Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2018 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

6-26-2018

Identifying Impact Factors of Question Quality in Online Health Q&A Communities: an Empirical Analysis on MedHelp

Jiatong Shi
Renmin University of China, jiatong_shi@ruc.eud.cn

Wei Du
Renmin University of China, ahduwei@ruc.edu.cn

Wei Xu
Renmin University of China, weixu@ruc.edu.cn

Identifying Impact Factors of Question Quality in Online Health Q&A Communities: an Empirical Analysis on MedHelp

Completed Research Paper

Jiatong Shi
School of Information, Renmin University of China, Beijing, 100872, P.R. China
jiatong_shi@ruc.edu.cn

Wei Du
School of Information, Renmin University of China, Beijing, 100872, P.R. China
ahduwei@ruc.edu.cn

Wei Xu
School of Information, and Smart City Research Center, Renmin University of China, Beijing, 100872, P.R. China
weixu@ruc.edu.cn

Abstract

Online health Q&A communities help patients, doctors and other users conveniently search and share healthcare information online and have gained much popularity all over the world. Good-quality questions that raise massive discussions could trigger users' engagement online, which is beneficial for platform operation. However, little attention has been paid to the antecedents of question quality in online health Q&A communities. To have a deep investigation of healthcare question quality, this research aims to investigate the impact factors from two special aspects that are neglected in previous research, i.e., user's structural influence and questions' sentiment. Using a dataset collected from MedHelp, one of the largest online health Q&A communities, we found that users with high structural influences and questions with negative sentiment have positive associations with the answer number of questions. Our research would offer meaningful suggestions to platform managers and users.

Keywords: Online Health Q&A Community, Sentiment Analysis, Q&A network, Question Quality

Introduction

Online Q&A Communities is a popular form of online social communities which aims to help users to solve their problems through asking questions to the crowd. Millions of people are asking questions every day in successful online Q&A communities, such as Yahoo! Answers, Google Answers and StackOverflow. Some of them would like to brace all kinds of questions, while others focus on specific domains. Among them, online health Q&A communities are specialized on health problems. In the community, users can share their similar experience and construct their own social networks. The community acts as a harbor that users could have a rest and gain comfort or encouragement from others. Online health Q&A communities are similar with generic Q&A communities by providing questioning and answering functions for users to interact and seek for help. But questions on health Q&A communities tend to be more emotional because of the seriousness of health problems (Lu et al. 2017). Besides, users in healthcare communities are easier to establish a strong relationship given their similar treatments or illness experience (Yan et al. 2015). Online health Q&A communities have become popular tools for patients and doctors to help each other in recent years. For example, MedHelp, as one of the most popular websites of online health Q&A communities, has nearly 300 sub-communities for different diseases and more than 12 million visitors browsing information online each month (Yang et al. 2012).

However, low user engagement in online health Q&A communities needs to be concerned. We randomly chose 5000 users from Medhelp, and surprisingly found more than 50% of the users are not active and about 90% users never posted questions or answers. For online Q&A communities, valuable questions that could generate a great amount of answers would help encourage user engagement online (Attfield et al. 2011). As far as we know, a few studies investigated the reasons why users post questions or the measurement of question quality in Q&A communities, but impact factors of question quality in online health Q&A communities were barely investigated in previous research. By knowing why some health questions gain popularity online, this research would offer meaningful suggestions to both users and platform managers.

This paper proposes to identify impact factors of question quality from asker-based features and question-based features by leveraging social network analysis and sentiment mining. In our paper, we analyze the question from two aspects— question-based features and asker-based features. To profile users, we constructed a Q&A social network with Q&A history of each user and applied social network analysis to measure structural influences of users. In addition, we also measure user's accessed information through their concerning communities. For questions, we employed sentiment analysis method to measure sentimental polarities and subjectivities to measure their sentiment and professionalism. We perform a Zero-inflated Poisson Regression model (ZIP) for modeling. Regression results show that there is an interesting paradox for the effect of accessed information. Besides, we found that structural influences significantly and positively affect question quality. Negative sentiment has positive effect for question quality, while professionalism has no significant effect.

Our paper will be organized as follows. Firstly, we will review existing research on question quality for online Q&A communities and identify our contribution to the program. Next, we will list several possible impact factors and explain why they may influence question quality. After elaborating our research methodology, we discuss our result from the model. At last, conclusions and future work are provided.

Related Work

Online Q&A Community and its research field

Online Q&A Community is a special type of online community that spreads wisdom of the crowd and provides chances for users to be free to ask problem solving methods and answer questions (Harper et al. 2008). It developed from some traditional online Q&A services such as straight knowledge sharing sites, experts answering sites and sites acting as an index set. Mostly, those communities are free to users (Jin et al. 2009) and are without organizers (Liu et al. 2011). Besides online Q&A communities with general questions, many Q&A communities are specialized in certain domain, such as StackOverflow in coding and MedHelp in health care. A typical process of asking a question in an online Q&A community is simple. At first, a man with a problem posts a question in certain community. And then, the question will be shown in a board of the webpages on PC or Apps on mobile devices. People with answers will post their answer on the related page of the question. For some communities, the process of knowledge exchanging is over, while for others, the question may be closed for the asker to choose the best one. In addition, nearly all the communities have option for following viewers to vote for the best answer or questions (Lou et al. 2013).

