The Influence Factors of Collective Intelligence Emergence in Knowledge Communities Based on Social Network Analysis

Zhihong Li[1], Ya'nan Xu[1], Kexin Li[2]

[1]School of Business Administration, South China University of Technology, Guangzhou, China
[2]School of Journalism and Communication, Shandong University, Jinan, China
Email: bmzhhli@scut.edu.cn, keepmovingxyn@163.com

Abstract

The advent of the age of Information shifts the environment we live in from off-line to on-line. The prospect of Collective Intelligence (CI) is promising. Based on this background, the aim of this paper is to discover the emergence mechanism and influence factors of CI in knowledge communities using the method of quantitative and qualitative analysis. On the basis of the previous research work, our model theorizes that the two dimensions of social network (*i.e.*, interactive network structure and participant's characteristics) affect two references of effectiveness (*i.e.* group knowledge production and participation of group decision). And this hypothetical model is validated with simulation data from "Zhihu" community. Our model has been useful since it offers some inspirations and directions to promote the level of CI in knowledge communities.

1. Introduction

With the advent of Web 2.0 era, different types of virtual communities are emerging increasingly. As a new generation formation of network community, Knowledge Community, like Quora, is characterized by knowledge sharing and exchanging, and always emphasized to motivate users to generate more contents by adding social features. Since it provides a platform for group cognitive, group collaboration and group decision-making, this kind of community is often rec-

ognized as a prototype of Collective Intelligence (CI) [1].

The related study of collective intelligence had mushroomed in the field of sociology in the past, which paid many attention to the content of collective intelligence existed in off-line organizations or enterprises. Psychologist Muller (1970) proved the important role of collective intelligence in education choice [2]. Surowiecki (2004) regards collective intelligence as a new management philosophy and management model, which sets off major changes in the enterprise management [3]. However, with the evolution of the Internet, a dramatic increase in users and easy access to online environment bring more complicated group of people. In addition, although the amount of online information received is extremely growing, the amount of information that an individual can handle is limited. Therefore, it is necessary to further study the collective intelligence in novel scenario, especially in knowledge community, due to the characteristics of collective intelligence has changed enormously.

There are many researches on collective intelligence of knowledge community. Tapscott [4] indicated that the Internet provides a platform for large-scale collaboration, and everyone can participate in the process of knowledge production. Therefore, the relevant studies on the emerging of collective intelligence mostly focused on the qualitative research, theories and methods of quantitative researches are few. Meanwhile, many knowledge communities emphasize its social attributes to motivate their continued participation. For example, people can easily communicate in the social network, which could promote the spread of knowledge [5]. Yang (2014) exploring the influencing factors of knowledge sharing among members in community by studying the social network structure in virtual community [6]. However, few studies have examined the sociality of knowledge community.

In this paper, we use a new idea that exploring the effect of group elements on collective intelligence by calculating the structure index in the social network method, to find group structure and member characteristics which are beneficial to the collective intelligence. Social network analysis is a kind of network method endowed with sociological significance [7]. So, we can explore the sociological connotations of collective intelligence by this method. At last, experimental data from "Zhihu" community is adopted to modified and verify the hypothetical model. The result of this study reveals the emerging mechanism and driving factors of collective intelligence, which provides important implications for improving the level of collective intelligence in knowledge community.

The remainder of the paper is organized as follows. Section 2 provides a brief literature review. Section 3 describes the hypothesis model and research methods. In Section 4, the verification process of proposed model from a sample database "Zhihu" is presented. And finally, we conclude the paper in Section 5.

2. Related Works

This chapter is an introduction to the relevant fields covered in this paper and a

review of the previous research literature. It mainly focuses on the following three aspects: collective intelligence, knowledge-based community, and social network analysis methods. In defining the knowledge community, we first introduce the evolution of Internet Web technologies and virtual communities, and the emergence of knowledge communities. Finally, the research methods and social network analysis used in this paper are summarized.

2.1. Collective Intelligence

Collective Intelligence is a process that a group of members use their knowledge to adapt the environment, put forward different views or methods, and eventually, give a better explanation or solution to the problem [8]. Based on the results of previous studies, we make a definition for collective intelligence which is individuals through participation and collaboration led to the emergence of collective intelligence, as well as the ability to solve problems [9] [10]. From this definition can be seen that the collective intelligence is more focus on a capacity, not simply a superposition of individual ability.

At present, there is little research on the measurement of collective intelligence, but more on the factors that influence the formation and development of collective intelligence. For example, Len Fisher argues that it could make better decisions in the community by drawing on mechanisms for coordinating of biological populations in nature [8]. Cai (2012) think that the collective intelligence will be affected by the following factors such as group size, membership heterogeneity, group cohesion, communication technology and group conflict [11]. Scholars believe that group size, membership diversity, independence and decentralization, and even technical factors have a significant impact on collective intelligence, but these factors are more like the case of abstract expression, there is no inclusion in a theoretical framework. Mačiulienė (1970) integrated the measurement of collective intelligence into three dimensions: 1) Capacity Dimension which is representing that the integration and creation of knowledge, group decision-making process. 2) Emergence Dimension which is representing that indicates the self-organization, adaptability and gathering effect. 3) Social Mutuality Dimension which is represent that the social impact of members and their social motivations [12]. These are some other methods in Table 1.

Table 1. Methods of research on collective intelligence.

Year	Scholars	Research Method
1991	Dorigo [13]	Ant Colony Optimization
2005	Wasko [14]	Social Capital Theory
2006	Jeppesen [15]	Statistical Analysis Approach
2010	Wagner [16]	System Modulation and Simulation
2011	Stefan Krause [17]	Social Experimentation

2.2. Knowledge Communities

The popularity of World Wide Web provides people with a more convenient platform for communication and sharing, it also boosts the flourish of kinds of knowledge communities. Howard (1993) thinks that knowledge community is social gathering place where people discuss public events, express emotions, and form online relationships [18]. The biggest characteristic of knowledge community is that people could break through time and space to communicate freely by the way of network interconnection.

