Knowledge Subgroups in Virtual Communities Based on Affiliation

Zhihong Li[*], Yushan Wu and Cheng Yu

South China University of Technology, School of Business Administration, Guangzhou, China

Abstract: With the rapid development of virtual communities, lots of knowledge has been expressed or produced many times. This paper focuses on how to collect and organize the knowledge in communities, with the theory of social network. According to the existing data of affiliation, there are numerous relationships between the questions and their answers in the community. We take "Zhihu" for example, and build the net model, which could avoid the problem that the connection between notes is based on subjectivity. Meanwhile, based on the cohesion of social network, we put forward the concept about "knowledge subgroups". The result shows that as an effective analytical method, social network could explore the inner connection and similarity among the notes in knowledge network, thus achieve the aim of knowledge discovery and classification.

Keywords: Affiliation, knowledge subgroups, social network, virtual community

INTRODUCTION

Knowledge acquisition is one of important step in the knowledge management and application. Borgatti and Cross suggest that knowing, access, and cost are included in knowledge acquisition in their model [1]. Knowing means that knowing the one who processes relative knowledge. Formerly, most of the research was focused on teachers and professors [2], whose knowledge structure has been systematized. However, with the rapid development of Internet, a huge amount of information and knowledge has been created by User-generated Content (UGC) in virtual communities. Even though the knowledge is mostly dominant, there are differences in diversity and granularity [3]. It is how to acquire and clear up the knowledge created by virtual communities that is the main question in this paper.

Researchers acquire knowledge mostly by knowledge map or knowledge network. Knowledge map focuses on the position of knowledge, while the latter adds the connection among knowledge. However, the most connection has been linked by subjectivity on knowledge network [4]. We found that if we use the affiliation to build the knowledge connection, it will reduce the influence of subjectivity. Furthermore, we propose the concept about "knowledge subgroups" based on the cohesion of social network. We take the "Zhihu" for example, building the knowledge network based the affiliation and existing UGC, acquiring the knowledge in the virtual communities. Then we explore the knowledge subgroups to find the connection and the accordingly of knowledge classification in order to clear up the knowledge.

REVIEW

SNA (Social Network Analysis) has been increasingly utilized in the field of KM, especially in the shard, translation,

innovation of knowledge. The network cohesion is one of attributes in social network, which indicates the cohesion level of part of nodes relative to other nodes [5]. Xi Yunjiang had put forward the research on knowledge finding based on knowledge network in 2005 [2], which aggregated the knowledge points by searching the groups and subgroups of network. Zhang Wei researched the measurement of the knowledge sharing colleges' capacity of scientific teams, which made cohesive subgroups as an index [6]. "Knowledge recommendation based on social network theory" by Bei Fan, modeled the users search records into net graph, analyzed the centrality and cohesive subgroups of the network, and then accomplished the knowledge recommendation [7]. Xiang Xi-Yao researched the knowledge spillover by utilizing the conception of cohesive subgroups [8]. Yang Wan-Shiou researched cohesive subgroups from social networks for targeted advertising [9], meanwhile Comu S proposed the study on transactional memory system, also applied the method about cohesive subgroups [10].

As one of the important conception, affiliation could be ascended to the Social Circle Theory of Simmel [11]. Kadushin thought the social circles cannot be measured directly but by the behavior similarity of individuals and groups [12]. Nowadays, there are several aspects as follow referring the affiliation. First, affiliation network is a bipartite network. Second, it is consisted of actors' subsets rather than the relation of actors merely. These conceptions show that we can observe the relationship among individuals or groups by the relation between individuals and groups. The reason why this paper researches the classification of knowledge is as follow. First of all, Wasserman even point out that, when a pair of participants belongs to the same events, the probability that they will become membership is higher [13]. Likewise, if a pair of events attracts the same members to take in, it could conclude that the events are overlapped. Secondly, the relation of affiliation is more common in real society or virtual communities. At the same time, the data of affiliation is easier to obtain, which are all the advantages of affiliation network.

