Article ID: 1671-8860(2016)S1-0101-05 **Document Code**: A

Exploring the Relationship Between Information Characteristics and Users' Behavior in a Chinese Social Q&A Community: Evidence from Zhihu

CHEN Xiaoyu[1] DENG Shengli[2] GUI Sisi[1] FU Shaoxiong[1]

1 School of Information Management, Wuhan University, Wuhan 430072, China

2 Center for the Study of Information Resources, Wuhan University, Wuhan 430072, China

Abstract: This paper analyzed the information in one SQA community and drew a conclusion on the characteristics of user information seeking and adopting behavior. Zhihu was chosen as an example and information crawled on it was analyzed by statistics and content analysis method. This study found that the information distribution in Zhihu had the "Matthew Effect", that is to say, just a few questions had the dominant proportion of answers. In addition, this study found that there was significant correlation between questions with many followers and answers with many adopters which shed light on the influencing relationship between them. Furthermore, the information sought by users in SQA communities was related to their information needs and different needs would lead to seeking different types of information for them (i. e. expertise knowledge, suggestions for making decision). There were close relationship between information adopted by users and information quality, specifically, users tended to adopt information with knowledge, special ideas, language charm and stories. At the same time, the information adopted by users could also be regarded as the reflective criterion of information quality, that is to say, the more seeker searched the information in SQA, the better information quality the SQA had.

Key words: social Q&A; information characteristic; user behavior; Zhihu

1 Introduction

With the development of social networks, various social media tools and services have allowed people to ask for help from others. Social Q&A communities present a new information seeking environment that differs from traditional search engines in numerous ways, including a wide range of questions, the relative and unstable nature of information being exchanged, and the active constructive role of the users[1-4]. Social Q&A communities have made great impact on how people access and share information. They allow askers to pose questions in natural language rather than submit a few keywords to a search engine, as well as to receive personalized answers from other people instead of a list of results determined by an algorithm[1,5]. Over the past three years, new social Q&A communities based on online communities, which produced more high-quality answers and may have attracted more people to join, and Quora (www. quora. com) and Zhihu (www. zhihu. com) are two typical examples[5,6]. Zhihu is regarded as one of the representatives of fast emerging Chinese social Q&A communities in recent years and its relations and content operation have aroused popular attention, and many professionals, with similar or different interests and knowledge, participate in asking and answering questions of these communities [7].

Most currently research on information behaviors in social Q&A has focus on information seeking and sharing, while little is known about the relationships of users seeking, sharing behavior and user comments (response rate) , especially in Chinese ones[8]. This study examines the questions and answers distribution in a Chinese social Q&A community, and

Received on July 16, 2016

First author: CHEN Xiaoyu, postgradate, specializes in information behavior and knowledge management. E-mail: chenxiaoyu0928@126. com

Corresponding author: DENG Shengli, professor. E-mail: victorydc@sina. com

Foundation support: The Key Research Institutes of Philosophy and Social Science by Ministry of Education, No. 15JJD870001; the Wuhan University Academic Development Plan for Scholars after 1970s for the Project of "Research on Internet User Behavior"; the Wuhan University Youth Fund of Humanities and Social Sciences on "Research on Image Needs of Users Base on Social Questions and Answers"; the Luo Jia Youth Scholar of Wuhan University.

the possible correlation of users seeking, sharing behavior and user comments. The following research questions are listed in follows:

RQ1: What is distribution of questions and answers in a Chinese social Q&A community?

RQ2: Is there correlation between numbers of followers and numbers ofupvote?

RQ3: Is there correlation between numbers of response rate and answer type?

2 Related Work

One of the most convenience ways to seek for information from personal sources is to propose a question to them. Recent studies have begun to investigate the quality and characteristics of questions, answers and answerers in social Q&A communities[9-10]. Nam et al.[10] cite participation from users as the basis of the success of social Q&A sites, but identify the motivations for answerer participation to be "one of the biggest puzzles" in our understanding of these resources. Through interviews with users of the South Korean Q&A site Naver, the authors found that "by far, the most often stated reasons were the wish to help others, to learn and review material, or to participate as a hobby". In drawing upon both the work of social Q&A researchers and research on knowledge sharing in the business sector, Oh[8] outlines the following ten motivations for health answerers on social Q&A sites: self-enjoyment, self-efficacy, learning, personal gain, altruism, empathy, community interest, social engagement, reputation, and reciprocity. In investigating the types of sources that these (motivated) answerers draw upon to respond to questions, Fichman[11] find that human sources are the most common resources that are referenced.