There are few studies on the knowledge communities. Song (2015) use SNA (Social Network Analysis) to study social Q&A platform, which is applied to the network density, cohesive subgroup analyses, structural holes, the network centricity, has great reference value for this article [19]. Wang (2013) use quantitative analysis to study famous social Q&A platform Quaro, it reveal that three kinds of network structure in this kind of platform, which has great guiding effect on this research [20]. J. Koh & Y.G. Kim (2004) conducted a questionnaire survey and analysis of members of the knowledge community, the result is that knowledge sharing will stimulate more community members to participate in communication and increased the knowledge community members' loyalty to community service provides [21]. Alexander Ardichvili & Vaughn Page (2003) suggests that the willingness to share knowledge is due to community culture, self-interest, and trust among members [22].

Scholars have analyzed and discussed various factors that affect the knowledge community, but there is little research on the social attributes of the knowledge community. According to the research report of iResearch.cn, the focus of the development of the knowledge community has been on the SNS (Social Network Service), and user in the SNS website has a strong sharing. This article research the effect of social attributes on the emergence of collective intelligence in the knowledge community.

2.3. Social Network Analysis (SNA)

The SNA is the synthesis of many disciplines. It is the combination of social theory and application, formal mathematics and statistics, which is the combination of graph theory and sociology [12]. The most representative of the study is the theory of Granovetter (1973), in which he distinguished the strength of the connection and pointed out the importance of weak relationship [23]. Since the 1970s, in addition to the discussion of the method itself, the social network analysis also explored the special network forms of small groups, cohorts, social circles, and internal networks of organizations, market networks [24].

There is little research on knowledge community by using SNA. Jankowski-Lorek (2016) use social network modeling to study the behavior patterns of authors in Wikipedia, in order to provide insights into how Wikipedia can better manage false information and write conflicts in communities [25]. Yang (2014) explores the influencing factors of knowledge sharing among members in the

community by studying the social network structure in the virtual community [5]. Fu (2009) apply the network quantification method to analysis the social network structure of the virtual community, and the differences in network structure based on virtual relation and reality are discussed through the network density and centrality [26].

The reason that we use SNA is the social network attribute is a significant feature in knowledge community compared with the traditional Q&A platform, this paper mainly use the SNA to quantify the influencing factors of the collective intelligence. There are many excellent social network analysis tools, such as UCInet, Pajek, Gephi, which greatly facilitate the researchers, and quickly promote the application of network method in the field of sociology. In this paper, we use Pajek software to calculate link strength, cohesion, centrality and social capital in social network.

3. Hypothesis and Research Method

The content of this chapter is based on the theory of the predecessors, combined with the results of the case analysis of the community, to summarize the research framework. Secondly, the research hypothesis is put forward to determine the corresponding variables and observation indicators to explore the factors that influence the emergence of collective intelligence in knowledge-based community groups. Finally determine the indicator calculations and main research methods, and the analytical software tools used.

3.1. Research Theoretical Framework

The theoretical framework of this paper revolves around the emergence of collective intelligence. Emergence is considered to be a phenomenon related to complex systems, which refers to the appearance of new structures and new properties during the self-organization in a complex system [27]. On the construction of the theoretical framework, we combine Maizeiulienė's three dimensions of measuring collective intelligence into two dimensions: social network and the collective intelligence effectiveness. The dimension of social network is as factor variables in theoretical framework, and the collective intelligence effectiveness as outcome variables.

In the dimension of social network, this paper firstly makes reference to the research of peer production system, because it is closely related to the emergence of collective intelligence, which can be seen from the definition of peer production [28]. This article mainly use Zhu's definition about peer production: with the help of the Internet, people from all over the world use relevant tools to participate large-scale knowledge production activities [29]. Furthermore, the social network dimension is subdivided into interactive network structure and participant's characteristics. Therefore, this paper studies the influence of social network dimension on the collective intelligence effectiveness dimension, which can be decomposed into the influence of the interactive network structure and the participation's characteristics on the collective intelligence effectiveness.

The effectiveness of collective intelligence refers to the knowledge, ability, and psychological encouragement which are generated by group collaboration, so we use collective intelligence effectiveness to measure the power of knowledge production and participation of group decision. Knowledge production refers to the activities of create knowledge, not only the emergence of new knowledge, but also on the integration of the original knowledge and innovation [30]. Group decision-making has always been a popular research, and focusing on factors which affecting the quality of group decision-making. Our study adopted Lan's viewpoint that the reliability of group decision-making is proportional to the scale of group [8]. Therefore, the ability of the group's decision-making can be directly measured by the participation scale of group, that is, the increase of group size will improve the effectiveness of group decision-making results.

To sum up, the theoretical framework of the emergence of collective intelligence as shown in Figure 1, the arrows represent the relationship among the dimensions.

3.2. Research Hypothesis

According to the framework of the above research, the following research hypotheses are put forward from two aspects: the impact of the interactive network structure on the collective intelligence effectiveness, the influence of the participant feature on the collective intelligence effectiveness.

3.2.1. The Influence of Interactive Network Structure

In this paper, the interactive network structure variables are determined as follows: link strength, degree of cohesion and centrality [31]. Although these three variables are important concepts in social network analysis methods, the focus is different. The link strength mainly focuses on the number of connections in the network, and the degree of cohesion focuses on the distribution of the connections in the network [32]. In the case of similar number of connections, if the distribution of the connection is more concentrated, the overall cohesion of the
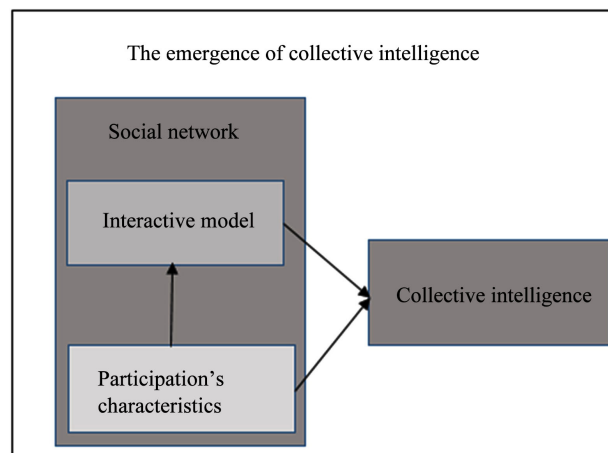

Figure 1. The framework of theory.

network will be reduced. The centrality is concerned with the degree of differentiation between the core nodes and the edge nodes in the network. The higher the value of the centrality, the higher the level distribution in the network [33].

