

Research on the Internal Relationship Characteristics and Their Influences of Knowledge Sharing Multilevel Network in Q&A Community

Chen Xiaohui^{a,1} and Hu Ping^a

^a*School of Management, Xi'an Jiaotong University*

Abstract. The virtual communities have become the main position for people to create and share content in today's society. It not only realizes the dissemination of knowledge and information, but also promotes the formation of the relationship between users. The traditional related studies treat all information in Internet as knowledge, which deviate from the real situation. Therefore, this paper uses text classification technology to classify the answer texts under the topic of "English learning" in the "Zhihu" Q&A community, and extract the real knowledge under the topic. On this basis, a multilevel network about answer-users' knowledge sharing is constructed, and three subgroups with different users' node degree are divided. The multilevel network exponential random graph models are used to explore the influence of local structural characteristics formed by the relationship between users on the whole multilevel network. The results show that: When the node degrees of answer-users are small and the network structure is stable, the initiative of sharing knowledge is small and the homogeneity of knowledge content is high; if there are structural holes in the network, answer-users will create an obvious clustering effect, and the heterogeneity of shared knowledge is high; for the subgroup with the largest answer-users' node degree, the relationship between users is tight and the network structure is stable, then the shared knowledge is more heterogeneous.

Key words. Knowledge sharing; multilevel network ERGMs; text classification

1. Introduction

The transmission of information and knowledge resources based on the Internet has become a mainstream trend. In online Q&A communities, users can ask questions freely and other users will participate in answering questions, which realizes the transfer of knowledge and information. Users' activities may affect the complex information flow within the platform, thus forming a huge social network with levels of nesting. Based on the theories of social capital, social exchange and social cognition, many studies analyze the influencing factors of knowledge sharing in virtual communities [1],[2] and the factors that may affect users' motivation and willingness of

¹ Chen Xiaohui, School of Management, Xi'an Jiaotong University, Xi'an 710049, China; E-mail: 842430599@qq.com.

knowledge sharing in virtual communities [3],[4],[5],[6]. Moreover, the researchers also explore the factors that promote the knowledge sharing behavior of community members [7],[8]. It can be seen that the studies mainly from the perspective of "users" and pay less attention to the impact of local structure generated by the relationship between users on the whole network.

In addition, most of them ignore the interaction between different networks and treat all information as knowledge. Thus, this paper uses the text classification method to classify the answer texts from "Zhihu" Q&A community and extract the real knowledge. In addition, it constructs a multilevel network of answer-users' knowledge sharing and uses the multilevel network exponential random graph models (ERGMs) to explore the possible relationship characteristics created by different network levels' interaction and the influences of the local network structures represented by these relationship characteristics on the overall network.

2. ERGMs and Multilevel network ERGMs

ERGMs can investigate the influence of some local substructures on the whole network through the statistical analysis of network structures. It can draw a conclusion by comparing the statistical results of real network data with the expected values of the variables in the model simulation results [9]. At present, people have gradually extended this kind of researches to more complex fields, such as ERGMs for affiliation network [10] and multilevel network [11].

Multilevel network ERGMs examine multiple networks and the interaction effects between these networks. Due to the complexities and difficulties in this method, the main study is about the model specification and variable selection [12]. Recently, there are few studies apply this method to the real networks [13],[14]. It could explore the possible local structure variables exist in the multilevel network and the interaction effects among different network levels. Different from the traditional empirical research methods, it is an exploratory empirical study method, which helps us to discover the internal features of real networks. Therefore, we use it to achieve the research objectives.

3. Research data and method

3.1 Data

In this study, we use Python to crawl the relevant users' information, questions and answers of "English Learning" topic in Zhihu from 9th November of 2017 to 13th January of 2018. Finally, 6296 answer texts and 6296 answer-users are collected.

At present, there is no standard classification thesaurus for Chinese text classification. Therefore, based on the classification system of Zhihu platform, combined with the classification modes of Netease, Sina Weibo and other virtual communities, this study constructs the "English learning" answer texts classification vocabulary with 13 tags (Examination, Grammar, Word, Institution, Speaking, Pronunciation, Method, Translation, Materials, Software, Listening, Reading and

Writing). We use the sequence to sequence (Seq2seq) algorithm and combine with the Encoder-decoder module of LSRTM model to improve the algorithm, and finally get 3768 answer texts with tags.

3.2 Multilevel network construction

The multilevel network ERGMs include three levels, Network A and network B are two different levels of network, and the affiliation network generated by their relationship, namely network X, is the middle network [11]. The specific model equation is as follows,

$$P(A = a, B = b, X = x) = \frac{1}{k(\theta)} \exp(\sum \theta_Q Z_Q(a) + \theta_Q Z_Q(x) + \theta_Q Z_Q(b) + \theta_Q Z_Q(a, x) + \theta_Q Z_Q(b, x) + \theta_Q Z_Q(a, b, x)) \quad (1)$$

Where, $Z_Q(a)$, $Z_Q(b)$ and $Z_Q(x)$ represent the network effects of networks A, B and X respectively. $Z_Q(a, x)$, $Z_Q(b, x)$ and $Z_Q(a, b, x)$ are the interaction effects between networks A and X, networks B and X, and three networks A, B, and X, respectively.

