

Summer 7-5-2020

Study of Influencing Factors on Users' Knowledge Contribution Behaviors in Social Q & A Communities

Ying Zhong

Department of Information Management, Peking University, China

Shengli Li

Department of Information Management, Peking University, China, lishengli@pku.edu.cn

Follow this and additional works at: <https://aisel.aisnet.org/whiceb2020>

Recommended Citation

Zhong, Ying and Li, Shengli, "Study of Influencing Factors on Users' Knowledge Contribution Behaviors in Social Q & A Communities" (2020). *WHICEB 2020 Proceedings*. 6.

<https://aisel.aisnet.org/whiceb2020/6>

This material is brought to you by the Wuhan International Conference on e-Business at AIS Electronic Library (AISeL). It has been accepted for inclusion in WHICEB 2020 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Study of Influencing Factors on Users' Knowledge Contribution Behaviors in Social Q & A Communities

Ying Zhong¹, Shengli Li^{1*}

¹Department of Information Management, Peking University, China

Abstract: Social Q&A communities are important platforms for knowledge sharing among online users nowadays. Therefore, it is of theoretical and practical significance to understand the motivations behind users' knowledge contribution behaviors in social Q&A communities. Drawing on a dataset from Stack Overflow, one of the largest Q&A sites worldwide, this paper aims to explore the factors that may influence users' knowledge contribution behaviors. In particular, we examined factors related to three aspects: social interactivity, social capital and questions readability. Based on related theories we proposed several hypothesis and then tested these hypothesis using an econometric model. Our research results established the relationship between user's knowledge contribution behaviors and factors related to social interactivity, social capital and questions readability. This paper contributes to literature related to studies on Social Q&A communities.

Keywords: social Q&A communities; knowledge contribution behaviors; social capital

1. INTRODUCTION

The increasing size of Internet users and the rapid development of Internet technology, especially the Web 2.0 technology, have transformed the network information sharing model from a traditional one-way transfer model to a user-centric communication model that emphasizes collaboration and sharing. Social Q & A communities are booming as a result ^[1]. The social question and answer communities, or Q&A communities for short, are social media platforms where users seek information and share knowledge ^[2]. In particular, social Q&A communities provide platforms and tools for users to ask questions on various topics. Other users may provide answers to these questions and therefore knowledge sharing is completed among these users ^[3]. Social Q&A communities usually design various mechanisms to improve the efficiency of knowledge sharing ^[4]. For example, users' questions and answers can be rated by others to motivate them to provide high quality content. Users can make friends and interact with each other on social Q&A communities and thus social interactivity is an important feature of social Q&A communities. Reputation systems is another commonly used mechanism. Users can earn high reputation scores or medals by posting questions and answers.

Social Q&A communities have rapidly grown into popular knowledge sharing platforms with large number of active users worldwide ^[5]. For example, Stack Overflow was founded in 2008 and has become the world's largest English IT question and answer community. It has more than 11 million users worldwide and an average daily traffic of more than 10 million. As of October 2019, more than 16.5 million questions were answered on Stack Overflow ^[6]. Furthermore, Stack Overflow has well-designed systems to motivate users to contribute high quality questions and answers. That includes questions and answers rating systems, social interaction design and reputation systems mentioned above. Due to its significant influences and well-designed features, this research draws data from Stack Overflow to conduct research on motivations behind users' knowledge contribution behaviors in social Q&A communities.

One major motivation for most users in Q&A communities is to receive answers to their posted questions. Therefore, number of answers that one users' question receives can largely influence his satisfaction with the

* Corresponding author. Email: lishengli@pku.edu.cn

platform. To identify factors that may influence number of answers one question receives is thus an important task and has become a popular research topic in literature. For example, in one previous research on the answerability of questions in Stack Overflow, it was found that the level of detail, specificity, and clarity of questions, as well as the popularity and participation of questioners, have an impact on whether questions are answered ^[7].