The higher link strength between members of the group, the greater the likelihood of generating new knowledge. Chang (2011) think that when members of the virtual community are more connected, the community members will have a stronger knowledge-sharing willingness [34]. The Tsai *et al.* (1998) suggested that interaction and association among members had a positive impact on knowledge sharing [35]. Yli-Renko *et al.* (2002) argue that the more frequent communicate in the group, the more knowledge will be exchanged [36]. The node average value is the average node degree in the network, when the number of nodes in a network is constant, the more the network is connected, the stronger the degree of network connection is, and the average value of nodes in the whole network increases, the connection between nodes in the network will also increase [32]. So make the hypothesis:

H1: The link strength of network has a positive effect on collective intelligence effectiveness.

H1a: The scale of network connection has a positive impact on collective intelligence effectiveness.

H1b: The average value of network node has a positive effect on collective intelligence effectiveness.

In the past, the research on the degree of cohesion is mainly through the exploration of network cohesive subgroup [37]. Tirado experimented with 6 weeks and 10 comparative groups, it was concluded that the cohesive index had a positive effect on the social dimension and cognitive dimension of knowledge construction [38]. In addition, you can also use the average path length of the network to measure the degree of network cohesion. Strongatz (2001) proposed small-world network model that is defined by the average path length [39]. The shorter the average path length is, the closer distance between nodes is, the higher degree of network cohesion. The higher degree of cohesion, the higher the participation of community members, more knowledge will be generated which is conducive to the formation of collective intelligence. Hence the following hypothesis:

H2: The degree of network cohesion has a positive effect on collective intelligence effectiveness.

H2a: The number of network cohesive subgroups has a negative effect on collective intelligence effectiveness.

H2b: The average path length of the network has a negative effect on collective intelligence effectiveness.

In the past, the research on network centrality is to study the relationship between the center position of the network and knowledge transfer. Most research agree that the individuals are closer to the center of the network, the more positive their impact on knowledge sharing or transfer. However, there is little relevant literature on the impact of network centrality on the emergence from the

perspective of the overall central degree of the network. From the previous research, it can be concluded that network heterogeneity is the key factor for the emergence of collective intelligence. When the network is more evenly connected, the network structure is more flat, it means that the network is difficult to find significant impact on other members of the opinion leaders. The centrality of a network is usually quantified by three network-wide indicators: degree centrality, closeness centrality, and betweenness centrality. Hence the hypothesis:

H3: Network centrality has a negative impact on collective intelligence effectiveness.

H3a: The degree centrality has a negative effect on collective intelligence effectiveness.

H3b: The closeness centrality has a negative effect on collective intelligence effectiveness.

H3c: The betweenness centrality has a negative impact on collective intelligence effectiveness.

3.2.2. The Influence of the Participant's Characteristics

The reason for considering a participant in a community is that the level of collective intelligence may be different even for similar community network structures, because that the members who make up the community are different. To understand the members of the community, the most important factors are the individual attributes, behavior preferences and knowledge structure. As we know that the scale of group participation and the diversity of members are very important for the emergence of collective intelligence. As to the measurement of the diversity of group members, the mainstream view is that the dominant and recessive features of members should be measured. The dominant features include membership age, gender, education and occupational background, and recessive features include membership attitudes, values, knowledge and ability. Because this research is experiment in a virtual knowledge community, it is difficult to obtain the true dominant and recessive features of the members in the community. In addition, member networks cannot be constructed by capturing the mutual attention behavior of members because "Zhihu" protect the user's data. To this end, the membership network Mi is constructed on the basis of the member-question network Qi, and it is considered that there are interactions among the members if they answer the same question, and then from the member network Mi, based on the use of social network analysis of the island algorithm to measure the degree of diversity of community members.

Research on the other side of the participant is the demarcation of the core and non-core members of the community. The method is to find the core and non-core members of the community by exploring the membership network Mi. The core and non-core members are divided to assess the impact of social connections on collective intelligence. According to Gang Wang *et al.* (2013), because the core members have more followers and each core member's response

will push a message to their followers, their responses are higher which represent the core members tend to obtain more votes on their responses [20]. At the same time, Lan (2009) also believes that the effectiveness of group decision depends on whether the members are independent of each other [8]. It can be seen that social factors contribute to the prosperity of knowledge communities, but also inevitably have negative effects. In this paper, this effect is defined as the core member effect, and that is conducive to the improvement of collective intelligence in a community which core member effect is non-significant. Therefore, the following hypotheses were made:

H4: The diversification of community members has a positive impact on collective intelligence effectiveness.

H5: The network core member has a negative impact on collective intelligence effectiveness.

4. Network Model and Hypothesis Testing

4.1. Data Collecting

This paper chooses the "Zhihu" as the research scene, which has three kinds of network structure. These three network structure are the member of the social network, members concerned about their interested question which form member-question network, members concerned about their interested topic which form members-topic network. Zhihu will screen out the elite response for each question, which are the result of the group's efforts, and represent the integration and creation of knowledge. In some way, Zhihu reflect the collective intelligence, which encourage members to produce knowledge, and then through the collective voting select high-quality knowledge, also include interaction between members.

This paper uses the LocoySpider to obtain data of Zhihu community. In order to get more samples, we choose the secondary and tertiary sub-topics under the topic of "social science", "natural science" and "formal science". At last, we obtained 27 samples. The collected data includes topic number, the number of follower, the number of question and elite answer. As to December 8, 2015, a total of 1,246,830 questions were collected, the total number of elite questions is 15,852, and 31,003 community members participate in the elite question. The detailed data are shown in Table 2.