A multilevel network about answer-users' knowledge sharing is constructed as follows: **Answer-users network (A)**: If two answer-users answered a same question, their matrix position in the network is marked as 1, otherwise it is 0. **Knowledge network (B)**: According to the results of text classification, the answer texts of users can be divided into 13 knowledge tags. If two different tags appear in a same answer, the matrix position in the network is marked as 1, otherwise it is 0. **Affiliation network (X)**: If one answer user's text contains one Knowledge tag, they are defined as related, and the matrix position is marked as 1, otherwise it is 0.

Network subgroup extraction: Due to the low density of the answer-users network (0.0035), it is difficult to analyze the overall network after text classification. Therefore, we use the node degree method in Pajek to partition the answer-users network. Moreover, the FR algorithm is used to visualize the network [15]. Then, we can extract three subgroups with 54, 66 and 116 node degree, respectively.

4. Model Estimation

Since one of the difficulties of multilevel network ERGMs is model convergence, this paper analyzes the possible network effects in the model and finally obtains the corresponding convergence model of three subgroups. Table 1 shows the parameter estimation results of each model. If the absolute value of the parameter is greater than 2 times of the standard deviation, it is considered to be significant.

Table1: Estimation results of three multilevel networks

Network level	Effects	subgroup 1	subgroup 2	subgroup 3
Network A	Cycle4A	--	0.0001*	--
	EdgeB	-8.7794	-0.0832	1.6366
	Star2B	1.8891	--	-0.2822
Network B	Star3B	-0.4082	--	--
	TriangleB	--	-0.2468	--
	ATB	--	--	0.3114
Network X	XEdge	-2.6599*	--	--
	XACB	0.0938*	-0.035*	-0.0283*
Network B&X	Star2BX	0.4375	--	--
	StarAB1X	-0.2728	--	--
	L3XBX	0.0029	--	--
	EXTB	0.0031	--	--

Note: “*” represents that the parameter estimation result of the variable is significant.

(1) Answer-users network (Network A): The results in Table1 show that under the premise of model convergence, no network effect of network A is found in subgroup1 and 3. This is because the network (A) density of the two subgroups is 1, which means that all nodes are fully connected and the network structures are stable. The network density of subgroup 2 is 0.5, there are unconnected nodes and the parameter estimation result of variable Cycle4A is positive significantly. This parameter reflects the clustering effects in network A. It shows that when there are unrelated nodes in network A, there will be more clusters in the multilevel network. In these small clusters, the relationship between answer-users will be closer, and the knowledge contribution behavior of answering-users will be affected by other users associated with them. Knowledge is transferred to each other in these small clusters to form a closure. Compared with other network structures, this closed structure can make users in the clusters receive more kinds of knowledge, and the transmission performance of knowledge in such clusters would be better. The results of subgroup2 show that when there are obvious structure holes in network A, the aggregation characteristics among answer-users in the multilevel network are more obvious.

(2) Affiliation network (Network X): Network X reflects the knowledge sharing behavior of answer-users. Table1 shows that XEdge participates in the converging of subgroup1 model and is negative significantly. It means that in the subgroup with the smallest node degree of answer-users, people are less inclined to share knowledge. According to social capital theory, the interaction between users has a positive relationship with the amount of knowledge sharing [3]. The parameter XACB indicates the homogeneity tendency of knowledge content shared by different answer-users. When XACB is positive and significant, the more answer-users share the same kind of knowledge. A negative XACB indicates that users pay more attention to the heterogeneity of content when sharing knowledge. Sharing original knowledge can improve users’ status and popularity in the platform, and the competition among similar users also makes the answer-users output different knowledge from others. According

to social network analysis theory, the characteristics of users in social network can be expressed as their status in the network, that is, the network reputation of users [16]. The higher the network reputation, the more popular the users are in the network, and the more other users have established relationships with them. For answer-users, sharing original knowledge can make them stand out in the platform and improve their popularity. The recognition of other users to these original knowledges will promote answer-users to get more attention, thus changing their relationship structures, enhancing their reputation in the network and gaining more authoritative network status. On the other hand, if answer-users want to improve their network reputation, they need to create original knowledge which is different from other competitors, improve the attraction of their output content, so as to obtain more recognition and attract more people to establish relations with them.

To sum up, in subgroup 1, the relationship between answer-users is less close than the other two subgroups, and users share less knowledge. At the same time, the knowledge sharing content of users in the corresponding multilevel network of this subgroup is relatively high. For subgroups 2 and 3, the users pay more attention to the heterogeneity of their knowledge output so as to get more attention in the platform. According to the results of network A, there are obvious aggregation characteristics among the answer-users in subgroup 2, and the knowledge heterogeneity shared by these users is higher, which once again confirms the relationship between the closeness of the answer-users and the heterogeneity of knowledge output.