Founded on the Online Q&A Communities, questions and answers are the centers of previous studies. Some studies stressed on why people ask questions in the online Q&A platform (Zhang. 2010). To elevate the efficiency of users, previous researches mainly focus on three parts. First, they considered how to perform Q&A retrieval efficiently (Jeon et al. 2005; Cao et al. 2010; Duan et al 2008). Second, they paid much attention to Q&A recommendation just like product recommendation system in e-commerce (Jeon et al. 2005; Qu et al. 2009). The last part is about evaluation the quality of questions and answers. There have been many prior studies concentrating on answer qualities. But unlike those focusing on answer quality (Neshati. 2017; Harper et al. 2008), question quality didn't gain such attention. For studies focusing on the connection between Q&A qualities, prior researches have shown that question quality may cast a great effect on the answers (Agichtein et al. 2008). A question with high quality could also generate more user flow by the means of attracting more users. Previous research devoted a lot on how to form a suitable model to define the question quality. For example, Li et al. combined the number of tag-of-interests, number of answer and a reciprocal of the minutes to reach the best answer together (Li et al. 2010). Based on their ground truth from domain experts, they gave a definition of question quality. Another instance, Ravi performed a topic model to extract textual features of certain questions and then based on votes for questions from StackOverflow, they raised a connection between topics and question quality (Ravi et al. 2014). Nevertheless, those models are generally to present question quality from perspective of users which stresses more on questions academic or daily usefulness. Neither of them considered question quality on user engagement.

Unlike those general Q&A sites which drew much attention of researchers, Online Health Q&A Communities are seldom concerned. For those did study health Q&A sites, they mainly studied the adoption decision of users (Jin et al. 2016), hot topics and user influences (Lu et al. 2013; Yang et al 2012). Nonetheless, comparing to other Q&A sites, there remain many differences in online health Q&A communities. First, sentiment based features have more effects on both askers and answerers (Lu et al. 2017). Second, patients are easier to become friends since they have similar ill experiences (Yan et al. 2015). Thus, their Q&A network and question sentiment are valuable to study.

Research Gap and Our Contribution

Based on the summary of previous studies, there remains a severe research gap between current studies' fields and the question qualities for gaining user engagement. Besides, as previous researches focused mainly on subjective features, the result is neither objective nor easy to understand. In this paper, we followed the insight of social networking and sentiment mining, and we found several interesting factors that may affect question quality. Through leveraging social network analysis and sentiment mining, we proposed a Zero-inflated Poisson (ZIP) Regression model to analyze how those impact factors may influence question quality (Lambert. 1992).

In conclude, our model innovatively focused on question quality for user engagement, which offers a new point of view for platform managers and users. Besides, we proposed several factors extracted by means of social network and sentiment analysis. Based on the factors, we developed a comprehensive model on the issue.

Hypothesis Development

This research examined the impact factors of to question quality based on question askers and the question itself. Based on the feature of the personal profiles and the question content, we performed a Q&A network for feature extraction and employed sentiment analysis for influencing factors. Figure 1 shows the framework of our research.

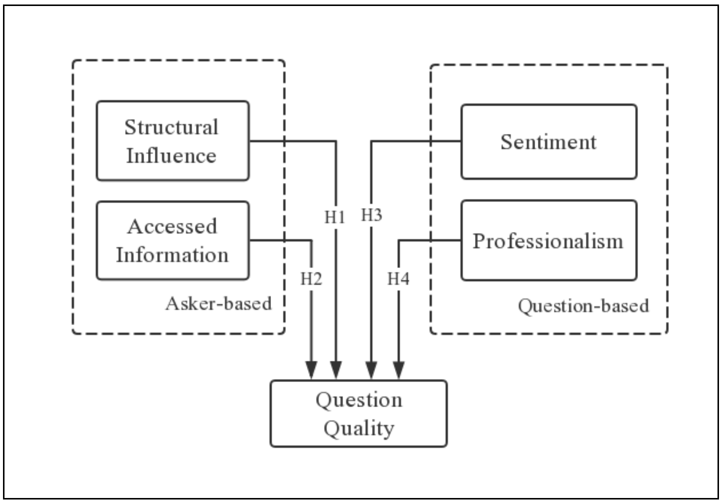

Figure 1. Research Framework

Learning is one of the main purposes for users in an online Q&A community which showed in a study on Baidu Knows (zhidao.baidu.com) (Lou et al. 2013). Based on a prior study on online social Q&A community, self-presentation, peer recognition and social learning are positive associated with knowledge contribution (Jin et al. 2015). So if a user has more access from more communities, they may be eager to self-present themselves and their information may lead to more valuable questions. Therefore, we assume that

