

# Research groups of oncology co-authorship network in China

Qi Yu · Hongfang Shao · Zhiguang Duan

Received: 5 March 2011 / Published online: 6 August 2011  
© Akadémiai Kiadó, Budapest, Hungary 2011

**Abstract** This paper aims at analyzing and extracting the research groups from the co-authorship network of oncology in China. By use of centrality, component analysis, K-Core, M-Slice, Hierarchical Clustering analysis, and Multidimensional Scaling analysis, we studied the data from 10 Core Chinese Oncology journals between 2000 and 2009, analyzed the structure character of the Chinese Oncology research institutes. This study advances the methods for selecting the most prolific research groups and individuals in Chinese Oncology research community, and provides basis for more productive cooperation in the future. This study also provides scientific evidences and suggestions for policymakers to establish a more efficient system for managing and financing Chinese Oncology research in the future.

**Keywords** Oncology · Co-authorship · Network · Research group

**Mathematics Subject Classification** 68R10

**JEL Classification** C38

## Introduction

In recent years, the increasing prevalence of collaboration has become one of the most interesting features of the research process (Bordons and Gómez 2000). Cooperation has been the primary way to advance scientific research. Increasingly, scientific research has been carried out with research groups, which are the most important work floor entity in science, particularly in the fields of the natural sciences (Van Raan 2008). Biomedical research development is no exception to this trend and is increasingly becoming a collaborative activity; one in which researchers form groups that go beyond the formal

---

Q. Yu · H. Shao · Z. Duan (✉)  
School of Public Health, Shanxi Medical University, No. 56, South Xinjian Road, Taiyuan, China  
e-mail: dzg528@yahoo.com.cn

institutional boundaries of the centers with which they are affiliated (González-Alcaide et al. 2010; Beaver and Rosen 1978).

Co-authorship network analysis has been widely used to study the cooperation of research groups in many disciplines. However, few researches have focused on the emergence of medical co-authorship networks in China. The medical sciences have experienced dramatic changes during recent years along with policy of opening and reform in China and the rapid growth of Chinese economy. New Chinese research policies offered a platform for Chinese researchers to collaborate in a broader scope. It is necessary to analyze medical research groups in China. Currently, of the nearly 7 million people die from cancer each year in the world, 24% are from China. The malignant disease is the number one killer in China. Developing innovative ways to prevent and to treat cancer is a top medical priority in China and there are now more than 40 different monthly publications devoted to oncology. So, we choose oncology, an important branch of medicine, as our study object.

This paper advances the techniques to study and extract cohesive research groups from Chinese Oncology co-authorship network. With these techniques and the data extracted from 10 major journals in the field of oncology in China from 2000 to 2009, we study the biomedical research structure in China at individual and group level, and provide scientific evidences for policymakers to establish a more efficient system for managing and financing Chinese Oncology research in the future.

## Methods

### Definition of research group

One of the most precise definitions of research group is that a community of scientists who work together in the approach to and development of research, sharing material and financial resources, but not necessarily organized along the lines of the formal structure of the institution or institutions where the activity is conducted (Zulueta and Bordons 1999). Cohen identified two methodological approaches to deal with research groups: output-based and input-based (Cohen 1991). The output based approach, researchers who cite each other or coauthored the same paper, regardless of their affiliation, are regarded as the same group members. In this case, the groups are not necessarily affiliated in the same administrative institutions. But one of the defects of the method is that it does not count the non-publishing-scientists. The input-based approach, on the contrary, requires that the members of one group have an administrative or institutional reality and includes all members, whether or not they publish. This paper adopts the output based approach to generate network apt for detecting research groups in the discipline of oncology.

### The co-authorship network

Co-authorship of articles is not a perfect measure of the processes driving collaboration among researchers in a specialty (Melin and Persson 1996; Subramanyam 1983; Vuckovic-Dekic 2003). However, there are many advantages when studying collaboration through co-authorship in collections of publications as this method is inexpensive and practical compared to surveys and interviews (Katz and Martin 1997).

Social network analysis focuses on ties among actors like, for example, people, organizations, and countries. We use graphs to detect and interpret patterns of social ties among

actors. The graph's vertices represent social actors and the graph's edges connect pairs of vertices and thus represent social interactions. This representation allows us to apply graph theory, a branch of mathematics, to the analysis of what would otherwise be an inherently elusive and poorly understood problem: the tangled web of our social interactions.

The co-authorship network is an important kind of social network and has been widely used to detect the structure of scientific collaborations and the status of individual researchers. Co-authorship implies a strong social bond, because authors coauthoring one paper always have a temporal and collegial relationship, which places it more squarely in the realm of social network analysis.