With regard to how askers view quality and select a "best answer,"Jeon & Rieh[12] findings suggest that there is a preference for longer responses. They also found that the answers from a user with a good track record are more likely to be selected as the best, leading to questions as to whether it is the quality of the answer or the track record of the answerer that influence ranking. In an effort to determine the reasons for the selection of "best answer," Kim et al.[1] examined the comments that askers can post in response to the best answer. After eliminating comments that were a "simple expression of appreciations" from their sample, the researchers performed a content analysis of the comments in the context of the questions and the best answers. From this analysis, they developed a framework of best answer selection criteria that is composed of seven values: socio-emotional value, content value, cognitive value, extrinsic value, information source value, and utility. The researchers found that askers most frequently selected best answers based on the socio-emotional value they offered. In the comments to the best answerers, askers responded to the socio-emotional support value of the best answer by expressing an appreciation for the answerer's attitude, shared experience, and effort.

In our study, we regard the question type (number of followers) and answers type (number ofupvotes) as information characteristics, and in addition, we also regard the comments of each answer as the response rate. We are interested in their relationships and explore the influence of them.

3 Method

This paper chose the topic of "Wumai" (Haze in Chinese) in Zhihu as the research object and the data was collected from Zhihu. Since Zhihu does not provide API for researchers and developers, a web crawler in python was written for our data collection. The program conducted a top-down extraction, beginning with each questions under the topic of "Haze". For each question, the following information was extracted and recorded: URL (including questions, answers and answerers), numbers of user following, numbers of answers and numbers of up-votes in each answer, the title of each question and descriptive content of questions. All the initial information was saved in a MySQL database. The data collection completed on 13 April 2015. We also analyzed the information in EXCEL.

There are three types of information in the homepage: personal information, participation information, and feedback information. Personal information consists of education, work, location and career. Participation information refers to member's behaviors in Zhihu, including the number of questions asked, answer provided, edits made, and members followed. Feedback information refers to other members' response to a member's behaviors, like votes of usefulness or thanks for answers.

The study firstly focuses on the basic characters of the data, including questions, answers, best answers and their own distributions. Also we used statistical approach to analyze their possible correlation.

4 Results

The full data set contains 1540 questions with 2863 answers. We give each question and answer an ID according to their URL, e. g. if a question's ID is "22211349", (URL: http://www. zhihu. com/question/22211349). Among the 1540 questions, there are 636 questions with no answers and 44 questions without followers, accounting for 41. 3% and 28. 6% of total questions respectively. Table 1 shows a basic summary of the data.

Table 1 Data Summary

data collection time	2015-04-13
questions	1 540
answers	2 863
Question' ID of the most followers	22 211 349
Number of the most followers	2 781
Question' ID of the mostupvotes	22 854 438
Answers' ID of the mostupvotes	4 364 667
Number of the mostupvotes	2 630
Number of the followers of the mostupvotes	1 143
Number of questions with no followers	44
Number of questions with no answers	636

In addition, we also make a picture of the distribution of the numbers of answers in each question is showed in Fig. 1 As we can see, there are more questions with few answers than ones with many answers. That is to say, only a few questions received very high number of answers and many of the questions received either very few or no answers at all. It is called Matthew Effect and it also predicts that many questions inZhihu have not been solved. We analyzed the contents of questions without answers and finds that the reasons are as follows which are the same with the findings of Fichman[11] and Savolainen[13].

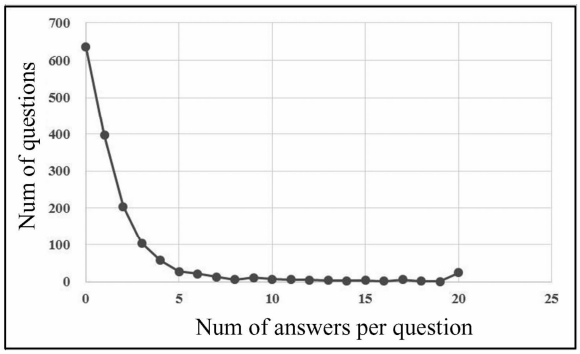

Fig. 1 Distribution of Answers and Questions

1) not clear, 2) too complex, 3) asking multiple questions, 4) lack of necessary information

There are 25 questions (1.6%, $n=1540$) are posted by best answers. Best answerers have moreupvotes (more than 100 upvotes) than ordinary answers (mean upvotes is 8.85). We calculated the comments in each answers, because comments can reflect the response rate to some extent, and ten we use Chi-square test to compare the response rate for best answers with ordinary ones. The results shown that the two kinds were significantly different ($\chi^2=4.318$, $p<0.001$).