Affiliation networks are reflected by "Bipartite Network", which means the nodes of net graph have been separated into two subsets, actors and events, and only nodes of different subsets can be linked. Researches inland more focus on empirical study, such as "Study on the relationship of bloggers and commentators based on the social network analysis" by Qiu Jun-ping [14], "Study on absorptive capacity of employees in know ledge transfer network between university and enterprise" by Zhou Bo [15], and building network by means of affiliation in order to research the knowledge diffusions between the enterprises and institutions by Hu Ping [16]. Meanwhile, the researchers oversea more focus on the innovation of theories. R. Breiger published the paper titled "The duality of persons and groups" [17], and "Network analysis of 2-mode data" by S. Borgatti [18], discussed how to use the structural indexes to analyze the bipartite network.

Zhihu, as the example in this paper, is a Chinese social question and answer site. Lots of high-quality knowledge has been explicit or recreated in several years. If we can find the subgroups of events set from the view of social network, classifying the knowledge furthermore, it will play an important role in accurate searching and the recommendation of knowledge.

MODEL AND ASSUMPTION

Social network means modifying the conversation among individuals, making the person into the node of the network model virtually. If there are links between nodes, it proves that the persons corresponding the nodes are closed. Simmel, as a social expert, ever pointed out that we can call it "Social Circus" that the crowd has been get together around the organizations or events. The society may be regarded as a net structure interweaved by many social circus [11]. Corresponding the social network, many scholars have put forward the concept of knowledge network in area of knowledge management. Accordingly, we propose the "Knowledge Circus" and assume as follow:

Assumption 1: People have been gathered by some pattern of manifestation of knowledge, and we call it "Knowledge Circus" that the networks bond the knowledge by the medium of people.

Affiliation indicates the relationship between events and participators in social network. In theory, the formation of affiliation is the result of participators are interested in some events in a large extent. Virtual communities do not constraint the members of the network too much, so that people will take part in some events without much more external factors. In our research, the pattern of manifestation of knowledge is the questions arose by members in virtual communities. Meanwhile, interested in the questions then answering them is the reason of the aggregation of members. The reason of the interest is that members are confident of answering these questions, or enjoy participating in answering. Furthermore, there are similarity about the knowledge which people specialize in. Thereby, we assume as follow:

Assumption 2: The members answering some question is as the result of interest, that is to say, there are similarity in some way among these questions.

Cohesive subgroups could indicate the cohesion of social network. We need to divide the net into several subnets, making the similar nodes gathering and others separated. Cohesive subgroups are the subsets that they have steady, directed, dramatic, frequent, and positive connection each other, and the links of events or actors are concerned in subgroups of affiliation networks [19]. Thereby, we put forward the concept of "Knowledge Subgroups" then assume as follow:

Assumption 3: By means of finding the cohesive subgroups of knowledge networks, we call the results as "Knowledge Subgroups", which could be the basis of knowledge classification.

In order to reflect the affiliation of networks, many scholars use the method of bipartite network modelling. Bipartite network is the concept according to the 1-mode network, meaning that the nodes of network have been divided into 2 sets, sets of actors and events. Only nodes from different sets could they be linked. The actors participating into some events together often mean they are similar. Based on the quantities and strength of shared events, we could detect the similarity of members [19]. At the same time, if different events possess the same actors, it suggests that the events could be similar. Using this method and the assumption above, we build the research model as follow:

Based on the virtual communities of knowledge transition, building this network: the nodes are divided into two classes. One is question, and the other is member who answers the questions. Only nodes from different classes could be connected, and the condition of connection is that a member answers a question. Considering two different questions, the knowledge of questions will be similarity if their duplication is high. Furthermore, we could classify the questions of the network, and extract the knowledge from the classes, which would be the basis of knowledge classification.

