Examining User Roles in Social Q&A: the Case of Health Topics in Zhihu.com

Zhe Wang
Department of Information Management
Peking University
5 Yiheyuan Rd, Haidian District
Beijing 100871, China
zhewang@pku.edu.cn

Pengyi Zhang
Department of Information Management
Peking University
5 Yiheyuan Rd, Haidian District
Beijing 100871, China
pengyi@pku.edu.cn

ABSTRACT
Social question answering (SQA) sites have become a popular online community platform for information seeking and social interaction. Prior SQA research has focused more on questions and answers, but has not examined user roles and their characteristics and activity patterns. In this poster, we identified four collaborative roles (starter, answerer, technical editor and follower) in Zhihu.com, a popular expert SQA site in China, and examined their characteristics and activity patterns. Our data set contained 14,779 users, 2,362 questions, and 8,486 answers from 12 health-related topics. Results show that, starters and technical editors ask more questions, make more technical edits, answers a few questions, and are not well connected to other users. Answerers contribute more content, attract the most followers and received the most likes and votes. Followers contribute little content, and receive the least likes and votes, but follow more topics and users. Content contributing activities such as answering questions gain more likes and votes, but posting a question and making a technical edit can get more pageviews. Users have overlapping roles and more research is needed to fully understand the interaction and interplay among user roles in SQA collaboration.

Keywords
Social question answering sites, SQA, user roles, social collaboration, Zhihu.com

INTRODUCTION
Social question answering (SQA) sites are web-based communities for information seeking by asking natural language questions to other users in a network (Shah, 2008). Social question answering may occur by a user posting a question in designated SQA services or systems such as Yahoo! Answers, Quora, or Zhihu (Harper, Moy, & Konstan, 2009; Kim, 2010), or through status messages posted as questions on other social networking sites such as Facebook, Twitter, and Sina Weibo (Morris, Teevan, & Panovich, 2010; Zhang, 2012). Prior SQA research has found that collaboration and interaction among users are able to elicit more answers, improve user satisfaction, and provide a sense of community (Gazan, 2010).

Research in user collaboration in SQA has focused on mostly micro-collaboration in answers (Gazan, 2007). Gazan (2007) has identified roles of seekers and sloths in SQA. However, social collaboration also occurs at the macro level in SQA sites, in which users play various roles that are more complex and interconnected. A large body of research has identified various roles in online communities (Preece, 2000; Gleave, 2009; Haythornthwaite, 2005; Turner, 2005; Geiger & Ribes, 2010), ranging from administrative roles, content contributors, and marginal roles. Role assignment has been found to have a positive effect on knowledge construction in online groups (De Wever, 2010). However, most of the prior work of user roles applies to general online communities, while little research has examined user roles and characteristics of such roles in SQA sites.

In this exploratory study, we examined user roles on Zhihu.com, a Chinese equivalent of Quora. Zhihu encourages knowledge exchange and social collaboration among its users and has attracted more than 17 million users by March 2015. As a start, we chose health topics since SQA is growing to become a major source of health information. 72% Internet users in the U.S. look for health information online (Fox, 2013), and seek for health information by posting questions on SQA sites such as Yahoo! Answers (Bowler et al., 2012).

This poster aims to address the following research question: what are the characteristics and activity patterns of different user roles in SQA?

ASIST 2016, October 14-18, 2016, Copenhagen, Denmark.

RELATED WORK

Collaboration in SQA

SQA sites contain both informational questions to solicit specific facts and conversational questions to carry on discussions (Harper et al., 2009). Research on Yahoo! Answers has found 64% of the questions are informational whereas the other 36% are conversational (Kim, 2010). Conversational questions are used to stimulate discussions, and to get opinions.

A large body of research is concerned with assessing the quality of answers generated in SQA sites. Higher rated answers demonstrate features including: quality, accuracy, clarity, currency of the information provided (Kim, 2010), using citations and supporting evidence such as links to other sites (Gazan, 2007; Harper et al., 2009). In addition to information provided in an answer, social aspects such as the author's reputation or rank also affect the perceived rating of an answer (Gazan, 2010; Shah & Pomerantz, 2010).