An early example of co-authorship network analysis is the Erdős Number Project, in which the smallest number of co-authorship links between any individual mathematician and the Hungarian mathematician, Erdős, were calculated (Castro and Grossman 1999). Newman studied and compared the co-authorship graph of arXiv, Medline, SPIRES, and NCSTRL and found a number of network differences between experimental and theoretical disciplines (Newman 2001a, b). With the graph containing all relevant publications of members in an international collaboration network COLLNET, Yin et al. found that this scientific community displayed many aspects of a small-world network and was vulnerable to disruption (Yin et al. 2006). With the Science Citation Index (SCI) data of 1990–2000, Wagner and Leydesdorff found that in the period of 1990–2000, the global network had expanded to include more nations and became more interconnected (Wagner and Leydesdorff 2003). Co-authorship analysis had also been applied to various ACM conferences: Digital library (Liu et al. 2005), Information Retrieval (SIGIR) (Smeaton et al. 2003), Management of Data (SIGMOD) (Nascimento et al. 2003) and Hypertext (Chen and Carr 1999). Cronin and Shaw found physical location played an important role in collaboration in their study of Rob Kling's intellectual impact and influences (Cronin and Shaw 2007). Hou et al. (2008) has revealed the co-authorship pattern of Scientometrics using the data from SCI.

### Analyzing research group cohesiveness

Several techniques are used for extracting and analyzing cohesive research group from oncology co-authorship network in China, such as centrality, component, K-Core, M-Slice, hierarchical clustering and multidimensional scaling (MDS).

#### *Centrality*

Centrality, which is one of the oldest concepts in network analysis, was used in our present study. Viewed from a sociocentered perspective, the network as a whole is more or less centralized. In highly centralized network, information spreads easily and the individuals in the center are indispensable for the transmission of the information.

We have adapted three common centrality metrics, namely degree centrality, closeness centrality, and betweenness centrality (Wasserman and Faust 1994) to analyze the oncology co-authorship network in China.

Degree centrality of a vertex is defined as the number of lines incident with it. Degree centrality represents the simplest notion of centrality since it is just the number of neighbors of a vertex in the network.

The closeness centrality of a vertex is the number of others vertices divided by the sum of all geodesic distances between the vertex and all others, where larger distances yield

lower closeness centrality scores. The closer a vertex is to all other vertices, the easier information may reach it, the higher its centrality.

Degree and closeness centrality are based on the reachability of a vertex within a network, while betweenness centrality rests on the idea that a vertex is more central if he is more important as an intermediary in the network. The betweenness centrality of a vertex is the proportion of all geodesics between pairs of other vertices that include the vertex.

### *Component*

Component is a sub-network with the characteristic that there is a path from any vertex to any other vertex of the sub-network. A co-authorship network usually consists of many disconnected components, the largest of which is the one we usually focus on. Component can be used to learn about the structure of co-authorship network.

### *K-Core*

A K-Core is a maximal sub-network in which each vertex has at least degree  $k$  within the sub-network. In a K-Core, we can use degree to identify groups of authors that are tightly connected because each author has a particular minimum degree within the group. K-Cores are nested: a vertex in a 4-Core is also part of a 3-Core, but not all members of a 3-Core belong to a 4-Core.

### *M-Slice*

K-Core will not take line value into account when extracting subgroups. In fact, line values in co-authorship network are considered more important because they are less personal and more institutional. From this point of view, we may define cohesive subgroups on line value rather than on the number of neighbors. The larger the line-value between two authors, the stronger or more cohesive their tie, the more similar or interdependent they are. This brings us to the concept of an M-Slice (Scott 1991): A maximal sub-network containing the lines with value equal to or greater than  $m$  and the vertices incident with these lines. The M-Slice is nested like K-Core.

### *Hierarchical clustering*

Hierarchical clustering was used to extract subgroups in the co-authorship network in many studies. Hierarchical clustering creates a hierarchy of clusters which can be represented in a tree structure called a dendrogram. The root of the tree consists of a single cluster containing all authors, and the leaves correspond to individuals.

### *Multidimensional Scaling (MDS)*

MDS is a set of data analysis techniques often used in information visualization for exploring similarities or dissimilarities in data. An MDS algorithm starts with a matrix of item-item similarities. Then it assigns a location to each item in  $N$ -dimensional space, where  $N$  is specified a priori. For sufficiently small  $N$ , the resulting locations may be displayed in a graph or 3D visualization.

## Data

We choose ten major Oncology journals from Chinese Science Citation Database (CSCD: <http://sdb.csd.ac.cn/>) and China National Knowledge Infrastructure (CNKI: <http://www.cnki.net/>) for the period 2000–2009. CSCD, regarded as “China’s SCI”, covers many higher quality journals published in China. We obtain bibliographic records, such as author name, title, affiliation, citation etc., from CSCD. However, CSCD only offers the first author’s affiliation. In order to solve the problem of ‘homonymy’ below, we must know all author’s affiliations, so we abstain non-first authors’ affiliations from CNKI. We believe these ten publications suffice to reveal the collaboration structure of China’s Oncology research.

These core publications are: (1) Chinese Journal of Cancer (2) Journal of Practical Oncology (3) Chinese Journal of Clinical Oncology (4) Chinese Journal of Radiation Oncology (5) Chinese Journal of Oncology (6) Tumor (7) China Oncology (8) Cancer Research on Prevention and Treatment (9) Chinese Journal of Cancer Biotherapy (10) Carcinogenesis, Teratogenesis and Mutagenesis.