METHOD

There are two method to explore cohesive subgroups by social network analysis. One is based on the number of adjacent nodes, which could be used in all kinds of nets. The other is based on the multi lines, which is better than the former since the multi lines ignore the individuality of nodes but emphasis the generality [19]. Meanwhile, the method of 2-mode network analysis is to transit 2-mode network into 1-mode network. The procedure is as Fig. (**1**).

In 1-mode network transited by 2-mode network, there are several multiple lines. Thereby, the method based on multi lines is more favourite in 2-mode network.

This paper focuses on undirected network. Wouter de Nooy even arose the decision tree of cohesive subgroups analysis [19]. We add the cluster analysis based on the former paradigm, because when going by the Nooy's method may gain the biggish subnet which goes against the analysis. Meanwhile, we do not use the cluster analysis at the very start in that the previous procedures could eliminate the alone or secondary nodes then remain the more important ones. It would receive the better result of analysis. Therefore, we put forward the procedures of exploring cohesive subgroups based on adjacent points as Fig. (**2**).
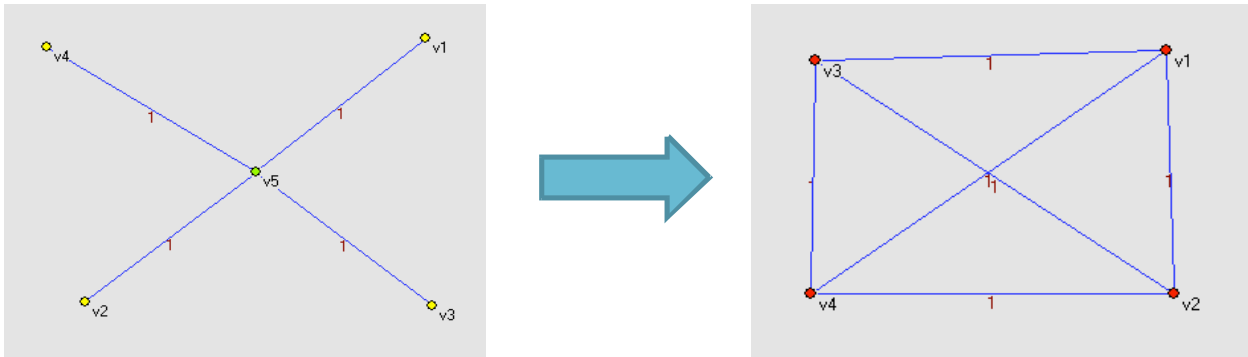
Fig. (1). The procedure of 2-mode network transiting into 1-mode network.