Table 2 Answer Response Rate Between Best and Ordinary

	Subtotal	Get comments	No comments	Mean comment number	χ^2	p
Best answers	25	25	0	65.8	24.569	0.000
Ordinary answers	1 515	758	757	5.6		

Correlation test suggests that there is no correlation between numbers of followers and numbers ofupvote ($r=0.145$, $p=0.258$). However, there is strong correlation between numbers of upvote and numbers of comment. An alternative account for the correlation of best answers and response rate may be that, when answers is talked about by many people, the answerers maybe achieved a stronger sense to improve the answer quality related to visibility of expertise, gratitude, and feeling of making a contribution to promote the upvote of the answer.

It may be speculated the difference between best answers and ordinary answers is due to larger numbers of followers of the best answers' questions. A Chi-square analysis was conducted to investigate the relationships between response rates and information seekers' social resource including numbers of followers. The numbers of followers are divided into 4 groups of similar sizes at their quartile values. The results in Table 3 show that there are significant differences in response rates in relation to the information seekers' number of followers ($\chi^2=271.93$, $p<0.001$). Information seekers with more followers get higher response rates than those with less followers.

Table 3 Response Rate by Followers of Questions

Followers	N	Response	Noresponse	df	χ^2	p
0～10	1 346	327	1 019	3	271.93	0.000
11～100	157	123	34			
101～999	29	29	0			
1 000～2 781	8	8	0			

The number of followers can be an indicator for access: the more followers an information seeker has, the more widely his question can be seen and the more widely he can access. The number of followings can be an indicator for knowing: the more people an information seeker follows, the more she knows about the expertise people have. The number of messages can be an indicator for activeness in one's social network: the more messages an information seeker posted, the more active she is. The results confirm the "know" and "access" factor social information seeking.

Chi-square tests are conducted to examine the relationships between answers types (be invited vs. volunteer) and response rates. Table 4 shows the results.

Table 4 Response Rate by Answer Types

Type	N	response	noresponse	χ^2	p
Invited	936	742	194	984.85	0.000
Volunteer	1 927	1 164	763		

936 answers (32.7%, n=2 863) were posted by using the sentence like"Xieyao", which means "acknowledgement of being invited to answer this questions". Answers targeted to individuals get a higher response rate than questions posted to a general network. Chi-square test suggests that the response rate for questions targeted at individuals differ significantly with network questions (χ^2=984.85, p<0.001). This finding suggests that the way the social resources (followers and followings) in one's social network get used influences the outcomes of information seeking. When signaling the attention of particular users, those users may feel a stronger sense of obligation to respond and thus result in a higher response rate.

Furthermore, we chose the answers which number ofupvotes were beyond 100 as analysis objects (N=63), and we found these answers were from 25 related questions. We used correlation method to analyze the relationship between the followers of the 25 questions and the adopters of the 63 answers and we found that they have a strong correlation effect (r=0.746). This result (Table 5) showed that the information sought by users were often positive associated with the information quality.

Table 5 Number of Followers of Question and of Best Answers

Questions' ID	Number of best answers	Number of followers of questions
26093366	8	1 367
22211349	6	2 781
28489755	5	1 691
22863429	5	1 017
22857128	5	742
22854438	5	1 143
23360556	4	1 275
28437244	3	631
22201366	3	1 923
26885159	2	241
22853878	2	1 306
22466087	2	332
28476627	1	373
28450404	1	291
28449150	1	330
28446942	1	24
28444819	1	63
27720788	1	116
26472781	1	339
23105466	1	471
22846068	1	329
22375822	1	697
22241612	1	48
22211886	1	72
21752895	1	458

5 Discussion

The ever-growing number of questions and answers available on the Web increasingly overburdens Internet users and impacts their information seeking experience. Besides relying on traditional electronic information seeking resources, such as search engines, online library catalogs, and IMs or emails, people have increasingly turned to social Q&A communities to satisfy their information needs[14-15]. The relationships of users seeking, sharing behavior and user response rate are one of interesting research topic in social Q&As. In the Chinese social Q&A community—Zhihu, the characteristics of questions and answers can reflect users' information behavior to some extent. In this study, we mainly analyze the questions' followers, answers' upvote, answer types, and user response rate. The study can tell us some important results as follows:

1) Only a few questions received very high number of answers and many of the questions received either very few or no answers at all. It also shows that many questions in Zhihu have not been solved.