As can be seen from Table 2, with 50 answers only in the minority, but there are one-third questions that no answer. For example, the topic of "Communication studies", with 333 questions which contain 50 and above the answers, accounting for 0.89%. But there are 15,495 questions that no answer, accounting for 41.50%. It can be seen that considerable number of questions have not been adequately discussed in the Zhihu community. Take "Communication studies" as an example, the distribution of the answer shown in Figure 2. Among them, there are 15,095 questions answered by no one, less than 10 answers about 94.6%, and the highest the numbers of answers is 4466, which indicates that most of the answers are concentrated in a few questions.

Next, in order to explore the relationship between the number of follower and the number of questions, the sample is divided into three layers according to the number of follower, a topic of less than 10,000 followers, a topic of 10,000 to 100,000 followers, a topic of more than 100,000 followers. The Pearson coefficients of the two variables were calculated using SPSS software. The results are shown in Table 3.

Table 2. Data collection results overview.

Topic	Topic number	The number of follower	The number of question	The number of elite answer	The ratio of 50 and above the answers	The ratio of no answer
Behavioral sciences	19,589,457	493	11,412	1000	2.16%	33.04%
Library science	19,571,396	886	3057	370	0.29%	39.91%
Environmental studies	19,896,182	886	10,740	999	0.62%	46.64%
Sex	20,020,091	968	61,220	999	1.51%	34.13%
Theoretical computer science	19,772,617	1484	10,320	1000	0.47%	32.47%
Information science	19,626,490	1606	2004	234	0.25%	39.47%
Social work	19,766,153	1771	100	95	0%	67%
Pedagogy	19,627,046	3488	1744	264	0.86%	45.87%
Journalism	19,570,521	4452	4200	1000	0.6%	47.86%
Politics	19,591,014	5146	7037	999	1.42%	35.21%
Logic theory	19,615,566	6146	1360	355	0.37%	33.31%
Anthropology	19,574,199	8670	72,352	1000	1.58%	34.31%
Communication studies	19,567,245	10,022	37,340	999	0.89%	41.5%
Earth Science	19,589,333	13,622	18,240	1000	0.57%	32.85%
Statistics	19,558,740	18,380	13,940	1000	0.27%	41.42%
Geography	19,571,338	20,272	8940	1000	1.16%	26.58%
Principles of management	19,576,831	26,390	160,320	1000	0.64%	42.21%
Sociology	19,559,218	43,161	122,218	999	1.56%	37.26%
Law	19,604,890	47,430	77,460	999	0.87%	35.86%
Math	19,554,091	69,379	34,560	1000	0.6%	34.01%
Astronomy	19,563,625	91,685	10,580	1000	0.76%	21.75%
Finance	19,551,003	92,458	30,511	1000	0.42%	38.49%
Chemistry	19,562,906	530,707	29,600	999	0.52%	33.17%
Biology	19,575,492	710,579	145,140	999	0.74%	35.59%
Physics	19,556,950	724,148	66,600	999	0.6%	31.01%
Economics	19,560,170	2,456,636	43,580	999	1.17%	38.41%
Psychology	19,551,432	3,608,791	262,255	1000	1.24%	41.05%

Table 3. Topics focus on correlation between volume and issue number.

The number of follower	The number of topic	Pearson coefficient	Significant (two-sided)
Less than 10,000	12	0.339	0.282
Between 10,000 and 100,000	10	0.124	0.687
More than 100,000	5	0.776*	0.024

Note: * = 0.05 significance level.
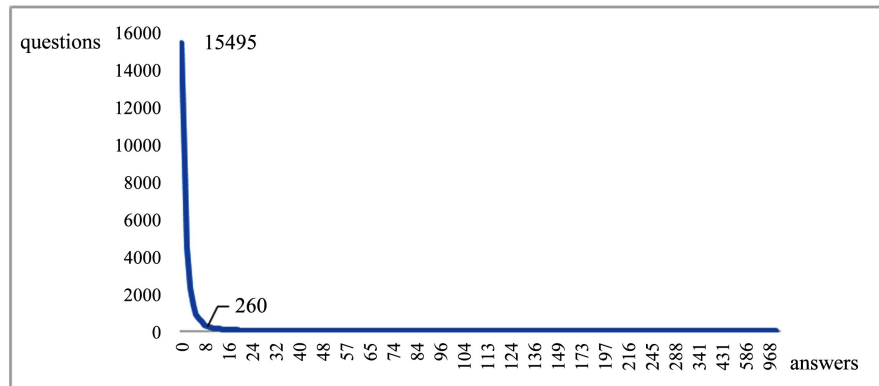
Figure 2. Topic "Communication Studies" answer weight distribution.

It can be seen that only the number of followers in more than 100,000 will appear significant correlation. This shows that the number of question will increase with the growth of attention to some extent. There are two points to explain above data. 1) Due to the LocoySpider can only get the first 50 answers for each question, so we only crawl on only the top 50 votes in the top answer. From the previous analysis, there are small proportion question that the number of answers exceed 50, and the number of votes after 50 votes received little votes. Therefore, it is not affect the later studies when we crawling only 50 answers. 2) Among all the questions on different topics, the proportion of unanswered questions is large. On the one hand because most of these questions have just been raised, no members have to answer, on the other hand because the question itself is little value. If these non-response questions are included in the network model, the network will lead to more missing values. The final decision is that only the elite answers as a basis for building the network model.