Hypothesis 1. *A question asker's accessed information amount will be positively associated with the number of answers achieved by his/her questions.*

Social network analysis is an analytical tool for sociology at first, to study how people connect with each other and how to define the strength of a connection as well as the impact strength in the network (Scott 2017). After forming a graph through particular relationships of people in certain groups, we can analyze the architecture of the network and get several features from the individual as well as the whole network. There are many applications on social network analysis, such as class structure (Milroy et al. 1992), cross-continent trades (Kim et al. 2002), academic citation and even the spread of certain disease (Otte et al. 2002; Ortiz-Pelaez et al. 2006). One of the basic fields in social network analysis is community network. Since the rapid development of the Internet, some sites specializing on certain social network have gotten spectacular success including Facebook and Twitter. There are several researches over them on their information spreading pattern and its unique features (Ellison et al. 2007; Ellison 2007). There are also social networks in online Q&A communities (Yan et al. 2015). In some online Q&A communities, you can add friends, send notes, and post statuses just like Social Network Sites (SNS). So it is reasonable to apply methods in social network analysis to generate impact factors of question quality. MedHelp is this kind of online Q&A communities with features from SNS. There are several studies on the user influence through the network in MedHelp (Yang et al. 2012; Liu et al. 2015).

In our paper, we constructed a Q&A network for users. In our network, every user contained in a community was regard as a vertex in the network. If one user answers another user's question, then our network would establish a directed edge from the answerer to the asker. Through the network, we computed closeness centrality and eigenvector centrality for every user to measure their final structural influences. The closeness centrality measures closeness of a vertex to all other vertex (Okamoto et al. 2008). The eigenvector centrality measures influence of a node in a network. A vertex with high eigenvector centrality value shows that the vertex is connected to many other vertexes which have high influence (Bonacich 1972; Bonacich 2007). Closeness centrality and eigenvector centrality are basic characteristics to represent individual's structural influence. Closeness centrality is a vital factor in spreading of the resources (Hansen 2002), while eigenvector centrality is frequently used in social capital measure and influence in the network (Smith 2008; Borgatti et al. 1998; Fowler et al. 2008). In a study based on MathOverflow, a mathematic online Q&A community, closeness centrality was found casting significant positive impact on the future active participation in the community (Montoya et al. 2013). Besides, there was a previous study focusing on relationship between reputation of question askers and its questions by means of closeness centrality and eigenvector centrality (Low et al. 2015), and they proved their positive effects on individual reputation. Therefore, we assume that

Hypothesis 2. *A question asker's level of structural influence (e.g., Eigenvector centrality, or closeness centrality) in the Q&A network will be positively associated with the number of answers achieved by his/her questions.*

It is well accepted that textual information could be divided into two general parts: facts and opinions. Facts are objective, while opinions are subjective (Liu 2010). Sentiment analysis (SA) with another name Opinion Mining (OM) is an essential approach to understanding textual opinions in natural language processing (NLP) (Pang et al. 2008). The primary purpose of SA is to analyze people's emotions which are latent under text. Due to the virtually of online community, text becomes the most important part for communication between each other. Therefore, to elaborately study the inner logic of plain text information on the websites, we applied SA to online question content. Previous studies on SA and online community were often by the means of topic model such as LDA or polarity and subjectivity (Ravi et al. 2014; Li et al. 2010). A traditional method to decide polarity and subjectivity is to use emotional dictionary (Wilson et al. 2015). The method is simple but useful and widely using. The primary reason for its usefulness is that the method is objective and is easy to explain comparing to other fancy machine learning algorithms.

Polarity of question content is always considered as a factor for textual data. By collecting votes of each questions to measure question qualities, prior study on StackOverflow examined the role of affective lexicon on the questions posted (Novielli et al. 2014). Similar study on StackOverflow validated the application of sentiment analysis to Q&A sites (Novielli et al. 2015). For another Q&A sites, Stack Exchange, a research studied the success factor of questions in the community and found negative sentiment tend to negatively affect the questions (Calefato et al. 2015). Nonetheless, most previous studies on Q&A sites are not for a health care site. Unlike other Q&A communities, health care online communities Health questions are mostly associated with negative sentiment (Jiarpakdee et al. 2016). One possible reason for the phenomenon is the sympathy effect from questions with negative sentiment. Therefore, we assume that

Hypothesis 3. *A question's level of negative sentiment will be positively associated with the number of answers achieved.*

There are many studies on the effect of subjectivity in social Q&A communities (Liu et al. 2016; Liu et al 2015). They modeled intent detection as a classification problem. By applying the model to their dataset, they found that subjective and objective questions have significant difference impact on the community. And a question with subjectivity has more potential to attract responses from other users. For health care field, text with objectivity often means a more professional answer (Cornwall et al. 1995). Similarly, a question would be professional if it is objective. Thus, we assume that

Hypothesis 4. *A question's level of professionalism will be positively associated with the number of answers achieved.*

To sum up, although previous studies have provided some useful information to users as well as platform managers and serve as a reference for our paper, their studies didn't have specifications for online health Q&A communities. Besides, they didn't manage to combine all the factors together, so they might get a result with omitted variable bias. Therefore, our study is necessary and is a nice supplement for theirs.