If the absolute value of the goodness of fit t -value is less than 2, it means that the goodness of fit of the model is good, which can fully explain the sample network [11]. The results show that all variables of three models have passed the goodness of fit test, so the results of model estimation are reasonable and meaningful.

5. Conclusion

To sum up, the structural characteristics and relationship compositions of the answer-users network have a significant impact on the knowledge sharing multilevel network, which not only affects the enthusiasm of users to share knowledge, but also affects the heterogeneity of the shared knowledge content. Meanwhile, the shared knowledge content also could affect the relationship between answer-users. In the multilevel network, the structural characteristics and relationship compositions of answer-users network are closely related to the network composed of their shared knowledge. Firstly, relationship structures of answer-users network affect their enthusiasm to share knowledge. When there are fewer associated users in the network and the less edges between answer-users, the narrower the scope of knowledge transmission they share. Therefore, the less sense of achievement the answer-users can obtain from the network, and they are more inclined to output less knowledge. On the contrary, when there are more edges between users, they share knowledge more actively in order to obtain higher sense of achievement and higher network reputation. Secondly, when there are more clusters in answer-users network, these aggregated users respond more closely, and knowledge transferred better between them. However, the higher the homogeneity of shared knowledge content, the lower the probability of being noticed and recognized. In order to stand out from the competition, users tend to output heterogeneous content. In addition, the greater the degree of nodes in answer-users network, the wider the range of knowledge dissemination. Similarly, in

order to distinguish from other answer-users and gain higher reputation in the network, they are more likely to share knowledge with high heterogeneity. This paper extends the social network analysis about knowledge sharing in virtual communities to the multilevel network's extent. The relationship generated by answer-users and the content characteristics of knowledge are integrated to analyze the relationship between local substructures in different network levels and its impact on the whole multilevel network. The relationship characteristics and structural variables that affect the knowledge sharing of users are explored.

In view of the fact that this paper is an exploratory study, it still has some limitations. First, the research does not consider the users' attributes, and how different attributes affect the multilevel network. Second, it only selects the sample data under the topic of "English learning". Whether there are obvious differences in the network under different topics needs further analysis and research.

References

- [1] Li Y, Shan SQ, Li H, et al. An Empirical Study on Knowledge Sharing towards Emergency Events in Virtual Communities. *Management Review*. 2012, 24(11): 88-98+130.
- [2] Moghavvemi S, Sharabati M, Paramanathan T, et al. The impact of perceived enjoyment, perceived reciprocal benefits and knowledge power on students' knowledge sharing through Facebook. *International Journal of Management Education*. 2017, 15(1): 1-12.
- [3] Chiu CM., Hsu MH, Wang G. Understanding knowledge sharing in virtual communities: an integration of social capital and social cognitive theories. *Decision Support Systems*. 2006, 42(3): 1872-1888.
- [4] Chen IYL, Chen NS, Kinshuk. Examining the Factors Influencing Participants' Knowledge Sharing Behavior in Virtual Learning Communities. *Journal of Educational Technology & Society*. 2009, 12(1): 134-148.
- [5] Chang HH, Chuang SS. Social capital and individual motivations on knowledge sharing: participant involvement as a moderator. *Information & Management*. 2011, 48: 9-18.
- [6] Zhang X, Liu S, Deng Z, et al. Knowledge sharing motivations in online health communities: A comparative study of health professionals and normal users. *Computers in Human Behavior*. 2017, 75: 797-810.
- [7] Sun SY, Ju TL, Chung HF, et al. Influence on Willingness of Virtual Community's Knowledge Sharing: Based on Social Capital Theory and Habitual Domain. *International Journal of Human & Social Sciences*. 2010, (53): 142.
- [8] Jin JH, Li YJ, Zhong XJ, et al. Why users contribute knowledge to online communities. *Information & Management*. 2015, 52(7): 840-849.
- [9] Frank O, Strauss D. Markov graphs. *Journal of the American Statistical Association*. 1986, 81(395): 832-842.
- [10] Wang P, Sharpe K, Robins GL, Pattison PE. Exponential random graph (p^*) models for affiliation networks. *Social Networks*. 2009, 31: 12-25.
- [11] Wang P, Robins G, Pattison P, Lazega E. Exponential random graph models for multilevel networks. *Social Networks*. 2013, 35: 96-115.
- [12] Wang P, Robins G, Pattison P, Lazega E. Social selection models for multilevel networks. *Social Networks*. 2016, 44: 346-362.
- [13] Smith M, Gorgoni S, Cronin B. International production and trade in a high-tech industry: A multilevel network analysis. *Social Networks*. 2019, 59: 50-60.
- [14] Mcglashan J, Haye KDL, Wang P, et al. Collaboration in Complex Systems: Multilevel Network Analysis for Community-Based Obesity Prevention Interventions[J]. *Scientific Reports*. 2019, 9. doi:10.1038/s41598-019-47759-4.
- [15] Fruchterman TMJ, Reingold EM. Graph drawing by force—direct placement. *Software-Practice and Experience*. 1991, 21(11): 1129-1164.
- [16] Wasserman S. *Social Network Analysis: Methods and Applications*[M]. China Renmin University Press, 2004: 125-160.