4.2. Network Model

Firstly, the Pajek is used to establish the bipartite network model Qi for each topic. Different topics are the boundaries of the network, a total of 27 members-question networks were established. For example, the network model "behavioral science" is shown in Figure 3. The network contains a total of 13,531 nodes, including 628 elite questions, with black nodes to represent; and 12,903 members, with a white node to represent. The network contains a total of 15,549 undirected connections, which represent those members to answer the question. And then use Pajek to calculate the structure index of each topic network. The results are shown in Table 4.
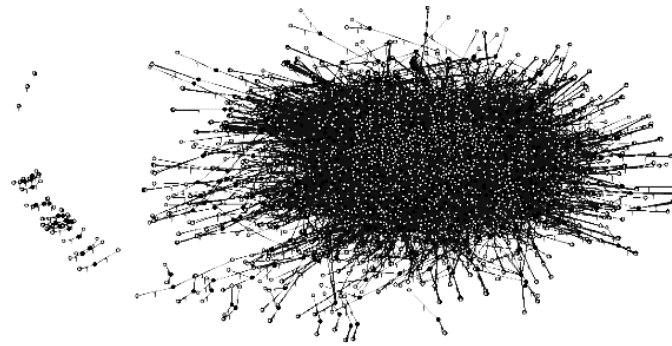
Figure 3. "Behavioral Science" topics bipartite network model Qi.

Table 4. Subject network Qi structure index results.

Topic	Participation scale	Maximum node degree	Average node degree	The number of subgroups	The average path length	Degree centrality	Closeness Centrality	Betweenness centrality
Behavioral sciences	12,903	26	2.30	36	6.61	0.0035	0.1354	0.0714
Library science	1899	22	2.01	80	8.75	0.0220	0.1477	0.5027
Environmental studies	9805	32	2.27	54	6.97	0.0046	0.1472	0.0876
Sex	18,196	39	3.05	2	5.42	0.0034	0.1653	0.0552
Theoretical computer science	6914	102	2.59	33	6.21	0.0130	0.2048	0.2438
Information science	1251	18	1.93	76	7.62	0.0327	0.1607	0.5644
Social work	547	11	2.17	6	6.29	0.0715	0.1633	0.4408
Pedagogy	1432	5	1.94	56	9.49	0.0309	0.1084	0.4488
Journalism	6192	16	2.15	99	7.43	0.0072	0.1179	0.0830
Politics	6688	41	2.34	84	6.53	0.0065	0.1576	0.1227
Logic theory	1392	35	2.08	46	7.32	0.0302	0.1693	0.3796
Anthropology	18,886	64	2.97	2	5.48	0.0031	0.1685	0.0548
Communication studies	16,599	20	2.37	17	6.56	0.0028	0.1291	0.0341
Earth Science	10,902	71	2.24	38	7.21	0.0059	0.1747	0.2423
Statistics	7521	28	2.46	38	6.72	0.0058	0.1423	0.0641
Geography	8889	26	2.16	93	8.05	0.0050	0.0951	0.0689
Principles of management	21,881	29	2.55	7	5.98	0.0021	0.1435	0.0290
Sociology	18,411	53	2.98	5	5.47	0.0026	0.1594	0.0390
Law	18,056	55	2.92	3	5.55	0.0028	0.1641	0.0457
Math	12,452	61	2.67	10	5.83	0.0044	0.1752	0.0929
Astronomy	8681	59	2.53	13	5.89	0.0060	0.1628	0.0768
Finance	11,086	37	2.48	21	6.34	0.0040	0.1486	0.0717
Chemistry	12,374	30	2.51	21	6.15	0.0036	0.1439	0.0442
Biology	21,966	78	2.59	9	5.87	0.0033	0.1848	0.1260
Physics	17,275	43	2.62	8	5.90	0.0026	0.1566	0.0509
Economics	17,109	81	2.72	8	5.65	0.0044	0.1873	0.1186
Psychology	20,706	67	3.01	4	5.48	0.0030	0.1688	0.0527

Collected data cover the whole topic in the community, which indicates that the sample is available. Among them, 14 sub-communities have more than 10,000 participants, and there are 18 answers on average, indicating that most sub-communities have a higher degree of participation. From the node degree, most of the average node degree of the network between 2 to 3, indicating that the average number of questions answered in three or less, and the maximum node degree value in the topic network has little effect on the average node degree. For example, in the "theoretical computer science" topic, although the some members answered 102 questions, the average value of the network is still 2.59. From the number of subgroups, it can be concluded that the number of subgroups decreases with the increase of participation scale. From the average path length, most of the network length is 6 or so, the overall average of 6.55. From the three indicators of centrality, the values are too small, especially degree centrality. Based on the above analysis, the maximum node centrality value has limited ability to reflect the overall situation of the network. Therefore, participation scale and maximum node centrality value are not included in the research hypothesis.

4.3. Define Participant Characteristics

4.3.1. Measurement of Member Diversity

On the measure of diversity of members, social networks island algorithm will be used. Island is an algorithm used to find the core network, firstly, the core part of the network is selected, and then the core nodes are divided into sub-networks. It can be seen that the nature of the island algorithm is to classify the members according to whether the members answer the same question. The subnets obtained by this method mean that there is a relative difference between members, and the number of categories represents the degree of the diversity of members. If the number of islands is larger, it indicates that the members of the community have higher degree of diversity.

In this paper, the process of measure final core members is as follows: 1) all members who involved the question are the participant; 2) based on the "member-question" network, the bipartite network model is established; 3) the core member are identified by using the social networks island algorithm; 4) find the important member whose answer has 5 or above votes; 5) find the intersection between the initial core member and the important member, which is the core member of the topic. The final result is shown in Table 5, and for the measurement of the diversity of the members, the number of network islands is main indicator. The more number of islands, the higher diversity of the members in the knowledge community.

4.3.2. Measurement of Core Member Effect

The measure of core member effect is achieved by the following two indicators. The vote ratio of core members is represented by μ_1 and A_i' indicates the number of votes about core member answer the questions. Correspondingly use

m represent the total response of the core member, A_j indicates the number of votes about each member answer the questions and *n* is that the total number of member, therefore, we propose the following equation:

Table 5. Definition of final core member.