## Data refinement

After the adoption of the assumption that a research group can be defined from a collection of published papers signed by a list of authors, we had to standardize their names for that purpose (Calero et al. 2006; Perianes-Rodríguez et al. 2010). We often encountered two problems when translating the author’s Chinese name into English: homonymous names (two authors having the same name) and synonymous names (two authors with different names sharing the same pinyin). To solve the problem of homonymy, we combine the author’s name and affiliation. There are 2,307 affiliations in this study, and each affiliation is assigned a unique ID. For example, we use ID ‘Org1’ to signify ‘the State Key Laboratory of Huanan Oncology’. Then, ‘Zhongzhen Guan\_Org1’ indicates Zhongzhen Guan coming from the State Key Laboratory of Huanan Oncology.

## Results and analysis

### Global view

With the data refined above, we generate a two-mode (author-paper) network, which contains 50,694 authors and 22,736 papers. Then, we extract the one-mode network, i.e. co-authorship network, from the two-mode network. The co-authorship network, with a low density of 0.000136, contains 174,465 edges (authors without collaboration are deleted). The highest line value, i.e. the co-author value, is 73.

In order to study the cooperation between “Administrative Divisions”<sup>1</sup> in China, we shrink all authors within an “Administrative Division” to a new vertex that represents the entire “Administrative Division” (Fig. 4, drawn with Pajek). Lines incident with shrunken vertices are replaced. For instance, all edges from Beijing to Shanghai in the original co-authorship network are replaced by one new line from Beijing to Shanghai, with its value equal to the sum of all original line values. Values of lines in the shrunk network represent

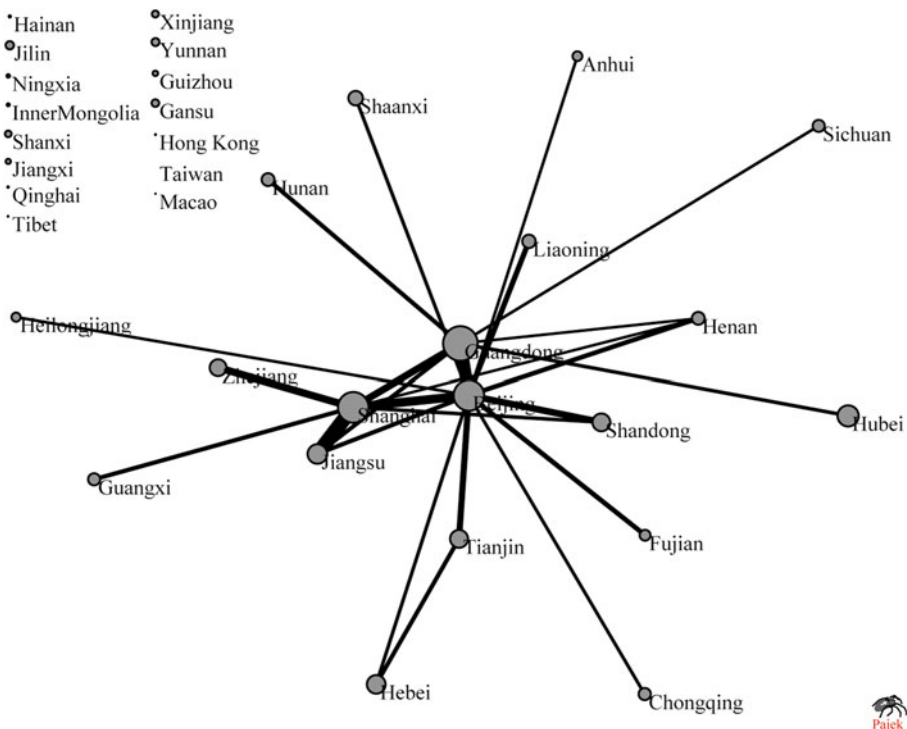
<sup>1</sup> China has 34 “Administrative Divisions”, including 23 provinces, 5 autonomous regions, 4 municipalities directly under the Central Government and 2 special administrative regions.

the value of total co-author frequencies between two “Administrative Divisions”. The vertex size represents total co-author value within one “Administrative Division”.

Figure 1 shows the co-authorship network that is shrunk according to China’s “Administrative Divisions”. To obtain a clear picture, we delete isolated vertices and lines with summed values below 100. The graph is plotted using the Kamada-Kawai algorithm, in which vertices with higher line value gather around in the center, while vertices with lower line value scatter in the periphery. We can see Beijing, Shanghai and Guangdong are in the centers of the graph, which conform to their preeminent economic and educational level in China. As shown in Table 1, the sum of the co-author value within these three “Administrative Divisions” is 231,626 articles, 21 times bigger than that among all other “Administrative Divisions” (10,936). The possible reasons are as follows: The three ‘Administrative Divisions’ are most developed areas in China, which offer a platform for researchers to collaborate in a broader scope; Researchers tends to cooperate within the same institution.