Previous research has reported social collaborations in SQA and collaborative information seeking (Hansen & Rvelin, 2005; Gazan, 2010; Wang, Gill & Mohanlal, 2013). For example, two thirds of social collaboration in collaborative information retrieval is document-related, while one third is human-related (Hansen et al., 2005). Research has found that in SQA sites, collaboration takes place in brief, informal episodes, and users with higher ranking are found to contribute more content (Gazan, 2010).

Factors influencing users' collaborative behaviors include willingness to share information, altruism and morality, perceived pleasure, social capital and resources, and affective factors (Hertzum, 2008; Gazan, 2010; Zhang, 2012).

User Roles in Online Community

Related studies have recognized three types of user roles: administrators, content contributors, and marginal roles.

Administrators

(a) *Mediators / moderators* ensure that users in the community behave reasonably by facilitating ongoing discussions, managing the lists, filtering spam, preventing flames, and so on (Preece, 2000; Gazan, 2010).

(b) *Vandal fighters / flame warriors* revert vandalism and sanction norm violators (Geiger & Ribes, 2010).

Content Contributors

(c) *Questioners* initiate a question by asking it for the first time (Gazan, 2010).

(d) *Answer people* respond to questions posted by other users and collectively donate vast amounts of valuable contributions to those who ask questions, resulting in the creation of valuable online resources (Gleave, 2009; Haythornthwaite, 2005; Turner, 2005).

(e) *Discussion people* are characterized by frequent reciprocal exchanges on opinions with a relatively high number of other participants (Gleave, 2009; Haythornthwaite, 2005; Turner, 2005).

(f) *Technical editors* correct small errors related to style or formatting of articles, e.g. spelling, grammar, faulty links, improper topic classification, and so on (Gleave, 2009; Geiger & Ribes, 2010).

Marginal Roles

(g) *Fans* express appreciation or affection on content or users by "liking" or "voting" (Haythornthwaite, 2005; Turner, 2005).

(h) *Lurkers* do not actively contribute contents or connect with others. They participate by following topics or other people. Some researchers see a good chance of lurkers turning into more active participants (Preece, 2000; Gleave, 2009).

RESERCH METHODS

Data Collection

We selected 12 topics in the health domain from Zhihu.com and used a Web crawler to collect questions, answers, question and topic editing logs. We then collected user profile information and activity data of those who participated in these topics by either asking a question, providing an answer, editing a question or topic description, or following a topic/question. We adopted four user roles that were present in our data including starter, answerer, technical editor, and follower. Table 1 shows the descriptions of the roles.

Roles	Description
Starter	Asking questions and possibly asking follow-up questions or commenting on answers when discussions get stuck.
Answerer	Providing answers to questions.
Technical editor	Editing topics and questions, such as adding or revising topic-question relationship, topic label, descriptions and images.
Follower	Following topics or questions.

Table 1. Descriptions of User Roles in Zhihu.com

Data was collected during the month of November 2015. Table 2 (next page) summarizes the data set. Examples of the twelve topics include: rare disease (T2), skin allergy (T5), rehabilitation medicine (T8), and so on.

Data analysis

We first analyzed the profile information and activity patterns of different user roles and conducted statistical tests to reveal any differences among the roles. User profile information included username, id, gender, residency, education, and job industry.