Research Methodology

Dataset

The experimental dataset used in our paper are real-world data crawled from famous online health Q&A communities – MedHelp. Our dataset contains two important communities – Stress Community and Cold&flu Community from communities in MedHelp. There are 3850 questions in total from the beginning of the two communities to Nov. 2017. 2601 of them are from Cold&flu community, while 1249 are from Stress community. A question on MedHelp has question content which is textual data, question askers, and its answerers, as shown in Figure 2.

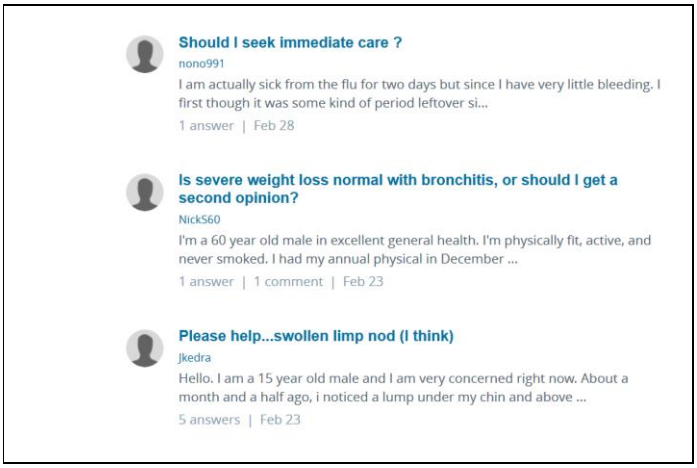

Figure 2. Screenshot of Questions on MedHelp

In total, we examine 6276 online users including answerers and askers. Besides, founded on the personal profile of each user, we collect all the users having posts in the two communities. The personal data including their history behavior of posts, status, notes received from others as shown in Figure 3.

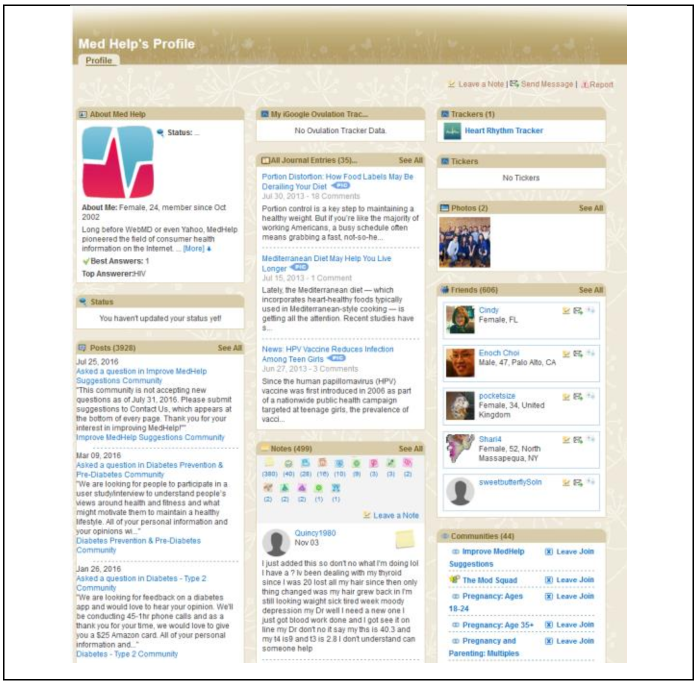

Figure 3. Screenshot of Personal Profile on MedHelp

For sentiment analysis, we apply English corpus of Textblob from Github which contains 2918 common words. With the emotion dictionary, we perform traditional sentiment analysis by Textblob and get polarity and subjectivity from textual data.

Variables

Dependent variable: Since answer number could encourage user engagement (Attfiled et al. 2011), we adopted answer number of certain questions as our dependent variable.

Independent variables: We extracted concerned communities for users directly from personal profile. Besides, we extracted eigenvectors centrality and closeness centrality from Q&A network constructed from the two communities. For each question's content, we used sentiment analysis techniques to compute polarity and subjectivity.

Control variables: we employed 6 control variables, including received note number of question askers, post (question and answer) number of question askers, status number of question askers, sentence number of a question, word number of a question body, word number of question title and question community (Stress and Cold & Flu in particular).