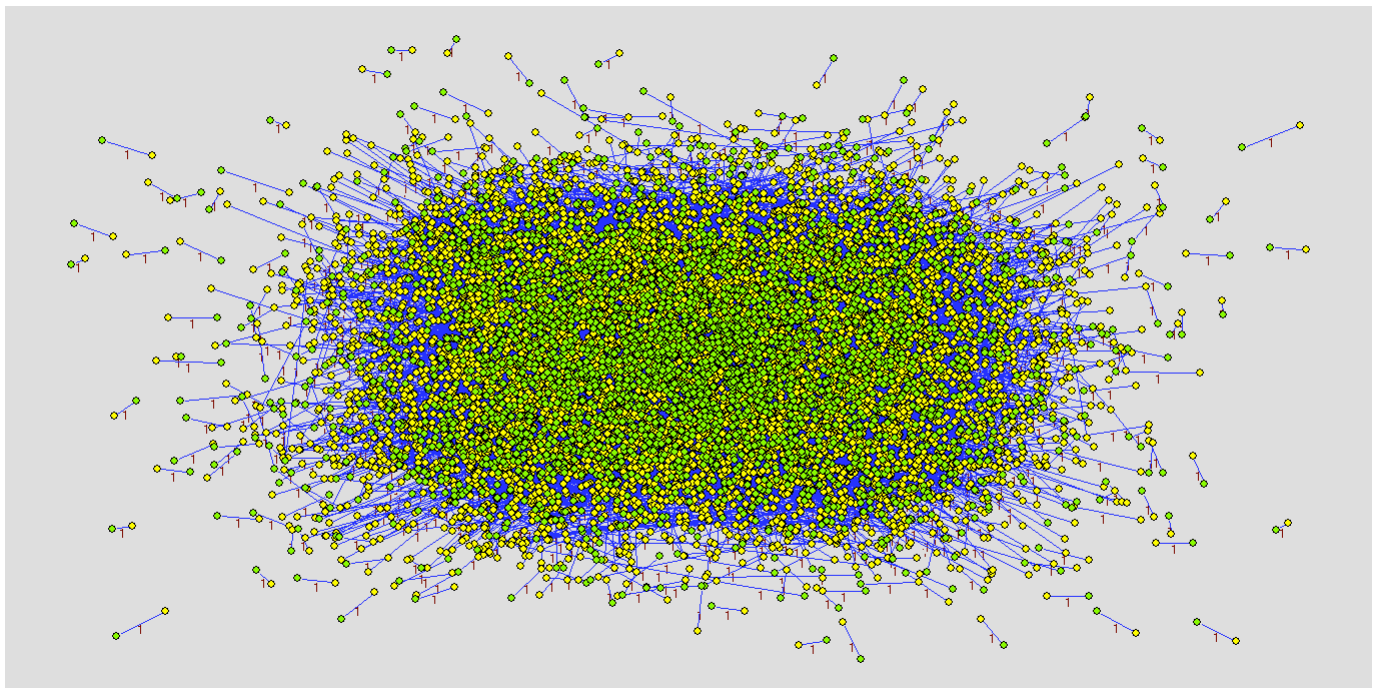**Fig. (2).** The procedures of exploring cohesive subgroups.

Fig. (3). The 2-mode network graph of affiliation data in Zhihu.

The definition of cohesive subgroups is increasingly strict from component, k-cores to clique. Component in undirected network means the biggest connected subnet. K-cores refers to the subnet that the degree of each node in subnet is not less than k. Meanwhile, the clique is the biggest complete subnet that is consist of 3 nodes at lease. The complete subnet means each node in subset links to the others directly [19]. Seeking out components, k-cores, or cliques do not mean the satisfied result, which would rely on the specific problem.

Cluster analysis is a familiar method in statistics. In social network analysis, we should calculate the dissimilarity values between nodes, in order to judge whether they could be in same class. Dissimilarity values are calculated by the network structure. That is, the more shared nodes they have, the more similar the nodes are. However, there is not a uniform standard that how many classes could be divided.

By Oct 26, 2014, we have collected 5481 questions, 8594 members and 15451 answers in topic "Douban" of Zhihu. It is worth mentioning that 2093 questions have no answers. In the meantime, we use Pajek2.0 as our analytical tool, which could transit 2-mode network into 1-mode network. The 2-mode network graph modelled by the affiliation data above is as Fig. (**3**), which the different colors mean the distinct data set.

There are two indexes in 2-mode network, scale of the events and participation of the actors. They measure the degree of events and actors respectively. We can get the distribution graph by data processing as Fig. (**4**).