Topic	The number network island	The number of initial core member	The number of important member	Important members as a proportion of all members	The number of final core member	Final core member as a proportion of the initial core members
Behavioral sciences	86	350	5684	44.1%	177	50.6%
Library science	77	491	507	26.7%	122	24.9%
Environmental studies	90	520	2578	26.3%	129	24.8%
Sex	65	152	9530	52.9%	129	84.9%
Theoretical computer science	56	216	1042	15.1%	32	14.8%
Information science	70	634	373	29.8%	174	27.4%
Social work	7	98	105	19.2%	14	14.3%
Pedagogy	46	361	592	41.3%	121	33.5%
Journalism	115	572	2148	34.7%	205	36%
Politics	92	398	3567	53.3%	200	50.3%
Logic theory	40	234	123	8.8%	17	7.3%
Anthropology	73	235	8513	45.1%	166	70.6%
Communication studies	114	377	6709	40%	266	70.6%
Earth Science	79	356	1381	12.7%	60	16.9%
Statistics	79	383	1020	13.6%	87	22.7%
Geography	115	514	1303	14.7%	74	14.4%
Principles of management	129	353	9614	44.7%	230	65.2%
Sociology	83	211	10,512	57.1%	160	75.8%
Law	83	210	8652	47.9%	157	74.8%
Math	57	155	2901	47.7%	70	45.5%
Astronomy	42	128	2319	44.5%	56	44.1%
Finance	80	286	3120	28.1%	138	48.3%
Chemistry	88	275	3472	28.1%	127	46.2%
Biology	98	243	4026	18.3%	79	32.5%
Physics	96	313	2757	16%	60	19.2%
Economics	74	184	6151	36%	123	66.8%
Psychology	90	219	10,760	52%	167	76.3%

$$\mu_1 = \frac{\sum_{i=1}^{m} A_i'}{\sum_{j=1}^{n} A_j} \tag{1}$$

The best ratio of core member represented by μ_2, N represent that the number of question which has highest responses of the core member, q represent the number of elite questions in the sub-communities, that is,

$$\mu_2 = \frac{N}{q} \tag{2}$$

Finally, μ_1 and μ_2 are summarized to get σ, and σ is normalized to get the score of core membership effect, that is,

$$\sigma = \mu_1 + \mu_2 \tag{3}$$

The results are shown in Table 6. The vote ratio of core member μ_1 mostly small, the largest 0.446, the remaining 17 sub-community value is less than 0.1, indicating that core members have limited influence; and the best ratio of core member μ_2 is also lower, no more than 0.4. The combination of μ_1 and μ_2, you can get each topic sub-community core member effect. The higher the core member effect value, the greater the impact of the core members on the other members in the sub-community. Due to the core membership effect of each sub-community is low, it can be concluded that the way of the community members select the optimal answer is more based on the answer content itself, instead of core members.

4.4. Collective Intelligence Effectiveness

According to the study of the knowledge community, the collective intelligence effectiveness can be embodied as two aspects: group knowledge production and group decision-making participation. In the aspect of group knowledge production, you can use the number of elite question and the cited frequency of best answer in the sub-community. And decision-making capacity can be measured by the decision-making frequency in the topic, that is, the total number of votes. The numerical value of collective intelligence production dimension index can be obtained by online collection or simple calculation directly, which the number of elite questions and the best answer cited frequency can be collected through the online, The average number of votes represented by λ, that is

$$\lambda = \frac{\sum_{j=1}^{n} A_j}{q} \tag{4}$$

The average votes of the question λ should be normalized, which is group decision participation score. And then integrated group knowledge production score and group decision participation score, ultimately obtain the collective intelligence effectiveness in Table 7. Among them, the higher the value of collective intelligence effectiveness, the higher the level of collective intelligence. In the selected samples, the topic of sociology is the highest in the community, while the topic of pedagogy is the lowest in the community.

4.5. Results

After the calculation of all the indicators in this study, the correlation analysis in SPSS was used to validate the hypothesis. The correlation analysis results are shown in Table 8.

Table 6. Core membership effect.

Topic	The vote ratio of core member μ_1	The best ratio of core member μ_2	Effect of core member
Behavioral sciences	0.066	0.126	0.166
Library science	0.083	0.213	0.260
Environmental studies	0.152	0.141	0.325
Sex	0.056	0.058	0.098
Theoretical computer science	0.011	0.029	0
Information science	0.446	0.368	1
Social work	0.013	0.200	0.131
Pedagogy	0.129	0.315	0.416
Journalism	0.098	0.165	0.251
Politics	0.186	0.240	0.457
Logic theory	0.043	0.191	0.176
Anthropology	0.125	0.123	0.266
Communication studies	0.111	0.125	0.243
Earth Science	0.058	0.118	0.146
Statistics	0.043	0.139	0.137
Geography	0.036	0.040	0.050
Principles of management	0.063	0.090	0.135
Sociology	0.082	0.081	0.160
Law	0.080	0.109	0.177
Math	0.100	0.142	0.236
Astronomy	0.149	0.229	0.385
Finance	0.091	0.170	0.242
Chemistry	0.096	0.165	0.247
Biology	0.059	0.074	0.116
Physics	0.020	0.035	0.020
Economics	0.122	0.148	0.279
Psychology	0.102	0.097	0.207

Table 7. The result of collective intelligence indicators.