Centrality

We calculated the degree, closeness, and betweenness centrality for the oncology co-authorship network. We find 122 authors rank in top 1% of all the three centralities, of which 29 authors came from the State Key Laboratory of Huanan Oncology, 26 from Cancer Institute and Hospital of Chinese Academy of Medical Sciences, 13 from Affiliated Oncology Hospital of Fudan University, ten from Cancer Hospital of Tianjin Medical



**Fig. 1** Co-authorship within and among “Administrative Divisions” in the discipline of Chinese Oncology

**Table 1** Vertex value for the shrunk network

Rank	Vertex	Value	Rank	Vertex	Value
1	Guangdong	39,818	18	Anhui	3,120
2	Shanghai	31,344	19	Heilongjiang	2,728
3	Beijing	30,996	20	Jilin	2,247
4	Hubei	14,772	21	Yunnan	2,043
5	Jiangsu	12,728	22	Gansu	1,857
6	Hebei	11,508	23	Xinjiang	1,523
7	Tianjin	10,627	24	Shanxi	1,401
8	Shandong	10,486	25	Guizhou	935
9	Zhejiang	9,284	26	Jiangxi	861
10	Shaanxi	6,950	27	Ningxia	358
11	Liaoning	5,935	28	Inner-Mongolia	274
12	Hunan	5,447	29	Hainan	148
13	Henan	5,422	30	Qinghai	65
14	Chongqing	4,904	31	Hong Kong	25
15	Guangxi	4,876	32	Tibet	16
16	Sichuan	4,797	33	Macao	3
17	Fujian	3,900	34	Taiwan	0

**Table 2** Correlation between centrality measures and author's productivity

	Productivity	Degree	Closeness	Betweenness
Productivity	1	0.870**	0.258**	0.497**
Degree	0.870**	1	0.347**	0.558**
Closeness	0.258**	0.347**	1	0.137**
Betweenness	0.497**	0.558**	0.137**	1

\*\* Correlation is significant at the 0.01 level (2-tailed)

University, the left from other institutions. All these authors and institutions play crucial role in forming and connecting subgroups.

Table 2 shows three centrality measures correlate with the rankings of author's productivity (i.e. the total number of papers published by the author), with degree centrality having the most significant correlation (0.870). The high correlation between author's productivity and centrality indicates that centrality measures to some extent be indicators for an author's scientific productivity.

### Component analysis

Consistent with a previous study in co-authorship network (Yan et al. 2010), the oncology co-authorship network is not an entirely connected graph, which can be divided into 4,145 components, then largest component of which contains 29,454 authors, about 58.10% of all the authors. Discipline like Biomedical may have bigger size of largest components,

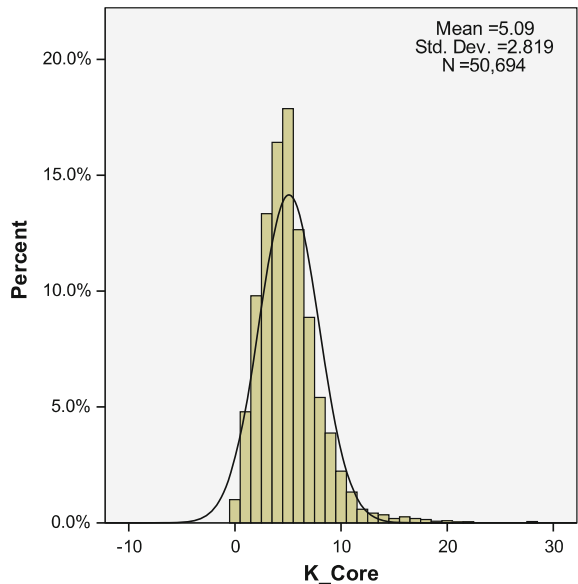
because more authors would be involved in an experimental research. The second largest component has 162 authors.

**K-Core**

K-Core distribution of the oncology authorship network is shown in Fig. 2. Most authors (90%) belong to small K-Core (smaller than 9), with 5-Core containing the most authors (9,059). The biggest K-Core has a degree of 28, in which all the 29 authors come from Cancer Institute and Hospital of Chinese Academy of Medical Sciences.

We remove the lowest 14-Cores from network. The remaining network, with 548 authors (about 1% of all the authors), breaks up into three relatively dense components (Table 3). The biggest component has 513 authors, who mainly come from the State Key Laboratory of Huanan Oncology, Cancer Institute and Hospital of Chinese Academy of Medical Sciences, Peking Union Medical College Oncology Hospital and Institution and Affiliated Oncology Hospital of Fudan University. The 18 authors in the second component all come from Anhui Medical University Affiliated Provincial Hospital. The 17 authors in the third component all come from Shanghai Jiao Tong University School of Medical Affiliated Ruijin Hospital. Then, each author in these components has at least 14 neighbors, so we consider them to be cohesive subgroups.

**Fig. 2** Distribution of K-Core



**Table 3** Component in 15-Core

Cluster	Freq	Freq%	CumFreq	CumFreq%	Representative
1	513	93.6131	513	93.6131	Zhongzhen Guan_Org1
2	18	3.2847	531	96.8978	Bing Hu_Org78
3	17	3.1022	548	100	Yu Zhu_Org21



## M-Slice

M-Slice distribution of the oncology authorship network is shown in Fig. 3. Most authors (92.25%) belong to M-Slices lower than four. The biggest M-Slice has a degree of 73, which has only two authors, coming from Tongji Medical College of Huazhong University of Science and Technology.