	# of Qs	# of As	# of Starters	# of Answerers	# of Technical editors	# of Followers	Total Users
T1	141	703	123	453	157	124	857
T2	84	1055	94	692	136	873	1795
T3	12	22	21	16	25	76	138
T4	797	2792	468	1213	538	903	3122
T5	441	1072	606	734	634	825	2799
T6	95	208	125	165	134	743	1167
T7	2	31	0	13	3	7	23
T8	197	586	217	369	239	175	1000
T9	292	1236	347	708	376	1009	2440
T10	286	749	366	511	376	234	1487
T11	13	23	6	15	5	142	168
T12	2	9	2	9	4	130	145
Total	2362	8486	2375	4898	2627	5241	15141

Table 2. Data Summary

User activity data include:

- *Questions*: number of questions a user has posted;
- *Answers*: number of answers a user has posted;
- *Technical edits*: number of technical edits a user has done;
- *Topic followings*: number of topics a user is following;
- *User followings*: number of users a user is following;
- *Followers*: number of followers;
- *Pageviews*: number of times a user's homepage is viewed;
- *Votes*: number of votes a user's answers have received;
- *Likes*: number of likes a user's answers have received.

To further investigate the similarity and difference among starters, answerers, technical editors and followers, we examined user characteristics by roles to see if there was any pattern. We divided the activities to three categories:

- *Content contribution*, including posting questions, answers, and topic or question editing;
- *Social connection,* including following topics, following other users and being followed by other users;
- *Popularity and recognition,* including pageviews, votes and likes received.

FINDINGS

User Distribution and Characteristics by Roles

Some users participate in multiple topics, and might serve same or different roles across topics. We removed 662 duplicate user IDs with same roles and 14,779 users remain. Figure 1 shows the distribution of starter, answerer, technical editor and follower in 12 health-related topics.

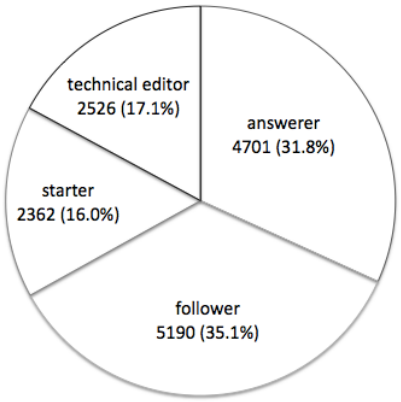

Figure 1. User Distribution of 12 Health-related Topics

2362 (16.0%) of the 14,779 users serve as starters, 4,701 (31.8%) are answerers, 2,526 (17.1%) are technical editors, and 5,190 (35.8%) of the 14,779 users are followers.

There are overlaps between different user roles, meaning that some users serve multiple roles. For instance, an answerer of one question may be the starter of another question. 150 (2.1%) answers also act as starters while 120 (1.7%) answerers also act as technical editors.

Gender

12,626 (87.2%) users have provided gender information in their profiles. Among them, 51.9% are male, and 48.1% are female, which suggests a nearly balanced distribution of gender in the users. However, gender distribution differs significantly among user roles (shown in Table 3).

	Male (N, %)		Female (N, %)	
Starter	1026	51.5%	967	48.5%
Answerer	2400	56.1%	1880	43.9%
Technical editor	1177	53.8%	1009	46.2%
Follower	2137	47.3%	2384	52.7%
Total	6740	51.9%	6240	48.1%

Table 3. Gender Distribution in User Roles

K-W Test suggests there is a significant difference in gender distribution with starters, answerers, technical editors, and followers (χ^2=72.17, p<0.001). Male users constitute a notable larger portion of answerers and technical editors while more female users serve the roles of starters and followers.

Job Industry

5,897 (39.9%) of the users filled their job information in their profiles. By applying the National Industries

Classification of China[1], we classified users' job industries into 19 categories. Figure 2 shows industries with more than 200 users.