Topic	The number of elite question	The cited frequency of best answer	Group knowledge production score	The average number of votes λ	Group decision participation score	Collective intelligence effectiveness
Behavioral Sciences	627	73	0.559	863.293	0.052	0.359
Library science	239	14	0.092	93.397	0	0.032
Environmental studies	665	72	0.569	707.089	0.042	0.358
Sex	465	68	0.465	12494.101	0.841	0.796
Theoretical computer science	712	88	0.672	359.152	0.018	0.408
Information science	163	13	0.055	319.693	0.015	0.018
Social work	55	11	0	746.909	0.044	0.002
Pedagogy	127	12	0.035	191.803	0.007	0
Journalism	496	77	0.525	657.567	0.038	0.328
Politics	629	77	0.580	338.097	0.017	0.349
Logic theory	199	16	0.086	230.593	0.009	0.034
Anthropology	675	48	0.449	8645.204	0.580	0.621
Communication studies	714	103	0.750	2037.651	0.132	0.529
Earth Science	720	147	0.980	647.394	0.038	0.614
Statistics	717	121	0.844	744.820	0.044	0.533
Geography	702	84	0.647	329.054	0.016	0.391
Principles of management	737	128	0.889	4462.018	0.296	0.720
Sociology	648	85	0.630	14844.185	1	1
Law	679	106	0.751	6238.353	0.417	0.709
Math	745	112	0.810	968.168	0.059	0.521
Astronomy	712	104	0.754	548.760	0.031	0.468
Finance	752	112	0.812	795.878	0.048	0.515
Chemistry	728	123	0.859	871.541	0.053	0.548
Biology	780	146	1	4851.342	0.323	0.806
Physics	752	129	0.900	2086.358	0.135	0.625
Economics	701	129	0.879	2234.832	0.145	0.619
Psychology	713	80	0.631	13768.482	0.927	0.955

Table 8. Observation in the model index SPSS related results.

The variable group	Pearson correlation between	Significant (two-sided)	Number of samples	Conclusion
Link strength → collective intelligence effectiveness	0.929**	0.000	27	Significant
The scale of network connections → collective intelligence effectiveness	0.928**	0.000	27	Significant
The average value of network node → collective intelligence effectiveness	0.873**	0.000	27	Significant
Degree of cohesion → collective intelligence effectiveness	0.724**	0.000	27	Significant
The number of network cohesive subgroup → collective intelligence effectiveness	−0.628**	0.000	27	Significant
The average path length → collective intelligence effectiveness	−0.758**	0.000	27	Significant
Centrality → collective intelligence effectiveness	−0.609**	0.000	27	Significant
The degree centrality → collective intelligence efficiency	−0.730**	0.000	27	Significant
The closeness centrality → collective intelligence effectiveness	0.292	0.139	27	Not significant
The betweenness centrality → collective intelligence effectiveness	−0.791**	0.000	27	Significant
Diversity of members → collective intelligence effectiveness	0.449*	0.019	27	Significant
Core member effect → collective intelligence effectiveness	−0.424*	0.027	27	Significant

Note: * = 0.05 significance level; ** = 0.01 significance level.

The result of correlation analysis shows that the hypothesis of this paper has been verified. In summary, the interactive network structure, participant's characteristics, and collective intelligence effectiveness in the social network are important to measure the emergence of collective intelligence, which has also been confirmed in previous studies. On this basis, this paper further analyzes the internal relations between these dimensions. Through the above experimental results, the topic sub-community with higher collective intelligence has the characteristics of high network connectivity and decentralization. Therefore, the conclusion of this paper is that the network structure of tight connection and uniform distribution would promote the emergence of collective intelligence, and the diversification and independence of community members leads to the formation of this network structure.

5. Conclusions

From the perspective of social network analysis, this paper studies the emerging mechanism and influencing factors of collective intelligence in knowledge community. Based on literature review and case study, this paper proposed the research model and used the social network modeling method to calculate the observation index and find the core members in the community. Finally, according to the conclusion of the experiment, we bring forward the network structure which is beneficial to the community collective intelligence, and summarize the emergence mechanism and influencing factors of the collective intelligence from the social network and the successful operation experience.

There are also the following shortcomings. It is necessary to further explore the relationship between variables, especially the variables in the social network dimension. The relationship between the variables and the indicators needs further demonstration. The relevant assumptions and indexes in the experiment process need to be optimized, such as the exploration process of network subgroups, the definition of community core members, the measurement of member diversity and independence. The sample data is slightly inadequate; the reliability of the model needs to be further tested.

In view of the above shortcomings, the proposed future research can be carried out in the following aspects. This article is only from the perspective of social network on the emergence of the impact of collective intelligence, this is not enough, and we should be combined with more research point of view. In the process of study, we improve the indicators of community network structure, member diversity and independence, collective intelligence efficiency, further optimize the research model, and deepen the interrelationship among influencing factors, to increase the amount of sample data to optimize the organization and cleaning of data to improve the credibility of the model assumptions.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Mattingly, K. and Ponsonby, A.L. (2014) A Consideration of Group Work Processes in Modern Epidemiology. *Annals of Epidemiology*, **24**, 319.
https://doi.org/10.1016/j.annepidem.2014.01.001

[2] Muller, P. (1970) Collective Intelligence Tests and Educational Selection-Neuchatel Experiences 12 Years Later. *Psychologie*, **29**, 194-202.

[3] Surowiecki, J. (2004) The Wisdom of Crowds: Why the Many are Smarter than the Few. Little Brown, New York, 10-20.

[4] Tapscott, D. and Williams, A.D. (2006) Wikinomics: How Mass Collaboration Changes Everything. Portfolio Hardcover.

[5] Rheingold, H. and Rheingold, H. (1993) The Virtual Community: Homesteading on the Electronic Frontier. MIT Press, Cambridge, 1114.

[6] Yang, M.W. (2014) Research on the Impact of Virtual Community Social Network Structure on Knowledge Sharing. Lanzhou University, Lanzhou.

[7] Borgatti, S.P., Mehra, A., Brass, D.J., *et al.* (2009) Network Analysis in the Social Sciences. *Science*, **323**, 892. https://doi.org/10.1126/science.1165821

[8] Fisher, L. (2009) The Perfect Swarm: The Science of Complexity in Everyday Life. Basic Books, New York.

[9] Gan, Y.C. (2004) Research on Knowledge-Building & Collective Intelligence in Virtual Learning Community: Perspective on the Integration of Knowledge Management & e-Learning. East China Normal University, Shanghai.

[10] Huang, X.B. and Zhou, Z.N. (2011) Realization of Collective Intelligence in the Environment of Web2.0. *Documentation, Information & Knowledge*, No. 6, 113-119.