We remove the lowest 10-Slice from network. The remained 480 authors, about 1% of all the authors, break up into 43 components, most of which have only two or three authors (Table 4). Then, the line value in each component is bigger than ten, so we consider them to be cohesive subgroup.

## Hierarchical clustering techniques

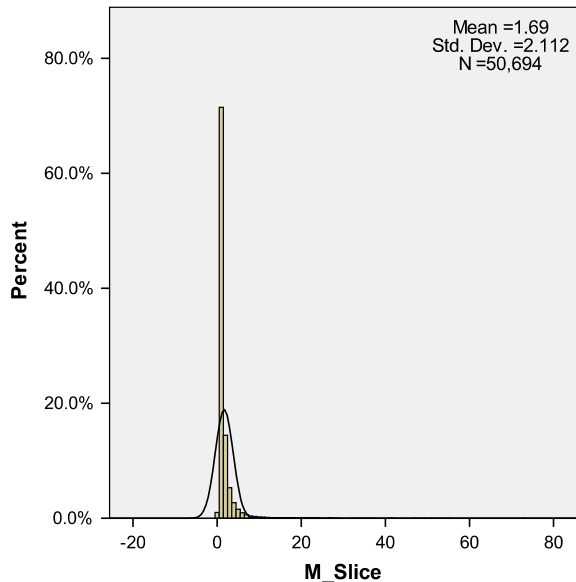
We select the 480 authors in 11-Slice mentioned above as a sample and use Euclidean distance indices to calculate the dissimilarity scores between each pair authors. Part of the dendrogram of the coauthor network is depicted in Table 5 and Fig. 4 (drawn with Pajek).

Twelve groups are forming according to the cluster analysis from bottom to top. Group 1, group 3, group 4, group7, group 8 all come from the State Key Laboratory of Huanan Oncology. Group 2, group5, group 10 and group 11 come from Cancer Institute and Hospital of Chinese Academy of Medical Sciences. Group 6 come from Affiliated Oncology Hospital of Fudan University. Group 9 come from Tongji Medical College of Huazhong University of Science and Technology. We find that the researchers with the same affiliation may be divided into different groups in cluster analysis, because they have different research topics.

## MDS

Figure 5 shows the MDS results of the 480 authors mentioned above. Again, we find that authors belonging to the same institutions tend to fall close to one another. Six groups are

**Fig. 3** Distribution of M-Slice



**Table 4** Components in 11-Slices

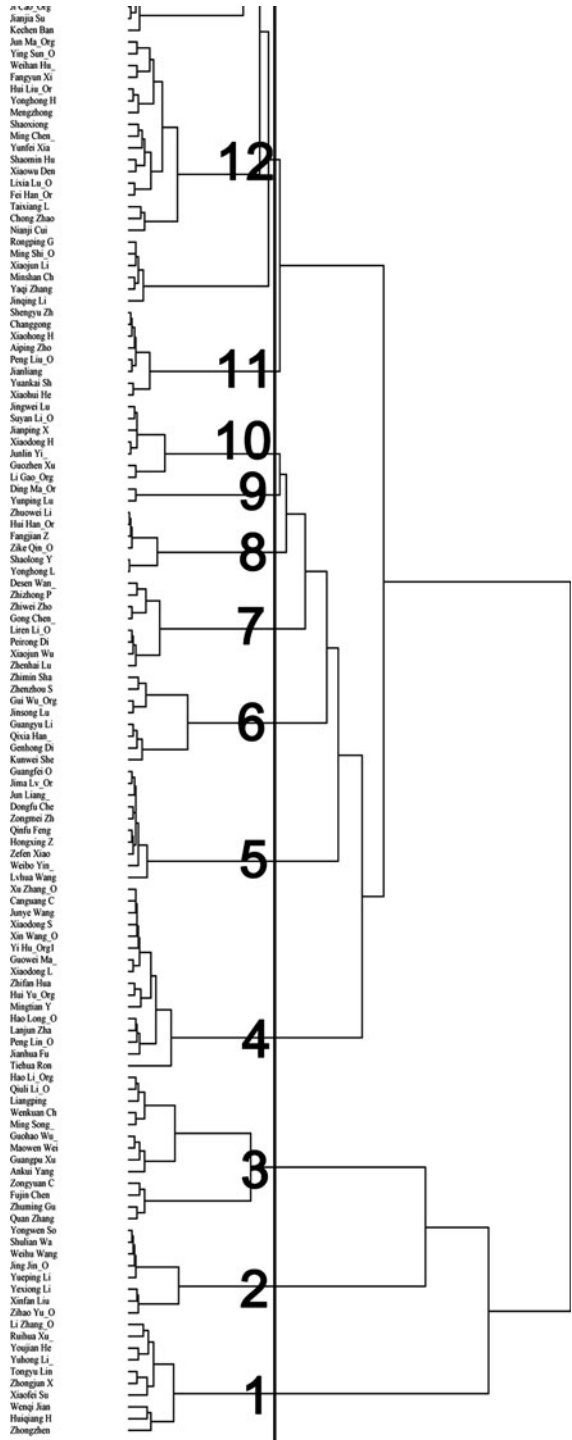
Cluster	Freq	Representative	Cluster	Freq	Representative
1	10	Jinming Yu_Org7	23	2	Baorui Liu_Org149
2	346	Zhongzhen Guan_Org1	24	2	Zixing Chen_Org33
3	21	Ding Ma_Org6	25	4	Zhengmin Yin_Org73
4	2	Mengchao Wu_Org39	26	2	Anmin Chen_Org6
5	3	Shu Zheng_Org9	27	4	Lin Shen_Org10
6	3	Caicun Zhou_Org62	28	2	Zhuchu Chen_Org42
7	2	Jia Cao_Org63	29	2	Sidong Xiong_Org37
8	4	Qiang Huang_Org72	30	3	Zhou Wang_Org47
9	2	Rongcheng Luo_Org20	31	2	Chunxu Hai_Org46
10	2	Shuyou Peng_Org9	32	2	Disheng Yang_Org9
11	2	Yutang Gao_Org18	33	2	Zhihai Peng_Org41
12	4	Enmin Li_Org35	34	2	Zenong Cheng_Org75
13	6	Mingqiu Liu_Org70	35	2	Caiping Huang_Org3
14	2	Tianhua Huang_Org35	36	2	Qian Huang_Org41
15	2	Guiyuan Li_Org42	37	2	Xiaochu Yan_Org60
16	2	Guisheng Qian_Org58	38	3	Xiaoqing Sun_Org130
17	5	Chunzheng Yang_Org2	39	2	Xuansong Cai_Org101
18	2	Huizhen Chen_Org17	40	3	Zuohua Feng_Org51
19	5	Shaokai Luo_Org29	41	2	Xinyong Luan_Org28
20	2	Longbang Chen_Org106	42	2	Wenxian Li_Org52
21	2	Hesheng Luo_Org38	43	2	Weilin Chen_Org183
22	2	Dechang Wu_Org30			