However, due to the fast innovations of social Q&A communities, many new features have not been studied, leaving rooms for further studies. To fill the gap in literature, this paper studies factors from three aspects: social interactivity, social capital and questions readability. Through an econometric model, we established the relationship between these factors and number of answers users' questions received. This paper has significant theoretical contribution to literature related to studies on social Q&A communities. Some findings are quite new. For example, we find that social interaction between questioners and responders (i.e., the number of responses from the questioner to the responder) positively increases the number of questions received. We related these new findings to theories and contribute to the understanding of the motivation behind users' knowledge contribution behaviors in social Q&A communities.

2. LITERATURE REVIEW

As a new type of knowledge sharing platform, social Q&A communities has achieved great success in helping users solve their information needs. However, to increase the response rate of questions still poses a huge challenge to social Q&A communities. It is therefore a popular research topic to identify factors that influence number of answers that a question receives. Previous studies are conducted from different perspectives. Some scholars examined factors that related to users. That is, to identify factors that influence one user to provide or receive answers. For example, Zhe Liu and Bernard J. Jansen's research on question recommendation from the perspective of users' willingness to answer reveals different roles of knowledge sharers and consumers in the Q&A community ^[8]. Based on this finding, scholars first identified knowledge sharers from non-sharers in order to improve the response probability of respondents.

Some scholars have pointed out that compared to social media such as Twitter and Facebook, the online social Q&A community focuses on knowledge contribution and dissemination, rather than information, news or commentary on an individual's current mood^[9]. A study by Zhang, Ackerman, and Adamic (2007) found that more than half of users on Yahoo! Answers usually only ask questions without answering them^[10]. This shows that in the Q&A community, users' willingness to contribute knowledge needs to be strengthened. Therefore, many scholars have also carried out rich researches on the influencing factors of users' knowledge contribution behavior in the community. For example, Liu Zhe and Bernard J. Jansen predicted the probability of answering a question on Sina Weibo, from the perspective of question-related features and questioner-related features. In particular, they selected question-related features including the length of the question, the uniqueness of the question, the urgency of the question, and the issue time of the question, and related features of the questioner. Using machine learning methods, the probability of receiving a response to a question posted on Weibo is predicted. They found that the probability of answering the question was more related to who the questioner was (user activity, posting style, and historical interaction), but less related to the characteristics of the question (amount of information, attractiveness, urgency, courtesy, and posting time) And subject ^[11]. Another research found that the questioner's popularity, participation, and questioning time have important effects on the response to questions ^[11]. On the question side, the level of detail, uniqueness, and clarity of the question will also affect the response.

Mechanisms design such as reputation system and social interaction design are important features of social Q&A communities. Therefore, it is also crucial to study their influences on users' knowledge contribution behaviors. Such research is yet rarely seen in literature. Some scholars have studied several related topics. For example, Zhe Liu and Bernard J. Jansen found that factors such as user participation, interest, and relevance

have a significant impact on users' willingness to answer in a social Q&A community^[11]. Studies by Jiahua Jin et al. also found that user self-presentation, peer recognition, and social learning have a positive impact on their knowledge contribution behaviors and willingness to answer^[9].

In summary, to identify factors that influence knowledge contribution behavior of users in social Q&A communities is an important research topic. However, the influence of factors that related to mechanisms design such as reputation system and social interaction design are yet to be examined. Our paper contributes to the literature by exploring the effects of these newly designed mechanisms along with question-related factors.