In MedHelp, notes are the other way to communicate apart from Q&A in communities. So notes number illustrates users' embedment in MedHelp. Well embedment may generate significant impact on user's connection with others, which influences question quality. Thus, we controlled notes number in our model.

Posts number indicates the number of Q&A for a user. Same as notes number, post number measure the activeness for a user. We controlled posts number in our model.

Status is a function in MedHelp, which is like sharing in Twitter where users can upload their emotions in a few words. Aside from embedment for communities, status number reveals users' personalities on self-presenting which may affect their question delivery. Therefore, we controlled status number in our model.

The length of a question is a basic feature for question. Lengthy question may lay difficulties on reading, while short question cannot illustrate a question well. Since sentence structure is another factor that may affect the question, we took sentence number, word number into consideration. Due to the functional difference between title and content, we separately controlled word number of content and word number of title.

As we consider two communities, we generated a dummy variable to control the effect of the community difference.

Table 1 provides a summary of all the variables we used in our model.

Table 1. Variable Summary

Features	Explanation
Answer number	The answer number of a question
Closeness centrality	Closeness to other users in the Q&A network
Eigen centrality	Influence of a user in the Q&A network
Subjectivity	Subjectivity of a question
Polarity	Sentiment polarity of a question
Communities	Concerned community number of a user
Notes	Number of notes a user received from others
Status	Number of statuses posted by a user
Post	Number of Q&A posted by a user
Sentences	Sentences number of question content
$\textbf{Wordcount}_{\textbf{content}}$	Word number of question content
$\textbf{Wordcount}_{\textbf{title}}$	Word number of question title

Model setup

We employed Zero-inflated Poisson (ZIP) Regression model to analyze effects of our independent variables. The first reason is that answer number is count data. Since traditional Ordinary Least Square (OLS) model can only deal with real variable and it is unwise to rescaled answer number to a set of categories for multi-classes Logit model which ignores comparable characteristics in answer number, we applied Poisson Regression model as our basic model. When exploring the data, we found that there remain 32.3% structural zeros in answer number. So to deal with the effect of excess zeros, we applied ZIP to fit our situation.

In a typical ZIP regression (Lambert. 1992), the responses Y_i are independent and they show as 0 with a probability p_i and show as Possion(λ_i) with probability of $1 - p_i$. In general, ZIP will estimate two sub-models. One is for prediction of structural zeros, while the other is for the count data. In our situation, the structural zeros are questions without answers. To be simplified in the rest of the paper, we use Target1 for whether the question has an answer and Target2 for the answer number with questions satisfied Target1. Through Maximum Likelihood Estimation (MLS), we could estimate p_i and λ_i. The p_i and λ_i estimation is modelled as

$$\begin{aligned} Log(\hat{\lambda}) = \beta_1 &+ polarity + subjectivity + communities + eigencentrality \\ &+ closenesscentrality + authority + notes + posts + status + sentences \\ &+ wordcount_{content} + wordcount_{title} + community_{dummy} + \varepsilon_1 \end{aligned}$$

$$\begin{aligned} \text{Logit}(\hat{p}) = \beta_2 &+ polarity + subjectivity + communities + eigencentrality \\ &+ closenesscentrality + authority + notes + posts + status + sentences \\ &+ wordcount_{content} + wordcount_{title} + community\_dummy + \varepsilon_2 \end{aligned}$$

Over-dispersion is an unexpected difference between expected features of the model and feature from real dataset. In the Poisson regression part of our model, we test if the model is over dispersed based on our dataset (Breslow. 1990). We got a p value similar to 0 in the test of over dispersed. So we apply Over-dispersed Poisson model for Possion(λ_i) (Gardner et al. 1995).

Results and Discussions

Descriptive statistics

Table 2 summarizes the descriptive statistics of our dataset. Several characteristics of the features are indicated from Table 2. Firstly, since the mean of subjectivity and the mean of polarity are 0.34 and -0.03, the main sentiment of question content is negative and objective. These may because people tend to feel worried when they have health problems and since the description of the illness is the main part in the content, it might contribute a lot to objective result. Secondly, most of our features are highly right-skewed, which is common in online community and can perfectly fit the assumptions of Poisson Regression.

Table 2. Descriptive statistics of the data

Features	Min	Median	Max	Mean	St. dev
Answer number	0	1	99	1.79	3.14
Closeness centrality	0.00	0.15	1.00	0.26	0.33
Eigen centrality	0.00	0.002	1.00	0.02	0.04
Subjectivity	0.00	0.33	1.00	0.34	0.18
Polarity	-0.74	-0.01	1.00	-0.03	0.15
Communities	0	1	65	2.83	5.48
Notes	0	0	1310	6.61	65.22
Status	0	0	240	1.05	10.79
Post	0	2	11978	74.88	508.96
Sentences	1	4	88	6.09	6.80

$Wordcount_{content}$	13	97	1519	133.03	123.78
$Wordcount_{title}$	5	9	49	10.00	3.35

Regression result

The parameter estimates of our model are presented in Table 3. Model 1 is Logit Regression that determines the p_i for Target1. Model 2 estimates λ_i based on assumptions on Poisson distribution of Target2. The Cox-Shell Pseudo R-square of Logit Regression and Poisson Regression of our ZIP is 0.380 and 0.274, which shows that user features and question features can help to explain the answer number of a question.