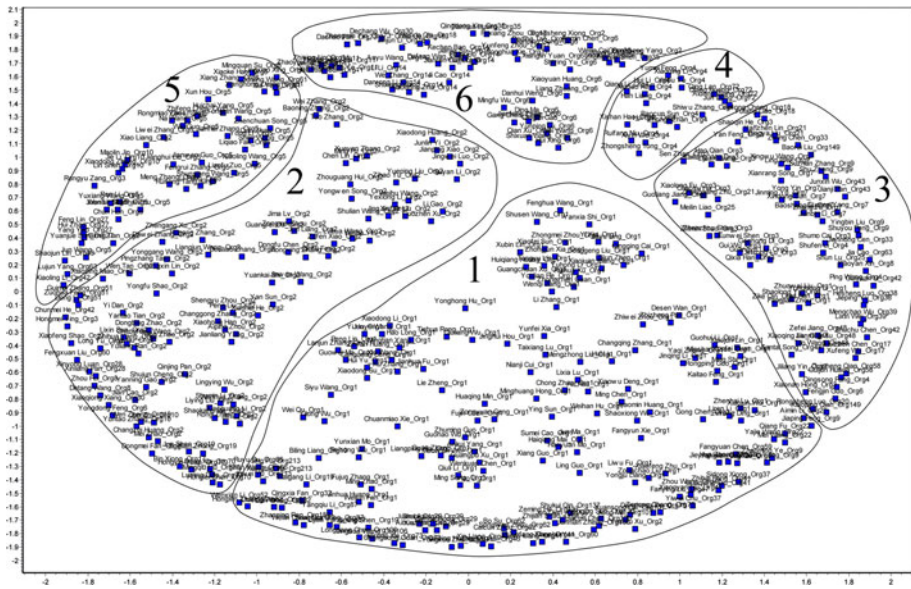
**Table 5** Groups in hierarchical clustering techniques

Cluster	Number of authors	Number of authors (%)	Representative
1	10	2.0833	Zhongzhen Guan_Org1
2	8	1.6667	Yexiong Li_Org2
3	13	2.7083	Zongyuan Ceng_Org1
4	16	3.3333	Tiehua Rong_Org1
5	10	2.0833	Lvhua Wang_Org2
6	8	1.6667	Zhimin Shao_Org3
7	8	1.6667	Desen Wan_Org1
8	6	1.25	Zike Qin_Org1
9	2	0.4167	Ding Ma_Org6
10	7	1.4583	Guozhen Xu_Org2
11	8	1.6667	Yuankai Shi_Org2
12	384	80	Jinming Yu_Org7

found: the No.1 group mainly come from the State Key Laboratory of Huanan Oncology, the No.2 group from Cancer Institute and Hospital of Chinese Academy of Medical Sciences, the No.3 group from Affiliated Oncology Hospital of Fudan University, the No.4 group from Tianjin Medical University Cancer Institute and Hospital, the No.5 group from

**Fig. 4** Dendrogram of the authors in 11-Slice





**Fig. 5** Groups in multidimensional scaling analysis

the Fourth Clinical Medical College of Hebei Medical University, the No.6 group from Tongji Medical College of Huazhong University of Science and Technology. We find that the researchers in the same group according to MDS may come from different institutions, because they have similar research topics.