3. RESEARCH MODEL AND HYPOTHESES DEVELOPMENT

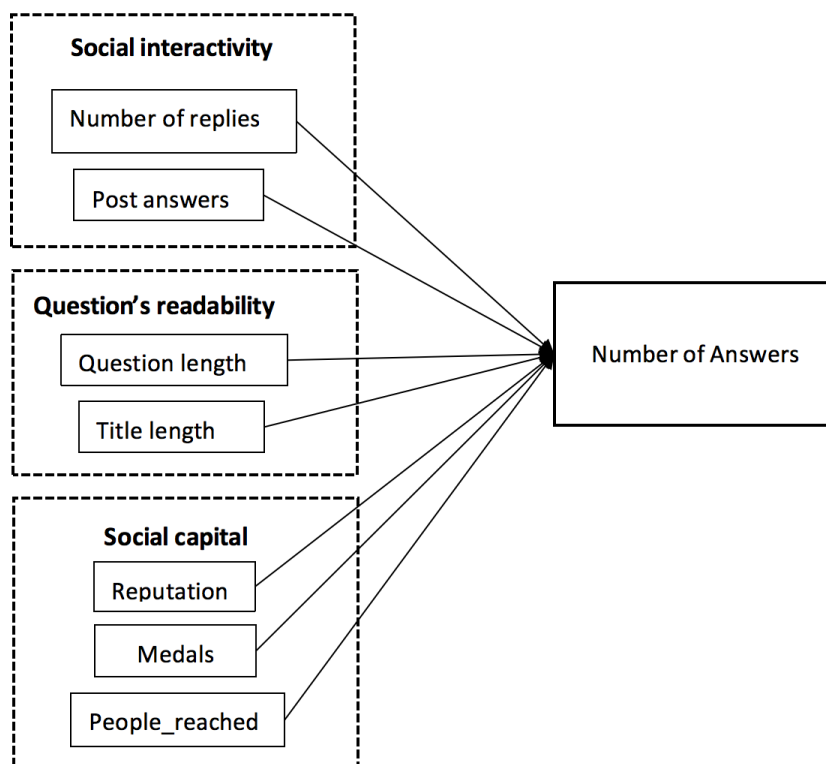


Figure 1. Research model

(1) Social interactivity

As a platform to facilitate information sharing between users, social interaction between users are important for Q&A communities^[12]. The most common type of users' interaction in Q&A communities is users' questioning and answering activities. These questioning and answering activities are also the foundation of the development of the Q&A communities. In this paper, we use the number of answers one users posted in Stack Overflow as a measurement of how active the user is in knowledge sharing on Stack Overflow. In addition, the social value and exchange benefits generated by interaction can make users experience positive emotional changes and promote their instant information sharing behavior^[13]. Studies also have shown that historical interaction can affect the social influence of questioners, and in the Q&A community, the social influence of questioners may affect the willingness of other users to answer^[8]. Further, according to the principle of reciprocity^[14], when users solve more problems for other users in the same community, their questions are more likely to receive answers by others. Therefore, we conjecture that the number of answers posted by users in the Q&A community is positively related to the number of answers received by their questions.

In addition to users' questioning and answering activities, there are other types of social interactions between users in Q&A communities. For example, the questioner may reply to some responders under his

question. Further, the responders may reply to the questioners' last round reply. That is, there may exist such "dialogue" type of interactions between users. Such interaction can bring more direct feedback to the respondent and can make the responder more active. Such "dialogue" type of interactions may generate high quality content and illicit other users' enthusiasm to reply. This important interactive behavior is often ignored by previous researches, which will be studied in our paper.

Therefore, we measure the level of social interactivity of users from the following two aspects: (1) the number of times the questioner responded to the respondent under his question (2) the number of answers the questioner has posted. The following hypotheses are made:

H1a: The number of responses from the questioner to the responder has a positive effect on the number of responses received.

H1b: The number of answers posted by the questioner in the community has a positive impact on the number of answers received by the questioner.

(2) Social capital

Bourdieu (1985) presented his own views on the definition of social capital. He proposed that social capital is related to the network of relationships it recognizes, and to the identity of members of a group^[15]. Social capital theory has been widely used to explain various social behaviors in Q&A communities, including information sharing and social achievements^[16].

In stack overflow, one user's identity in the community is reflected by his reputation and the number of medals earned. Based on their analysis on Stack Overflow, Chua et al. (2015) found that highly prestigious respondents generally perform better in providing detailed and high-quality answers, while novices perform better in terms of readability of answers. At the same time, they also found that novices answer questions such as facts, lists, and definitions more quickly. These results show the impact of user reputation on answering behaviors^[17]. Similarly, Zhang (2013) found in social media, compared with members of socially marginalized groups such as migrant workers, often people with higher social prestige have the right to speak^[18]. We thus test how users' reputation and number medals earned relates to the number of answers he received.

As for the relationship network in the community, one measurement of it in Stack Overflow is the visits of user's home page. When a user's homepage is visited more than others, it shows that he receives more attention in the community, which in turn indicates that his degree of recognition in the community network is high. Thus, we include the visits of user's home page as variable to measure the social capital of one user in Stack Overflow.