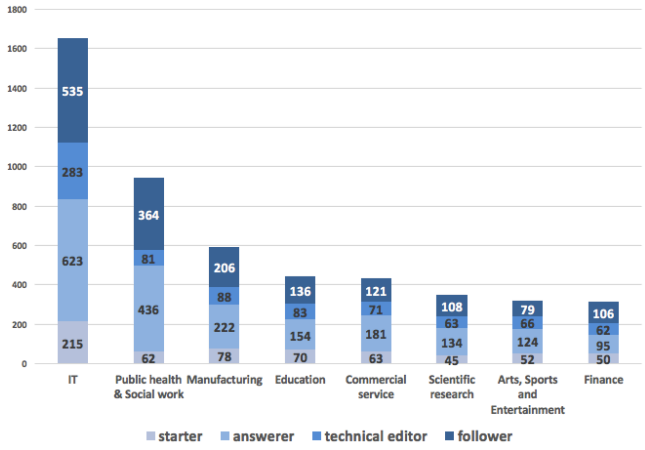

Figure 2. User Distributions in Job Industry

More than half of the users in the 12 health topics are either working in IT (27.9%), public health and social work (15.9%), or manufacturing (10.0%). There is no significant difference across roles in job industries.

SQA Activity Patterns by Roles

Content Contribution

We compared the number of posted questions, answers, and technical edits of a user by roles. Table 4 shows the results:

Mean / Median	Starters	Answerers	Technical editors	Followers
# of questions	13.29 / 2	4.99 / 1	17.36 / 3	1.47 / 0
# of answers	19.2 / 1	188.26 / 19	30.88 / 2	10.03 / 0
# of technical edits	123.86 / 7	29.48 / 3	722.5/ 8	9.42 / 0

Table 4. Content Contributions by Roles

There is significant difference in the number of questions (χ^2=3759.85, p<0.001), answers (χ^2=4249.05, p<0.001) and technical edits (χ^2=3703.36, p<0.001) across user roles. Technical editors and starters ask the most questions and make more edits. Answerers answer more questions than other roles. Followers have the least content contributions of all roles.

We then plotted the number of content contributions users make by user roles, including number of questions posted (Figure 3) and number of answers posted (Figure 4). X-axis indicates the rank of a user and Y-axis indicates the number of contributions he/she makes. Both X-axis and Y-axis are in log scale.

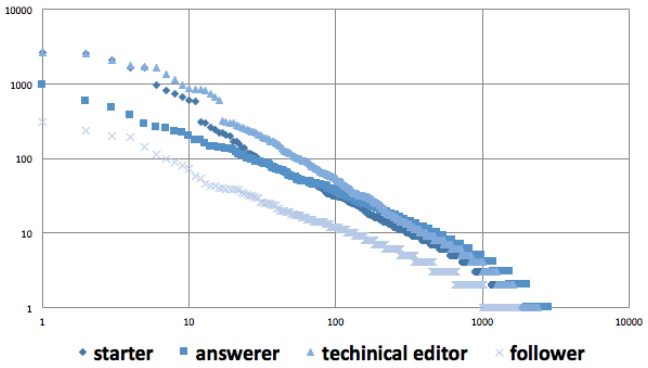

Figure 3. Number of Questions a Users Posted by Roles

As Figure 3 shows, the number of questions posted follows a power-law distribution: a few users posted most of the question while most users posted a few questions. Among higher-ranking users, starters and technical editors posted more questions, whereas answers and followers posted fewer questions. However, different user roles posted similar amount of questions among lower ranking users.

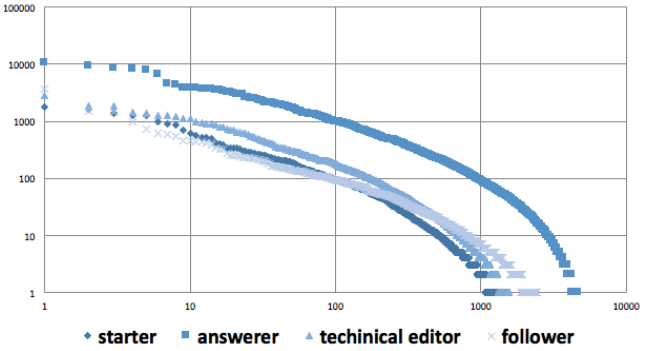

Figure 4. Number of Answers a User Posted by Roles

Figure 4 shows the number of answers posted by different roles follows a similar overall distribution. It seems that users who answer questions in one of these twelve topics answer a lot more questions in general than other roles.