Communities (i.e., concerned communities number) have a significant negative coefficient in Logit part and a significant positive coefficient in Poisson model, which partially supports Hypothesis 1. It shows that as users have more access to different communities' knowledge, it would be difficult for answerers to answer the question. However, once the question is answerable, the question would be more likely to achieve more answers. The paradox is possibly caused by the accumulated knowledge effect. For users who achieved amounts of healthcare information from more communities, they wish to achieve more valuable or professional information from others, and this explains why their questions are not easy to be answered. But once their questions get answered, the valuable information would attract more users' attentions and discussions. Generally, the effect of concerned communities' number involves a contradiction between having an answer and having more answers.

Network features including closeness centrality and eigenvector centrality are both significant and positive in Logit and Poisson model. Hence, Hypothesis 2 is supported. A high question asker's level of structural influence in the social network leads to a larger amount of answers. Their structural influence can help them gain popularities and attract more answers from other users. The result indicates that platform managers should focus more on willingness of those opinion leaders and provide some supports to them in order to develop Q&A communities.

Since polarity is significant and negative in Poisson part of our model, Hypothesis 3 is supported. Our dataset shows that in Stress community and Cold&Flu community of MedHelp, users' negative emotion in questions would generate sympathy of other users and as a result gain many answers. For platform managers, how to carve the community to a place where people could open up their deep heart to others would be very essential.

Since subjectivity is not significant and remains controversial in the both parts of the ZIP. So our hypothesis 4 cannot be supported for our dataset. The reason probably lies in the characteristic of online communities (Preece 2000). Online community tends to emphasize the sociality which is irrelevant to professionalism. So the professionalism may not be a central factor affecting answer number of a question.

Table 3. Regression Result

	Logit Regression	**Poisson Regression**
Closeness centrality	6.07 (17.118) ***	-0.29 (-4.331) ***
Eigen centrality	88.27 (11.878) ***	5.31 (15.697) ***
Subjectivity	0.17 (0.664)	-0.16 (-1.365)
Polarity	-0.15 (-0.466)	-0.27 (-1.839) *
Communities	-0.026 (-2.524) **	0.0011 (2.725) ***
Notes	-0.0038 (-3.601) ***	0.0011 (3.675) ***
Status	-0.025 (-3.082) ***	-0.0022 (-1.496)
Post	0.000012 (-0.058)	-0.00031 (-7.151) ***

Sentences	0.0095 (0.857)	0.0025 (0.577)
$\text{Wordcount}_{content}$	-0.00057 (-0.947)	0.000078 (0.311)
Wordcount_{title}	-0.0082 (-0.-0.571)	0.012 (2.030) **
Community dummy	-0.75 (-6.574) ***	-0.063 (-1.359)
_cons	-0.020 (-1.7133)	0.83 (10.892) ***
Pseudo R2	0.380	0.274

** $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$ (t statistics in parentheses)*

Overall, the results show that some of our proposed factor influences the answer number of a question. For elevation of the question quality, Q&A network should be focused and actions based on sentiment features should be taken.

Conclusions

In this paper, we explored several factors for question quality. Based on prior studies and specific context of online health Q&A communities, we investigated the factors from basic features, Q&A network features and sentiment features. We conducted an empirical study on the data crawled from MedHelp, a famous online health Q&A communities. The results show that the structural influences, question sentiment play important roles in question quality measured by answer number. To be specific, users with high structural influences, negative sentiment have a large amount of answers. Platform managers should provide supports for users with high network properties and create a comfortable environment for pouring negative emotions. Users, who want to get popularity in the community, need to accumulate their structural influences by asking more valuable questions.

Our paper provides several theoretical and practical contributions. For starters, we firstly study the question quality in an online health Q&A community on the perspective of platform managers. Even for online Q&A community, the literature on question quality is limited. Next, our paper is one of the first paper that combined network properties and sentiment features together for analysis. It brings insights for further study on Q&A communities. At last, our findings can provide practical advice for platform managers and users. These suggestions can help them handle the challenges of losing user engagement.

This research has several limitations. Our study is only based on MedHelp. However, there are still many other online health communities we didn’t consider. Since the features for each site may differ with each other, it is necessary to study more sites such as Yahoo! Health. Second, according to question content, question’s topics also affect question quality. Further study may examine the influence of the topics on question quality. Third, there are some users that only browse questions without answer or ask any question. So simply using answer number as a measure for question quality is not enough. Further study may consider browse number instead of answer number.