As the graph above, the participation of the actors, meaning the answers of the members in the network, is power-law distribution, which is similar to former researches. It suggests that minor members will answer more questions, and
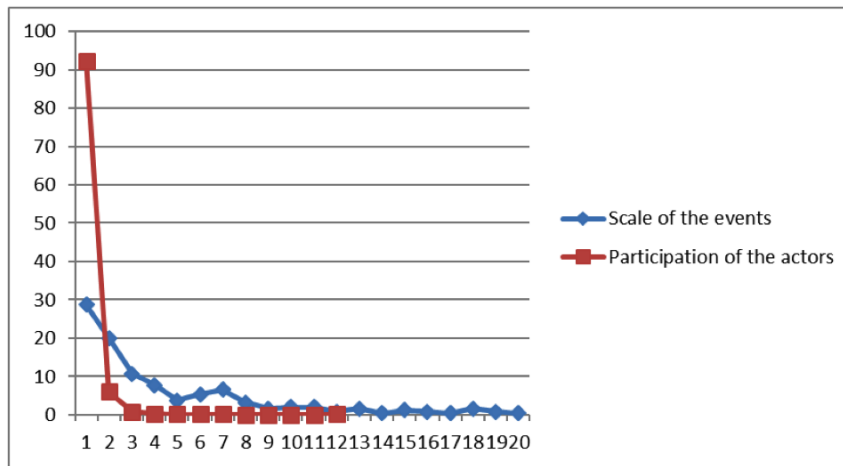
Fig. (4). The distribution graph of scale of the events and participation of the actors.
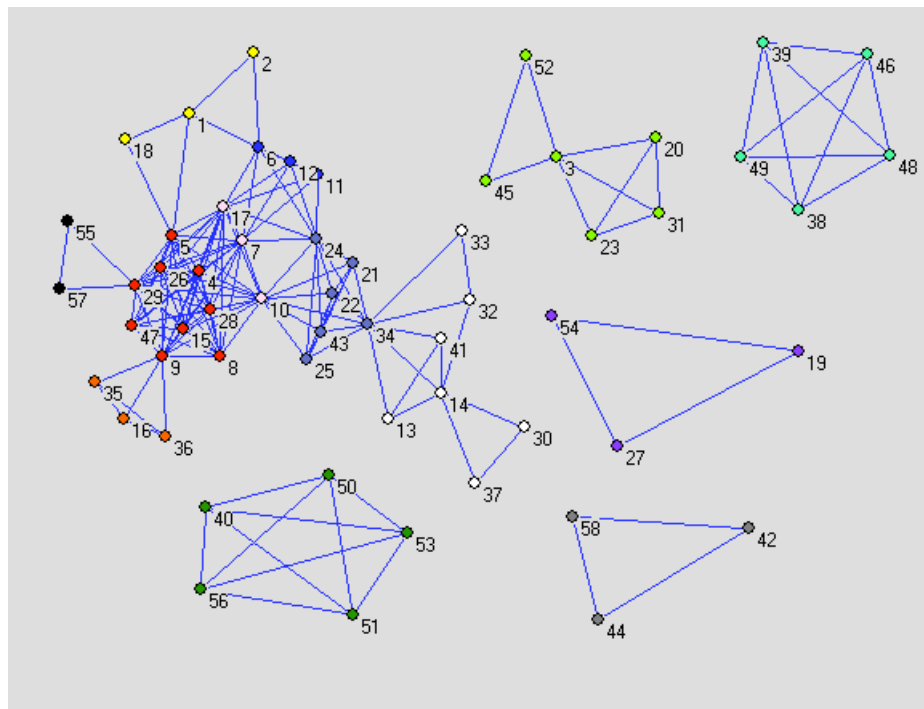

Fig. (5). The distribution of knowledge subgroups.

the major members will answer less. There are 7934 members answering only one question, accounting for 92.32% of the all. The most number of answers is 12, and only one member achieved it. Meanwhile, the distribution of scale of the events is near the power-law, though it is not too much obvious. There is only one question obtaining the 20 answers, as the most number of answers. 70.90% of all the questions gain the 5 or less answers, which means minor questions will be attractive and the major is left without anybody to care for them.

RESULT AND DISCUSS

We explore the cohesive subgroups of network by means of the method based on the number of adjacent points. First of all, we derive the 1-mode network from the 2-mode network on the events set. Secondly, we explore the subgroups. Finally, we gain the 58 more important questions, and the

result of analysis and classification of knowledge subgroups is as Fig. (**5**).