## Conclusion

With data from 10 high ranked Chinese Oncology Journals published between 2000 and 2009, we construct the Chinese Oncology co-authorship network. Several approaches are adopted to analyze and extract subgroups from this network.

This study reveals that researchers from Beijing, Shanghai, and Guangdong cooperate most closely with each other and that the collaboration within ‘Administrative Divisions’ is closer than that between them. The collaboration intensity and activity, however, largely depends on the economic and educational level within or between ‘Administrative Division’. This finding suggests that we should encourage scientific cooperation among regions with different economic levels, thus contribute to economic progress in these underdeveloped areas and ensure equitable standards of research.

Regarding the personal influence, we find that there are 122 authors ranking on top 1% in degree, closeness, and betweenness centrality and they are central in the Chinese Oncology co-authorship network. This finding indicates that these top researchers are well situated in the network to transfer information more quickly and easily amongst themselves. We also detect the three centralities value has a higher correlation with authors’ productivity, that further indicates the centrality could be an index to evaluate authors’ productivity.

Based on the K-Core and M-Slice analysis about the research institutions, we find that the closer collaboration groups are among the State Key Laboratory of Huanan Oncology, Affiliated Oncology Hospital of Fudan University, Cancer Institute and Hospital of Chinese Academy of Medical Sciences, Peking Union Medical College Oncology Hospital and Institution, Tongji Medical College of Huazhong University of Science and Technology, No 1 Affiliated Hospital Oncology of Zhengzhou University, Peking University School of Oncology, Anhui Medical University Affiliated Provincial Hospital, Shanghai Jiao Tong University School of Medical Affiliated Ruijin Hospital, Shandong Provincial Oncology University, etc. These institutes function as “incubators” for Chinese Oncology research and are major players in this field. They deserve to obtain more government support and to attract more investment for cultivating excellence in the field of cancer research.

In this study we divide the co-authorship network into 4,145 components. We assume authors in one component form a research group, because every pair of authors can reach each other along the path between them. The component analysis in the Oncology network is quite similar with previous study, with the biggest component containing almost 58.10% of all the authors. The concept of component can be generalized to k-component: a maximal sub-network, in which each pair of vertices is connected by at least k distinct paths. Subgroups extracted with k-component analysis are more cohesive than those with component analysis.

In a K-Core, all vertices are connected by k or more other vertices. A K-Core identifies relatively dense sub-networks, so they help to find cohesive subgroups in the co-authorship network. The K-Core does not take into consideration line values, which are the basis of the definition of M-Slices. In a M-Slice, vertices are connected by line value of m or higher to at least one other vertex. The M-Slice is superior to K-Core when analyzing cohesive subgroups in the co-authorship network, since the co-authorship network is a valued network.

It is important to bear in mind that a K-Core or an M-Slice does not need to be connected. If we assume that cohesive subgroups are connected, we regard component within a K-Core or an M-Slice as cohesive subgroups rather than the K-Core or M-Slice itself. So we can apply component analysis in K-Core or M-Slice to find the cohesive subgroups, and that is what we do in this study. We found three and 43 components in 15-Core and 11-Slice respectively.

Hierarchical clustering is a useful technique when grouping a small number of vertices in a network. We applied this technique to the 480 authors in the 11-slice extracted from the original co-authorship network. They are clustered into 11 subgroups. The dendrogram helps to visualize the clustering process of vertices. Meanwhile, we divided the 480 authors into six groups with the MDS. We find that hierarchical clustering and MDS partition researchers into different groups according to their research topics rather than their affiliations, so it will be helpful for the management and construction of input-base research groups if we make an in-depth study of the groups generated by hierarchical clustering and MDS.

## Discussion

Having analyzed the results, we know ten major Chinese Oncology journals could not represent authors' overall productivity in the field of Oncology in China, so, in a way, we should say there may be a few disadvantages for the group research based on their outputs.

Otherwise, compared to their disadvantages, this network has a lot of advantages such as more practical, easily to handle, and the effectiveness, and will be more instructive for scientific research evaluation work. However, with a comprehensive analysis to the authors' productivity and centrality ranks, we identified a certain number of outstanding researchers in their field based on their contribution and publication and influence on Chinese Oncology research. Based on this network author rank, we could more easily choose principle investigators or academic leaders.

Furthermore, having studied the academic institutions, team structure, and membership of existing research groups in the co-authorship network, we can identify institutions with higher output to target for closer collaboration. It can help the authority to allot and distribute scientific research finance reasonably.

In addition, if analyzing the research subjects of research groups, we may easily to grasp the hotspot in oncology research field. Furthermore, these hotspots will help researchers apply for scientific projects and funds more easily.