In the future, we would explore more features from Q&A network and develop more comprehensive suggestions for platform managers on more specific actions.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China (Grant No. 71301163, 71771212), Humanities and Social Sciences Foundation of the Ministry of Education (No. 14YJA630075, 15YJA630068), the People’s Livelihood Investigation Project of Social Sciences Development Fund (201701602), Fundamental Research Funds for the Central Universities, and Research Funds of Renmin University of China (No. 15XNLQ08), Research Funds of Renmin University of China (No. 2018030009).

References

Agichtein, E., Castillo, C., Donato, D., Gionis, A., and Mishne, G. 2008. "Finding high-quality content in social media." In Proceedings of the 2008 International Conference on Web Search and Data Mining, ACM, pp. 183-194.

Attfield, S., Kazai, G., Lalmas, M., and Piwowarski, B. 2011. "Towards a science of user engagement." In WSDM Workshop on User Modelling for Web Applications, pp. 9-12.

Bonacich P. 1972. "Technique for analyzing overlapping memberships." Sociological Methodology, (4) pp. 176-185.

Bonacich P. 2007. "Some unique properties of eigenvector centrality." Social Networks (29:4) pp. 555-564.

Borgatti S P, Jones C, and Everett M G. 1998. "Network measures of social capital." Connections (21:2), pp. 27-36.

Breslow N. 1990. "Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models." Journal of the American Statistical Association (85:410), pp. 565-571.

Calefato, F., Lanubile, F., Merolla, M. R., and Novielli, N. 2015. "Success factors for effective knowledge sharing in community-based question-answering."" In Proc. 10th International Forum on Knowledge Asset Dynamics.

Cao, X., Cong, G., Cui, B., and Jensen, C. S. 2010. "A generalized framework of exploring category information for question retrieval in community question answer archives." In Proceedings of the 19th International Conference on World Wide Web, ACM, pp. 201-210.

Cornwall, A., and Rachel J. 1995. "What is participatory research?" Social science & medicine (41:12), pp. 1667-1676.

Duan, H., Cao, Y., Lin, C. Y., and Yu, Y. 2008. "Searching questions by identifying question topic and question focus." Proceedings of ACL-08: HLT, pp. 156-164.

Ellison, N. B. 2007, "Social network sites: Definition, history, and scholarship." Journal of Computer-mediated Communication (13:1), pp. 210-230.

Ellison, N. B., Steinfield, C., and Lampe, C. 2007. "The benefits of Facebook “friends:” Social capital and college students’ use of online social network sites." Journal of Computer-mediated Communication (12:4), pp. 1143-1168.

Fowler, J. H., and Christakis N A. 2008. "Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study." Bmj (337), a2338.

Gardner, W., Mulvey, E. P., and Shaw, E. C. 1995. "Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models." Psychological bulletin (118:3), pp. 392.

Hansen, M. T. 2002. "Knowledge networks: Explaining effective knowledge sharing in multiunit companies." Organization Science (13:3) pp. 232-248.

Harper, F. M., Raban, D., Rafaeli, S., and Konstan, J. A. 2008. "Predictors of answer quality in online Q&A sites." In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, pp. 865-874.

Harper, F. M., Raban, D., Rafaeli, S., and Konstan, J. A. 2008. "Predictors of answer quality in online Q&A sites."" In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, pp. 865-874.

Jeon, J., Croft, W. B., and Lee, J. H. 2005. "Finding similar questions in large question and answer archives."" In Proceedings of the 14th ACM international conference on Information and knowledge management, ACM, pp. 84-90.

Jiarpakdee, J., Ihara, A., and Matsumoto, K. 2016. "Understanding question quality through affective aspect in Q&A site." Proceedings of the 1st International Workshop on Emotion Awareness in Software Engineering, ACM, pp. 12-17.

Jin, J., Li, Y., Zhong, X., and Zhai, L. 2015. "Why users contribute knowledge to online communities: An empirical study of an online social Q&A community." Information & management (52:7), pp. 840-849.

Jin, J., Yan, X., Li, Y., and Li, Y. 2016. "How users adopt healthcare information: an empirical study of an online Q&A community." International journal of medical informatics (86), pp. 91-103.

Jin, X. L., Lee, M. K., and Cheung, C. M. 2009. "Understanding users' continuance intention to answer questions in online question answering communities." In Proceedings of the International Conference on Electronic Business, pp. 679-688.

Kim, S., and Shin, E. H. 2002. "A longitudinal analysis of globalization and regionalization in international trade: A social network approach." Social Forces (81:2) pp. 445-468.

Lambert, D. 1992. "Zero-inflated Poisson regression, with an application to defects in manufacturing." Technometrics (34:1), pp. 1-14.

Li, N. and Wu, D. D. 2010. "Using text mining and sentiment analysis for online forums hotspot detection and forecast." Decision support systems, (48:2) pp. 354-368.