Social Connection

We compared the number of topic followings, the number of user followings and the number of followers a user had by roles. Table 5 shows the results.

In Table 5, there is significant difference in the number of topic followings (χ^2=2370.73, p<0.001), user followings (χ^2=359.40, p<0.001) and followers (χ^2=1563.14, p<0.001) among user roles. Overall, followers follow a lot more topics and users than the other three roles, while answerers tend to follow more topics and users than starters and technical editors and have more followers than the other three roles.

[1] http://www.stats.gov.cn/tjsj/tjbz/hyflbz/

Mean / Median	Starters	Answerers	Technical editors	Followers
# of topic followings	22.99 / 10	32.09 / 16	31.72 / 11	237.5 / 40
# of user followings	48.42 / 5	79.2 / 19	82.65 / 8	94.04 / 10
# of followers	1175.11 / 0	1103.21 / 8	2265.86 / 1	77.88 / 1

Table 5. Social Connections by Roles

Figure 5 shows the number of topics a user follows by different roles.

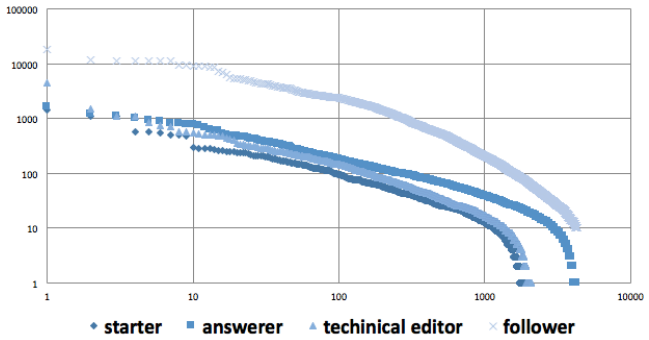

Figure 5. Number of Topic Followings by Roles

It seems that users who follow one of these 12 topics (followers) follow a lot more topics overall, while starters, answers, and technical editors follow similar amount of topics.

Popularity and Recognition

We compared the number of pageviews, votes and likes of a user received by roles. Table 6 shows the results:

Mean / Median	Starters	Answerers	Technical editors	Followers
# of pageviews	262.95 / 11	369.93 / 44	790.37 / 30	68.99 / 14
# of votes	382.14 / 0	1595.88 / 20	1160.24 / 0	123.57 / 0
# of likes	82.17 / 0	335.71 / 7	271.75 / 0	29.79 / 0

Table 6. Popularity & Recognition by Roles

There is significant difference in the number of pageviews (χ^2=1482.99, p<0.001), votes (χ^2=2976.92, p<0.001) and likes (χ^2=3100.68, p<0.001) across roles. In general, technical editors have more pageviews and answerers receive more votes and likes than other roles. While followers receive the fewest pageviews, votes and likes.

Figure 6 shows the number of votes a user's answers received by different roles.

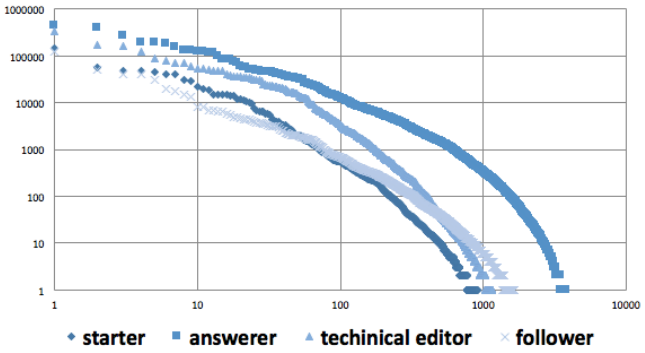

Figure 6. Number of Votes a User Receives by Roles

Answerers receive more votes than the other three roles. Among higher ranking users, answerers and technical editors receive more votes than starters and followers. It seems that users roles that provide more substantial content contributions gain higher popularity and recognition in the SQA.

