other general cities, and has become a growth pole in the paper collaboration network.

![Figure 3: Degree centrality diagram of paper collaboration network in 221 cities from 1982 to 2014](image)

2) Closeness centrality

The closeness centrality of each city node is calculated in the pajek software, and the network is drawn based on the closeness centrality of different nodes as shown in Figure 4. The closeness centrality values of different cities are relatively balanced, with the minimum value being 0.3514 and the maximum value being 0.7857. The average network closeness centrality is 0.5152, the standard deviation is 0.0593, and the overall fluctuation is small. This may be related to the characteristics of the computer science discipline itself. Scientific research activities related to the computer science discipline are bound to use a large number of Internet tools for research. Due to the high convenience of communication between different individuals on the Internet, the flow of information between two nodes is not affected by geographical distance. Limitations, all have relatively equal access to information resources, so each node has a high degree of connectivity between pairs, and its closeness to centrality shows a relatively even distribution.

![Figure 4: Closeness centrality diagram of paper collaboration network in 221 cities from 1982 to 2014](image)
3) Betweenness centrality

The betweenness centrality of each city node is calculated in the pajek software, and the network is drawn based on the betweenness centrality of different nodes, as shown in Figure 5. The betweenness centrality distribution of urban paper collaboration network is very different, with the minimum value below 0.0001, the maximum value 0.1369, the mean value 0.0044, and the standard deviation 0.0137. Similar to point centrality, cities with higher economic and educational levels such as Beijing and Shanghai are located in the core area of the paper cooperation intermediary centrality network, while other general small and medium-sized cities have smaller values and are located at the periphery of the network. Since betweenness centrality reflects the node's ability to control the network, that is, the node's ability to act as an intermediary or "transit station" in the network, these cities are rich in scientific research-related resources such as universities and scientific research units, and some excellent Scholars or researchers have a tendency to travel to these areas to conduct further research. On the other hand, scientific research exchange activities such as academic conferences, scientific and technological services, and technology transfer are mainly held in cities with relatively developed economies. A large number of scientific research cooperation between different cities are promoted by such scientific research exchange activities. Therefore, such cities often have higher betweenness centrality and distributed in the core of the network.

![Figure 5: Betweenness centrality diagram of paper collaboration network in 221 cities from 1982 to 2014](image)

4.2 Evolution trend of paper collaboration network in stages

By importing the paper collaboration data of different years from 1982 to 2014 into the pajek software in stages for calculation, we can obtain the paper cooperation network diagram of my country's computer science disciplines at different stages, as shown in Figures 6, 7, and 8. In the initial stage of the development of computer science in my country, from 1982 to 1993, my country's computer science paper cooperation network was relatively simple, with a small number of network nodes and short interconnection paths between nodes. Most cities' paper cooperation only remained within their own cities or within their own cities. The cooperation with a very limited number of specific cities has not formed a very close scientific research cooperation relationship. From 1994 to 2008, when the computer science discipline was booming, the number of nodes in the paper cooperation network increased significantly, and paper cooperation activities between cities became more frequent. A relatively dense paper cooperation network has basically been formed, but some cities are still distributed in the peripheral areas of the network. There is less interaction with other cities and there is room for further tightening of the network. From 2009 to 2014, the output of computer science papers was basically at a relatively stable and high level. The number of network nodes increased slightly, the network density further deepened, the number of nodes distributed in peripheral areas decreased, and relatively mature and stable papers were basically formed. Cooperation network.
Figure 6: Paper collaboration network from 1982 to 1993

Figure 7: Paper collaboration network from 1994 to 2008

Figure 8: Paper collaboration network from 2009 to 2014

Further calculate the degree centrality of the paper collaboration network at different stages and display it visually in the network. The degree centrality diagram of the paper collaboration network at different stages can be obtained as shown in Figures 9, 10, 11. Degree centrality between 1982 and 1993. The minimum value of centrality is 2, the maximum value is 21, the average value is 5.7037, the standard deviation is 4.0808, and the network density is 0.1056. The minimum value of degree centrality from 1994 to 2008 is 2, the maximum value is 137, and the average value is 23.2376. The
standard deviation is 22.8382, and the network density is 0.1271. From 2009 to 214, the minimum value of degree centrality is 1, the maximum value is 152, the average value is 25.1005, the standard deviation is 25.2075, and the network density is 0.1193. Generally speaking, with the continuous development of computer science, the connections between different nodes in the network have become closer, and the maximum value and average value have increased. However, in different stages of development, urban areas with relatively high economic levels, such as municipalities and provincial capitals, have always been at the core of the paper cooperation network.

5. Research conclusion

This paper conducts a phased complex network method analysis on the computer science paper collaboration data in 221 cities in my country from 1982 to 2014. The study finds: (1) Midpoint degree centrality and intermediary in the computer science paper cooperation network among cities in my
country from 1982 to 2014. Centrality has a relatively obvious core-peripheral structure. Economically developed areas are often in the core area of the network. With more colleges and universities and scientific research institutions inside, they have closer ties with other network nodes and more. It plays the role of "transfer station" for knowledge dissemination. (2) The distribution of closeness centrality of each node is relatively balanced, which may be related to the characteristic of computer science itself that information dissemination is not geographically restrictive. (3) In the initial, growth, and mature stages of the development of computer science in my country, the paper collaboration network gradually became more dense, the paper cooperation activities between cities and other cities became closer, and the paper cooperation network evolved in the direction of agglomeration, but the degree centrality It always maintains the more obvious core-peripheral characteristics.

References