Liu B. 2010. "Sentiment Analysis and Subjectivity." Handbook of natural language processing (2) pp. 627-666.

Liu, Z. and Jansen, B. J. 2016. "Understanding and Predicting Question Subjectivity in Social Question and Answering." IEEE Transactions on Computational Social Systems (3:1), pp. 32-41.

Liu, Q., Agichtein, E., Dror, G., Gabrilovich, E., Maarek, Y., Pelleg, D., and Szpektor, I. 2011. "Predicting web searcher satisfaction with existing community-based answers." In Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, pp. 415-424.

Liu, Z., and Jansen, B. J. 2015. "Subjective versus objective questions: Perception of question subjectivity in social Q&A." In International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, Springer, pp. 131-140

Lou, J., Fang, Y., Lim, K. H., and Peng, J. Z. 2013. "Contributing high quantity and quality knowledge to online Q&A communities." Journal of the Association for Information Science and Technology (64:2), pp. 356-371.

Low, J. F., and Davor S. 2015. "Data analysis of social community reputation: Good questions vs. good answers." 2015 IEEE International Conference, IEEE, pp. 1193-1197

Lu, Y., Wu, Y., Liu, J., Li, J., and Zhang, P. 2017. "Understanding Health Care Social Media Use From Different Stakeholder Perspectives: A Content Analysis of an Online Health Community." Journal of Medical Internet Research (19:4), e109.

Lu, Y., Zhang, P., Liu, J., Li, J., and Deng, S. 203. "Health-Related Hot Topic Detection in Online Communities Using Text Clustering." PLoS ONE (8:2), e56221.

Milroy, L, and Milroy J. 1992. "Social network and social class: Toward an integrated sociolinguistic model." Language in Society (21:1), pp. 1-26.

Montoya, L. V., Ma, A., and Mondragón, R. J. 2013. "Social achievement and centrality in MathOverflow." In Complex Networks IV, Springer, pp. 27-38.

Neshati, M. 2017. "On early detection of high voted Q&A on Stack Overflow." Information Processing & Management (53:4), pp. 780-798.

Novielli, N., Calefato, F., and Lanubile, F. 2014. "Towards discovering the role of emotions in stack overflow." Proceedings of the 6th international workshop on social software engineering, ACM, pp. 33-36.

Novielli, N., Calefato, F., and Lanubile, F. 2015. "The challenges of sentiment detection in the social programmer ecosystem." Proceedings of the 7th International Workshop on Social Software Engineering, ACM, pp. 33-40.

Okamoto, K., Chen, W., and Li, X. Y. 2008. "Ranking of closeness centrality for large-scale social networks." International Workshop on Frontiers in Algorithmics, Springer, pp. 186-195.

Ortiz-Pelaez, A., Pfeiffer, D. U., Soares-Magalhaes, R. J., and Guitian, F. J. 2006. "Use of social network analysis to characterize the pattern of animal movements in the initial phases of the 2001 foot and mouth disease (FMD) epidemic in the UK." Preventive Veterinary Medicine (76:1-2), pp. 40-55.

Otte, E., and Rousseau, R. 2002. "Social network analysis: a powerful strategy, also for the information sciences." Journal of Information Science (28:6) pp. 441-453.

Pang, B., and Lee, L. 2008. "Opinion mining and sentiment analysis." Foundations and Trends in Information Retrieval (2:1–2) pp. 1-135.

Preece, J. 2000. "Online communities: Designing usability and supporting socialbilty." John Wiley & Sons, Inc.

Qu, M., Qiu, G., He, X., Zhang, C., Wu, H., Bu, J., and Chen, C. 2009. "Probabilistic question recommendation for question answering communities." In Proceedings of the 18th international conference on World wide web, ACM, pp. 1229-1230.

Ravi, S., Pang, B., Rastogi, V., and Kumar, R. 2014. "Great Question! Question Quality in Community Q&A." ICWSM (14), pp. 426-435.

Scott, J. 2017. "Social network analysis." Sage.

Smith, M. S. 2008. "Social capital in online communities."" In Proceedings of the 2nd PhD workshop on Information and Knowledge Management, ACM, pp. 17-24.

Wilson, T., Wiebe, J., and Hoffmann, P. 2015. "Recognizing contextual polarity in phrase-level sentiment analysis" Proceedings of the conference on human language technology and empirical methods in natural language processing, Association for Computational Linguistics, pp. 347-354.

Yan, L., Peng, J., and Tan, Y. 2015. "Network dynamics: how can we find patients like us?" Information Systems Research (26:3), pp. 496-512.

Yang, C. C., and Tang, X. 2012. "Estimating user influence in the MedHelp social network." IEEE Intelligent Systems (27:5), pp. 44-50.

Zhang, Y. 2010. "Contextualizing consumer health information searching: an analysis of questions in a social Q&A community."" In Proceedings of the 1st ACM International Health Informatics Symposium, ACM, pp. 210-219.