2) There is no correlation between numbers of followers and numbers of upvote. However, there is strong correlation between numbers of upvote and numbers of comment.

3)There are significant differences in response rates in relation to the information seekers' number of followers. Information seekers with more followers get higher response rates than those with fewer followers.

4) Answers targeted to individuals get a higher response rate than questions posted to a general network. Users who feel a stronger sense of obligation always respond and thus result in a higher response rate.

6 Conclusion

This study investigated the characteristics of the questions and answers in the Chinese social Q&A community--Zhihu, and analyzed the relationships between response rate and question types and answer types in this Chinese Q&A community. A sample of 1 540 questions and 2 863 answers were analyzed and we explored correlation of information characteristics and user information behavior in a Chinese social Q&A community. We conclude that the information characteristics (questions' followers, answers' upvote, answer types) can get used influences the outcomes of user information behavior (user response rate).

This study is a first step in examining information behavior in Chinese social Q&A communities. In-depth understanding of why and how users use social Q&A to seek for information and contributing knowledge will lead to better use and design of such communities[16].

References

[1] Choi E, Scott C R, Shah C. Effects of User Identity Information on Key Answer Outcomes in Social Q&A[C]. iConference,

in iConference, USA, 2013

[2] Kim S, Oh J S, Oh S. Best-answer Selection Criteria in a Social Q&A Site from the User-oriented Relevance Perspective[C]. The American Society for Information Science and Technology, USA, 2007

[3] Gazan R. Social Q&A[J]. *Journal of the American Society for Information Science and Technology*, 2011, 62(12): 2 301-2 312

[4] Choi N, Yi K. Raising the General Public's Awareness and Adoption of Open Source Software Through Social Q&A Interactions[J]. *Online Information Review*, 2015, 39(1): 119-139

[5] Guan X, Deng S. Understanding the Factors Influencing User Intention to Continue Contributing Knowledge in Social Q & A Communities[J]. *Chinese Journal of Library and Information Science*, 2013, 6(3): 65-76

[6] Chen X, Deng S. Influencing Factors of Answer Adoption in Social Q&A Communities from Users' Perspective: Taking Zhihu as an Example[J]. *Chinese Journal of Library and Information Science*, 2014, 7(3): 81-95

[7] Jin J, Li Y, Zhong X, et al. Why Users Contribute Knowledge to Online Communities: An Empirical Study of an Online Social Q&A Community[J]. *Information & Management*, 2015, 52(7): 840-849

[8] Oh S, Oh J S, Shah C. The Use of Information Sources by Internet Users in Answering Questions[J]. The American Society for Information Science and Technology, 2008, 45(1): 1-13

[9] Shah C, Pomerantz J. Evaluating and Predicting Answer Quality in Community QA[C]. Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, USA, 2010

[10] Nam K K, Ackerman M S, Adamic L A. Questions in, Kowledge in: A study of Naver's Question Answering Community[C]. The SIGCHI Conference on Human Factors in Computing Systems, USA, 2009

[11] Fichman P. Information Quality on Yahoo! Answers[C]. Approaches and Processes for Managing the Economics of Information Systems, Business Science Reference, Hershey, PA, 2014

[12] Jeon G Y J, Rieh S Y. The Value of Social Search: Seeking Collective Personal Experience in Social Q&A[J]. *American Society for Information Science and Technology*, 2013, 50(1): 1-10

[13] Savolainen R. Providing Informational Support in an Online Discussion Group and a Q&A Site: The Case of Travel Planning [J]. *Journal of the Association for Information Science and Technology*, 2015, 66(3): 450-461

[14] Lou J, Fang Y, Lim K H, et al. Contributing High Quantity and Quality Knowledge to Online Q&A Communities[J]. *Journal of the American Society for Information Science and Technology*, 2013, 64(2): 356-371

[15] Yan Y, Davison R M. Exploring Behavioral Transfer from Knowledge Seeking to Knowledge Contributing: the Mediating Role of Intrinsic Motivation[J]. *Journal of the American Society for Information Science and Technology*, 2013, 64(6): 1 144-1 157

[16] Jin J, Li Y, Zhong X, et al. Why Users Contribute Knowledge to Online Communities: An Empirical Study of an Online Social Q&A Community[J]. *Information & Management*, 2015, 52(7): 840-849