Therefore, to test how social capital of one user influence number of answers he receives, this article makes the following hypotheses:

H2a: The questioner's reputation has a positive effect on the number of answers he receives.

H2b: The number of medals a questioner has a positive effect on the number of answers they receive.

H2c: The number of visitors to the questioner's homepage has a positive impact on the number of answers to their questions.

(3) Questions' readability

Obviously, the characteristics of questions will influence the number of answers it receives. One important aspect of questions is the readability of questions. Questions' readability indicates how much efforts other users take to read the questions^[11]. In Q&A communities, questions' readability can be measured by the length of the questions and the length of the titles.

Apparently, it requires more time and efforts for users to provide answers to a question with low readability. According to the least-effort principle, people always try to minimize the average work that can be done. People's various social activities are governed by this principle^[19]. Since they are economically rational, they always try to make judgments and decisions as quickly as possible. Moreover, a huge number of questions are

posted every day on Stack Overflow and users bear a large amount of cognitive costs when answering questions. Thus a question with low readability might be passed at the first impression. According to the principle of least-effort principle, we therefore conjecture that users prefer questions with higher readability. That is, questions with shorter content and title. Accordingly, we make the following hypotheses:

H3a: The length of the question has a negative effect on the number of answers received for the question.

H3b: The length of the question title has a negative effect on the number of responses received for the question.

4. RESEARCH METHODOLOGY

We collected data from Stack Overflow, one of the largest Q&A communities worldwide. Everyday, more than 10 million users will visit Stack Overflow to ask, answer or view questions. It includes huge amount of information. More than 16.5 million questions were already received answers by October 2019 on Stack Overflow. Due to the huge amount of data, we limit our data collection time window as from January 2017 to January 2018. That is, only questions (and their answers received) posted during this time window are collected. We used a web crawler to collect data, including number of answers received and related features from three aspects including social interactivity, social capital and questions readability. After data cleaning, we obtained in total 13,897 observations.

To test Hypothesis H1-H3, we test the following regression model:

$$\text{Number of answers} = \beta_0 + \beta_1 \text{reply_num} + \beta_2 \text{post_answers} + \beta_3 \text{post_questions} + \beta_4 \text{reputation} + \beta_5 \text{medals} + \beta_6 \text{logp_r} + \beta_7 \text{questionlength} + \beta_8 \text{titlelength} + \epsilon \quad (1)$$

Previous studies employed the number of answers to represent the level of user knowledge contribution^[9]. We therefore employ the number of answers as the dependent variable. *Number of answers* measures the number of answers. *reply_num* measures how many times the questioner reply to respondents under a specific question. The larger this variable is, the higher degree of “dialogue” type of social interaction between users exist. *post_answers* indicates how many answers the questioner has provided to other users’ questions. Similarly, *post_questions* measures the number of questions the questioner has asked within the same Q&A community. These three variables represent social interactivities between users. The variable *reputation* is given by Stack Overflow to measure one user’s “reputation” score in the community. Stack Overflow has a mechanism to calculate each user’s “reputation” score based on his past performances. The basic idea is, users with higher “reputation” scores are more recognized within the community. Another mechanism design is the “medal” system. Users can earn medals by answering questions, posting answers or providing ratings for questions. When users completed certain amount of tasks, they will be granted a medal. Thus, the variable *medal* measures a user’s past contribution to the community. *logp_r* measures how many visits the questioners profile page has attracted. Higher visits indicate stronger network connects. These three variables can measure questioner’s social capital. Lastly, Variables *questionlength* and *titlelength* measures the length of questions and titles respectively. As discussed above, these two variables can be used to measure the readability of posted questions.