[11] Cai, M.S. and Chen, S.J. (2012) On Influence Factors of Collective Intelligence from the Perspective of Reflection Sociology. *Academics in China*, No. 4, 23-30.

[12] Mačiulienė, M. and Skaržauskienė, A. (2016) Emergence of Collective Intelligence in Online Communities. *Journal of Business Research*, **69**, 1718-1724. https://doi.org/10.1016/j.jbusres.2015.10.044

[13] Colorni, A., Dorigo, M. and Maniezzo, V. (1991) Distributed Optimization by Ant Colonies. *Ecal91-European Conference on Artificial Life*, York, 20-24 July 2015.

[14] Wasko, M.M.L. and Faraj S. (2005) Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice. *MIS Quarterly*, **29**, 35-57. https://doi.org/10.2307/25148667

[15] Lars, J. and Lars, F. (2006) Why Do Users Contribute to Firm-Hosted User Communities? The Case of Computer-Controlled Music Instruments. *Organization Science*, **17**, 45-63. https://doi.org/10.1287/orsc.1050.0156

[16] Wagner, C. and Vinaimont, T. (2010) Evaluating the Wisdom of Crowds. *Proceedings of Issues in Information Systems*, **11**, 724-732.

[17] Krause, S., James, R., Faria, J.J., *et al.* (2011) Swarm Intelligence in Humans: Diversity Can Trump Ability. *Animal Behaviour*, **81**, 941-948. https://doi.org/10.1016/j.anbehav.2010.12.018

[18] Rheingold, H. (1993) The Virtual Community: Homesteading on the Electronic Frontier. MIT Press, Cambridge.

[19] Song, W. (2015) Social Network Analysis on Social Q&A Communities Illustrated By the Case of Zhihu. Southwest Jiaotong University, Chengdu.

[20] Wang, G., Gill, K., Mohanlal, M., *et al.* (2013) Wisdom in the Social Crowd: An Analysis of Quora. *Proceedings of the 22nd International Conference on World Wide Web*, Rio de Janeiro, 13-17 May 2013, 1341-1352.

[21] Koh, J. and Kim, Y.G. (2004) Knowledge Sharing in Virtual Communities: An e-Business Perspective. *Expert Systems with Applications*, **26**, 155-166. https://doi.org/10.1016/S0957-4174(03)00116-7

[22] Ardichvili, A., Page, V. and Wentling, T. (2003) Motivation and Barriers to Participation in Virtual Knowledge-Sharing Communities of Practice. *Journal of Knowledge Management*, **7**, 64-77. https://doi.org/10.1108/13673270310463626

[23] Granovetter, M.S. (1973) The Strength of Weak Ties. *American Journal of Sociology*, **78**, 1360-1380. https://doi.org/10.1086/225469

[24] Emirbayer, M. and Goodwin, J. (1994) Network Analysis, Culture, and the Problem of Agency. *American Journal of Sociology*, **99**, 1411.

[25] Jankowski-Lorek, M., Jaroszewicz, S., Ostrowski, Ł., *et al.* (2016) Verifying Social

Network Models of Wikipedia Knowledge Community. *Information Sciences*, **339**, 158-174. https://doi.org/10.1016/j.ins.2015.12.015

[26] Fu, L., Lv, B., Wu, Y., *et al.* (2009) Study on the Network Character of Relationship-Type Virtual Community. *Mathematics in Practice and Theory*, **2**, 119-129.

[27] He, X., Zhu, Y. and Wang, M. (2005) Knowledge Emergence and Complex Adaptability in Swarm Intelligence. *Information and Control*, **5**, 50-56.

[28] ATLAS Collaboration (2014) Measurement of Higgs Boson Production in the Diphoton Decay Channel in *pp* Collisions at Center-of-Mass Energies of 7 and 8 TeV with the ATLAS Detector. *Physical Review D*, **90**, 25-43.

[29] Zhu, T. (2011) Research on the Dynamical Mechanism of the Development of Peer Production-Baidu Encyclopedia Asa Example. South China University of Technology, Guangzhou.

[30] Fu, C., Qian, S., Chen, J., *et al.* (2009) The Literature Review on Knowledge Production. *Science & Technology Progress and Policy*, **2**, 155-160.

[31] Li, Z., Duan, Z., Chen, G., *et al.* (2010) Consensus of Multiagent Systems and Synchronization of Complex Networks: A Unified Viewpoint. *IEEE Transactions on Circuits & Systems I Regular Papers*, **57**, 213-224. https://doi.org/10.1109/TCSI.2009.2023937

[32] Wasserman, S. (1994) Social Network Analysis: Methods and Applications. Cambridge University Press, London. https://doi.org/10.1017/CBO9780511815478

[33] Scott, J. and Carrington, P.J. (2011) The SAGE Handbook of Social Network Analysis. SAGE Publications, New York.

[34] Chang, Y.P. and Zhu, D.H. (2011) Understanding Social Networking Sites Adoption in China: A Comparison of Pre-Adoption and Post-Adoption. *Computers in Human Behavior*, **27**, 1840-1848. https://doi.org/10.1016/j.chb.2011.04.006

[35] Tsai, W. and Ghoshal, S. (1998) Social Capital and Value Creation: The Role of Intrafirm Networks. *Academy of Management Journal*, **41**, 464-476.

[36] Yli-Renko, H., Autio, E. and Sapienza, H.J. (2001) Social Capital, Knowledge Acquisition, and Knowledge Exploitation in Young Technology-Based Firms. *Strategic Management Journal*, **22**, 587-613. https://doi.org/10.1002/smj.183

[37] De Nooy, W., Mrvar, A. and Batagelj, V. (2011) Exploratory Social Network Analysis with Pajek. Cambridge University Press, London. https://doi.org/10.1017/CBO9780511996368

[38] Tirado, R. (2012) The Effect of Centralization and Cohesion on the Social Construction of Knowledge in Discussion Forums. *Interactive Learning Environments*, **23**, 1-24.

[39] Strongatz, S.H. (2001) Exploring Complex Networks. *Nature*, **410**, 24-27.