Knowing from the figure, the core network has been separated into 6 parts by the analysis of cohesive subgroups. We number every node, and the nodes' color means their classes. There are 5 components independently in the processing net graph, so that they belong to the same class severally. It is improper to make all of the nodes into the only one class, since the remaining component possesses too much nodes. As we all know, the major reason of classification is the dissimilarity value in cluster analysis [18]. Concerning about the practical situation, we determine that the value is 0.7, which means it will be dissociated when the value of the class is more than 0.7. So, we reach the table of knowledge classification and their dissimilarity values as Table **1**. Because of limitation of the length of paper, we have to list the classes and their questions partly as table 2, such as class 1,

6, 7, 9, 11 and 13, which are typical in our opinion. It could be concluded that the result of classification is basically satisfied according to the description of question.

Table 1.   Knowledge classification and dissimilarity value.

CLASS	DISSIMILARITY VALUE
1	0.130
2	0
3	0
4	0
5	0
6	0.420
7	0.440
8	0.170
9	0
10	0
11	0.530
12	0.340
13	0.670

Table 2.   Knowledge subgroups and question.

CLASS	NUMBER OF QUESTION	QUESTION
1	1	Where could we gain the latest and best film news?
	2	Are there nice/hit movies recently?
	18	What movies are a European version and Hollywood remake?
6	13	What is the basis of your grade for films in Douban or Mtime?
	41	Why the new vision of Douban is less humane?
	14	How could we contract Douban and Mtime?
	32	Why are people willing to write film review in Douban? Why are there lots of superior film reviews in Douban?
	33	Why are such many people fond of browsing the film reviews in Douban? What are the advantages of this reviews?
	30	How many films are there in the world up to now?
	37	Is it a bias that the understanding about the spirit of USA western films by Quentin Tarantino?

Table 2. Contd……

Class	Number Of Question	Question
7	20	Why is not there IMDb in China?
	23	What classic movies would reflect the Jazz era in USA?
	31	Will the level of film comments in Zhihu be superior Douban or Mtime?
	3	How could we improve Douban?
	45	Is there a kind of films trending to gain higher grade in Douban? Why?
	52	What movies have lower grade but are excellent in fact?
9	40	Are long film reviews all unreasonable?
	50	What do Shutter Island want to tell the audiences?
	51	Why do the US/UK TV series or cartoons have a higher grade?
	53	Inspection is ranking 15 in IMDb, is it too high?
	56	Why are there film reviews and short reviews in Douban?
11	21	Is Titanic 3D vision popular?
	22	Why do Titanic 1997 score 8.9 but Titanic 3D score 9.5 in Douban?
	25	How do CAM earn money?
	43	Why are not there 0 points in Douban?
	24	How often do you use Douban?
	34	Why do people grade the higher scores for books than films?
13	7	What is the site of Douban official Android App?
	17	Why are not there the reminding of movies being on in Douban Movies for iphone?
	10	Are prevue videos uploaded by Douban or cooperative partners?

This paper focuses on the application of knowledge management by means of social network analysis. As we discuss above, SNA could be an effective method, which may explore the inner connection and level of similarity in improving virtual communities nowadays. We can find the connection of knowledge and the core of network *via* building the affiliation network model then exploring knowledge subgroups. Zhihu is as our example. The final of analysis is mainly consistent with classification of Zhihu providing. Knowledge classification and discovery effectively will play an important role in knowledge matching and recommendation. In addition, we also summarize the universal procedure of exploring cohesive subgroups, then put forward the con-

ception of "Knowledge circus" and "Knowledge subgroups", replenishing the theory of area in knowledge management and social network combined.