5. RESULTS AND DISCUSSION

5.1 Descriptive statistics

Table 1. Descriptive statistics

Variable	Mean	Std. Dev.	Min	Max
answers	1.013816	0.9221007	0	10
reply_num	0.5690127	0.9981565	0	10
post_answers	28.15457	202.8747	0	12687

Variable	Mean	Std. Dev.	Min	Max
post_questions	38.69144	106.3235	1	2407
reputation	1156.124	8648.17	1	513000
medals	4.827733	7.383647	0	263
logp_r	8.569213	2.785468	1.609438	18.61543
questionlength	1467.703	2024.676	35	28076
titlelength	55.57081	20.79852	15	161

According to the descriptive statistics of the data, it can be seen that the average number of answers received for each question is about one. the average number of questions posted by users is about 39, while the average number of posted answers is about 28. Such statistics shows that users usually tend to ask questions rather than answer questions. This further proves the importance of the research to explore factors that influence the number of answers received. Further, the average number of responses from users and responders was only 0.569. This shows that the social interaction between the questioner and the respondent exists but not very strong. Another statistic worth noticing is that the ranges of reputation is from 1 to 513000, indicating that users' reputation on Stack Overflow is distributed widely.

5.2 Regression analysis results

Table 2. Collinear Results

Variable	VIF	1/VIF
reputation	8.54	0.117103
post_answers	7.60	0.131549
medals	1.53	0.655487
log of people_reached	1.16	0.863011
questionlength	1.01	0.993776
titlelength	1.01	0.994906
reply_num	1.00	0.996282
Mean VIF	3.12	

Before performing regression analysis, we first performed a collinearity test on the research data. The results of the collinearity test showed that $VIF < 10$, which indicates that there is no multicollinearity in our experimental data.

Table 3. Testing The Hypotheses in The Research Model

Hypothesis	Relationship	Coef.	Std. Err.	P-values	Decision
H1a	reply_num->answers	0.2834891	0.0074462	0.000	Supported
H1b	post_answers->answers	0.0001875	0.0001009	0.063	Supported
H2a	reputation->answers	-6.91e-06	2.51e-06	0.006	Not supported
H2b	medals->answers	0.0024547	0.0015529	0.114	Supported
H2c	logp_r->answers	0.0198884	0.0031293	0.000	Supported
H3a	questionlength->answers	-0.0000188	3.68e-06	0.000	Supported
H3b	titlelength->answers	-0.0008798	0.0003576	0.014	Supported

As shown by Table 2, the final analysis results support the hypotheses H1a, H1b, H2b, H2c, H3a, H3b. Firstly, results show that the number of responses from the questioner, the number of answers posted, the number of medals users earned, and the number of visits to the questioner's homepage have a significant positive impact on number of answers. Therefore, as we have conjectured, results show that social interactivity and social capital have a significant positive impact on the answers to questions.

Secondly, results show the length of question and its title have a negative impact, indicating readability is a significant factor that influences users' willingness to answer as we have hypothesized.

However, H2a is not supported. That is, users' reputation score does not positively affect number of answers that their questions received. Rather, it significantly reduces the number of questions received. A close observation reveals that its magnitude is very small ($-6.91e-06$), although it is significant ($p=0.006$). This can be explained as follows. Users on Stack Overflow usually are senior users who has a long history with this platform. Stack Overflow focuses on technical questions and answers, such as programming techniques. These senior users usually have more in-depth knowledge with these techniques. As a result, their posted questions also will be more difficult for common users to provide answers. Therefore, their posted questions usually receive low number of answers. In contrast, novice users usually ask some questions that easy to answer and such that receive more answers.

6. CONCLUSIONS

Social Q&A communities are now important knowledge sharing platforms with large amount of users to ask and answer questions on various topics. However, users are more tend to ask questions rather than provide answers. Thus, to increase the number of received answers poses a challenge for these Q&A communities.

Drawing a dataset from Stack Overflow, one of the largest Q&A communities worldwide, this paper explores the factors that may influence users' activities to provide answers on Q&A communities. We specifically examined factors from three aspects including questioner's social capital, social interactivity between users and questions' readability. Our paper establishes their relationship with the number of received answers using a econometric model. For example, we find that the number of responses from the questioner, the number of answers posted, the number of medals users earned, and the number of visits to the questioner's homepage have a significant positive impact on number of answers; the length of question and its title have a negative impact. One interesting finding is that users' reputation score would reduce number of received answers, rather than increase. This might because users with high reputation usually are senior users with more technical knowledge and their posted questions are difficult for common users to answer.

This research contributes to the literature by establishing a framework to analyze influencing factors from three aspects. Some variables are first discussed in this paper. For example, we find that questioner's response to responder's answers will significantly increase the number of answers received. This proves the significance of social interaction. In addition, positive effect of the number of visits to the questioner's homepage provides the significant of social capital. This paper thus provides great practical and theoretical implications.