In conclusion, this study advances the methods to analyze cooperative research groups from the Chinese Oncology co-authorship network, selects the most prolific research groups and individuals in Chinese Oncology research community, and provides basis for more productive cooperation in the future. This study also provides scientific evidence and suggestions for policymakers to establish a more efficient system for managing and financing Chinese Oncology research in the future.

**Acknowledgments** The research reported in this paper is done as part of the 'Management Research of University Fundamental Research Innovation System on First-class Young Scientists—with Biomedicine for Example' (No. 70773072), which is supported by National Natural Science Foundation of China. And it is also supported by Program for New Century Excellent Talents in University (NCET-08-0887).

## References

- Beaver, D., & Rosen, R. (1978). Studies in scientific collaboration. Part I. The professional origins of scientific co-authorship. *Scientometrics*, *1*, 65–84.
- Bordons, M., & Gómez, I. (2000). Collaboration networks in science. In B. Cronin & H. B. Atkins (Eds.), *The web of knowledge: a festschrift in honor of Eugene Garfield* (pp. 197–213)., ASIS Monograph Series Medford, NJ: Information Today Inc.
- Calero, C., Buter, R., Cecilia Cabello Valdés, C. C., & Noyons, E. (2006). How to identify research groups using publication analysis: an example in the field of nanotechnology. *Scientometrics*, *66*(2), 365–376.
- Castro, R. D., & Grossman, J. (1999). Famous trails to Paul Erdős. *MATHINT: The mathematical intelligencer*, *21*, 51–63.
- Chen, C., & Carr, L. (1999). Trailblazing the literature of hypertext: Author co-citation analysis (1989–1998). In *Proceedings of the tenth ACM conference on hypertext and hypermedia: returning to our diverse roots* (pp. 51–60). New York: ACM Press.
- Cohen, J. E. (1991). Size, age and productivity of scientific and technical research groups. *Scientometrics*, *20*(3), 395–416.
- Cronin, B., & Shaw, D. (2007). Peers and spheres of influence: Situating Rob Kling. *Information Society*, *23*(4), 221–233.
- González-Alcaide, G., Aleixandre-Benaven, R., & Granda-Orive, J. I. D. (2010). A study of the bibliometry and areas of the research groups of Archivos de Bronconeumología (2003–2007). *Archivos de Bronconeumología*, *46*(2), 78–84.
- Hou, H., Kretschmer, H., & Liu, Z. (2008). The structure of scientific collaboration networks in Scientometrics. *Scientometrics*, *75*(2), 189–202.
- Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*, 1–18.
- Liu, X., Bollen, J., Nelson, M. L., & Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Information Processing and Management*, *41*, 1462–1480.

- Melin, G., & Persson, O. (1996). Studying research collaboration using co-authorships. *Scientometrics*, 36(3), 363–377.
- Nascimento, M. A., Sander, J., & Pound, J. (2003). Analysis of SIGMOD\_s co-authorship graph. *SIGMOD Record*, 32(3), 8–10.
- Newman, M. E. J. (2001a). Scientific collaboration networks: I. Network construction and fundamental results. *Physical Review E*, 64, 016131.
- Newman, M. E. J. (2001b). Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64, 016132.
- Perianes-Rodríguez, A., Olmeda-Gómez, C., & Moya-Anegón, F. (2010). Detecting, identifying and visualizing research groups in co-authorship networks. *Scientometrics*, 82(2), 307–319.
- Scott, J. (1991). *Social network analysis: A handbook*. London: Sage.
- Smeaton, A. F., Keogh, G., Gurrin, C., McDonald, K., & Sodring, T. (2003). Analysis of papers from twenty-five years of SIGIR conferences: What have we been doing for the last quarter of a century. *ACM SIGIR Forum*, 37(1), 49–53.
- Subramanyam, K. (1983). Bibliometric studies of research collaboration. *Journal of Information Science*, 6, 33–38.
- Van Raan, A. F. J. (2008). Scaling Rules in the science system: Influence of field-specific citation characteristics on the impact of research groups. *Journal of the American Society for Information Science and Technology*, 59(4), 565–576.
- Vuckovic-Dekic, L. (2003). Authoship-co-authorship. *Archive of Oncology*, 11(3), 211–212.
- Wagner, C.S., & Leydesdorff, L. (2003). Mapping global science using international co-authorships: A comparison of 1990 and 2000. In: Guohua, J. et al. (Eds.), *In Proceedings of ninth international conference on scientometrics and informetrics* (pp. 330–340). Dalian: Dalian University of Technology Press.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. London: Cambridge University Press.
- Yan, E. J., Ding, Y., & Zhu, Q. H. (2010). Mapping library and information science in China: A co-authorship network analysis. *Scientometrics*, 83, 115–131.
- Yin, L., Kretschmer, H., Hanneman, R. A., & Liu, Z. (2006). Connection and stratification in research collaboration: An analysis of the COLLNET network. *Information Processing and Management*, 42, 1599–1613.
- Zulueta, M. A., & Bordons, M. (1999). A global approach to the study of teams in multidisciplinary research areas through bibliometric indicators. *Research Evaluation*, 8(2), 111–118.