Meanwhile, we must recognize that it could not draw a reasonable classification only rely on the social network analysis. First, this paper believes that a member answers different questions means these questions are similar, which is not thorough. Furthermore, the analysis method of 2-mode network is immature and limited. Last, we just take a static network as example, it is not sufficient. These shortages are all the emphasis of the next research.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

REFERENCES

[1]   R. Cross, and P. B. Stephen "A Relational View of Information Seeking and Learning in Social Networks," *Management Science Journal of the Institute for Operations Research & the Management Sciences*, vol. 49, no. 4, pp. 432-445, 2003.

[2]   Y. Xi, and Y. Dang, "The discovery and representation methods of expert domain knowledge based on knowledge network." *System Engineering*, vol. 23, no. 8, pp. 110-115, 2005.

[3]   X. Cao, G. Liu, X. Fu, and B. Li, "The mechanism and simulation of the knowledge integration process based on perspective of network," *Studies in Science of Science*, vol. 30, no. 6, pp. 886-894, 2012.

[4]   Y. Zhao, Z. Fan, and Q. Zhu, "Conceptualization and research progress on user-generated content," *Journal of Library Science in China*, no. 5, pp. 68-81, 2012.

[5]   N. E. Friedkin, "Structural cohesion and equivalence explanations of social homogeneity," *Sociological Methods & Research* vol. 12, no. 3, pp. 235-261, 1984.

[6]   Z. Wei, Q. Zhang, W. Shan, "Research on measurement of knowledge sharing capabilities of university research team based on overall network perspective," *Science of Science and Management of S. & T.,* vol. 10, pp. 170-180, 2012.

[7]   B. Fan, L. Liu, M. Li, and W. Yin, "Knowledge recommendation based on social network theory," *Advanced Management of Information for Globalized Enterprises, 2008. AMIGE 2008. IEEE Symposium on. IEEE*, vol. 1, no. 3, 2008.

[8]   X. Y. Xiang, H. Cai, S. Lam, and Y.L. Pai, "International knowledge spillover through co-inventors: An empirical study using Chinese assignees' patent data," *Technological Forecasting and Social Change,* vol. 80, no. 1, pp. 161-174, 2013.

[9]   W. S. Yang, and J. B. Dia, "Discovering cohesive subgroups from social networks for targeted advertising," *Expert Systems with Applications,* vol. 4, no. 3, pp. 2029-2038, 2008.

[10] S. Comu, J. Iorio, J. E. Taylor, and C. Dossick, "Quantifying the impact of facilitation on transactive memory system formation in global virtual project networks," *Journal of Construction Engineering and Management,* vol. 139, no. 3, pp. 294-303, 2012.

[11] G. Simmel, "The sociology of Georg Simmel," *Simon and Schuster,* 1950.

[12] C. Kadushin, "The friends and supporters of psychotherapy: on social circles in urban life," *American Sociological Review*, pp. 786-802, 1966.

[13] S. Wasserman, "Social network analysis: Methods and applications," *Cambridge university press*, 1994.

[14] J. Qiu, and W. Li, "Study on the Relationship of Bloggers and Commentators Based on the Social Network Analysis --Take the Web of Science Blog as Example," *Information Science,* vol. 7, pp. 959-963, 2012.

[15] B. Zou, B. Yu, and L. Bu, "Study on absorptive capacity of employees in know ledge transfer network between university and enterprise," *Studies in Science of Science,* vol. 29, no. 1, pp. 106-111, 2011.

[16] P. Hu, Y. Lu, and Y. Jiao, "Research on the knowledge diffusions between enterprises and institutions in information service industry cluster," *Studies in Science of Science,* vol. 31, no. 2, pp. 250-257, 2013.

[17] R. L. Breiger, "The duality of persons and groups," *Social forces,* vol. 53, no. 2, pp. 181-190, 1974.

[18] S. P. Borgatti, and M. G. Everett, "Network analysis of 2-mode data," *Social networks,* vol. 19, no. 3, pp. 243-269, 1997.

[19] W. de Nooy, A. Mrvar, and V. Batagelj, "Exploratory social network analysis with Pajek," *Cambridge University Press*, 2005.