There are several limitations within this study. First of all, this research is conducted based on a cross-sectional dataset and some dynamics might not be captured. Future research may collect a panel dataset and conduct further study. Secondly, this paper only considers variables from three aspects. More variables might be related the number of answers one question receives, such as emotions of questions, usage of emoticon, when the question is posted. Thus, future research may take these variables into consideration. Lastly, the conclusion of this paper is solely drawn based on dataset from Stack Overflow. Future research may conduct analysis on other Q&A communities, such as CSDN, to test the robustness of our research.

REFERENCES

- [1] Jia Jia, Song Enmei, Su Huan. (2013).Research on Assessment of Answer Quality in Social Q&A Platform. *Journal of Information Resources Management*,3(02): 19-28(in Chinese).
- [2] Long T.Le, Chirag Shah.(2018).Retrieving People: Identifying Potential Answerers in Community Question-Answering. *Journal of the association for information science and technology*,69(10): 1246-1258.
- [3] Li Lei,He Daqing, Zhang Chengzhi. (2018).Survey on Social Question and Answer. *Data Analysis and Knowledge Discovery*, 2(07):1-12(in Chinese).
- [4] Huang Mengting, Zhang Pengyi.(2015). Collaboration Forms and Effectiveness in Social Q&A Communities: A Case Study of Zhihu.com. *Library and Information Service*, 59(12): 85-92(in Chinese).
- [5] Shah C,Oh S. (2009). Research agenda for social Q&A. *Library& Information Science Research*,31(4): 205-209.
- [6] Stack Exchange. (2019).2019 Stack exchange network overview. <https://stackexchange.com/about/>.
- [7] Chua AYK and Banerjee.(2015).Answers or no answers: Studying question answerability in stack overflow. *Journal of Information Science*, 41, 720-731.
- [8] Zhe Liu, Bernard J. Jansen. (2017).Identifying and predicting the desire to help in social question and answering. *Information Processing & Management*,53: 490-504.
- [9] Jiahua Jin,Yijun Li,Xiaoqia Zhong,Li Zhai. (2015).Why users contribute knowledge to online communities: An empirical study of an online social Q&A community. *Information & Management*,52(7): 840-849.
- [10] Zhang,J., Ackerman, M., & Adamic,L. (2007). Expertise networks in online communities: Structure and algorithms.*International Conference on World Wide Web ACM*,221-230.
- [11] Liu Z, Jansen B J. (2018).Questioner or question: Predicting the response rate in social question and answering on Sina Weibo. *Information Processing & Management*,54(2):159-174.
- [12] Lu H P, Lin J C, Hsiao K L, et al. (2010).Information sharing behavior on blogs in Taiwan: Effects of interactivities and gender differences. *Journal of Information Science*,36(3):401-416.
- [13] Wang C, Zhou Z, Jin X L, et al.(2017).The influence of affective cues on positive emotion in predicting instant information sharing on microblogs: Gender as a moderator. *Information Processing&Management*,53(3):721-734.
- [14] Komter Aafke. (1996).Reciprocity as a Principle of Exclusion: Gift giving in the Netherlands. *Sociology*, 30(2): 299-316.
- [15] Bourdieu, Pierre.(1985).The forms of capital, in Richardson Hand book of theory and research for the sociology of education. New York: Greenwood Inc.
- [16] Jabr W, Mookerjee R , Tan Y , et al.(2014).Leveraging Philanthropic Behavior for Customer Support: The Case of User Support Forums. *MIS Quarterly*,38(1):187-208.
- [17] Chua A Y K, Banerjee S.(2015). Measuring the effectiveness of answers in Yahoo! Answers. *Online Information Review*,39(1):104-118.
- [18] Zhang, Pengyi. (2013).Social Inclusion or Exclusion?: When Weibo (Microblogging) Meets the "New Generation" of Rural Migrant Workers. *Library Trends*,62(1):63-80.
- [19] Hardy, Andrew. (1982).The selection of channels when seeking information: Cost/benefit vs least-effort. *Information Processing & Management*,18(6): 289-293.