Correlation Analysis

We conducted correlation tests to see how the activities are related. Results are shown in Table 7.

	Q	A	E	T	U	F	P	V	L
Questions (Q)	1								
Answers (A)	0.14 **	1							
Technical Edits (E)	0.229 **	0.025 **	1						
Topic followings (T)	0.017 *			1					
User followings (U)	0.151 **	0.096 **	0.064 **	0.237 **	1				
Followers (F)	0.38 **	0.165 **	0.099 **		0.165 **	1			
Pageviews (P)	0.362 **	0.252 **	0.13 **		0.209 **	0.881 **	1		
Votes (V)	0.19 **	0.405 **	0.038 **		0.099 **	0.538 **	0.655 **	1	
Likes (L)	0.196 **	0.369 **	0.045 **		0.105 **	0.6 **	0.685 **	0.95 **	1

Table 7. Correlations of Activities (p<0.01, * p<0.05)**

Number of questions is positively correlated mostly with the number of followers (correlation=0.38) and pageviews (correlation=0.362), whereas number of answers is positively correlated mostly with number of votes (correlation=0.405) and likes (correlation=0.369). Number of technical edits is positively correlated to number of questions posted (correlation=0.229), possibly because users who post questions also tend to edit questions and topics.

Number of topics and number of users a user follows are positively correlated (correlation=0.209). Users who follow more topics tend to follow more people.

The three popularity and recognition factors are all highly correlated pairwise. Users who get more pageviews also get more votes and likes. The correlation between votes and likes is extremely high (correlation=0.95).

CONCLUSION AND DISCUSSION

In this poster, we examined characterisics of four different SQA user roles, their content contributions, social connections, popularity and recognition. We found that users seek information from SQA in both active and passive ways, by actively asking a question (starters), or by following a topic or a user of interest to receive updates (followers). Male users participated more as answerers and technical editors, whereas female users participated more as starters and followers. This seems to suggest that men take more substantial roles in SQA, confirming the gender gap found in Wikipedia (Reagle & Rhue, 2011).

Starters and technical editors seem to share some common activity patterns, for example, they ask more questions, make more technical edits, but answer a few questions. They follow a few topics and users, and have very few followers. Answers contribute more content, attract the most followers and received the most likes and votes. Followers contribute little content, and receive the least likes and votes, but they follow more topics and users, suggesting that they are using the SQA site as a information source (Gazan, 2007). Future research is needed to further understand the interaction and relationship between these roles as well as other roles such as commenters, fans, and so on.

REFERENCES

Bowler, L., Oh, J. S., He, D., Mattern, E., & Jeng, W. (2012). Eating disorder questions in Yahoo! Answers: Information, conversation, or reflection?. *Proceedings of the American Society for Information Science and Technology*, *49*(1), 1-11.

De Wever, B., Van Keer, H., Schellens, T., & Valcke, M. (2010). Roles as a structuring tool in online discussion groups: The differential impact of different roles on social knowledge construction. *Computers in Human Behavior*, 26(4), 516-523.

Fox, S., & Duggan, M. (2013). Health online 2013. *Health*, 1-55.

Gazan, R. (2007). Seekers, sloths and social reference: Homework questions submitted to a question-answering community. *New Review of Hypermedia and Multimedia*, *13*(2), 239-248.

Gazan, R. (2010). Microcollaborations in a social Q&A community. *Information processing & management*, 46(6), 693-702.

Geiger, R. S., & Ribes, D. (2010, February). The work of sustaining order in wikipedia: the banning of a vandal. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 117-126. ACM.

Gleave, E., Welser, H. T., Lento, T. M., & Smith, M. A. (2009, January). A conceptual and operational definition of 'social role' in online community. *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference*, 1-11. IEEE.

Harper, F. M., Moy, D., & Konstan, J. A. (2009, April). Facts or friends: distinguishing informational and conversational questions in social Q&A sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 759-768. ACM.

Hansen, P., & Järvelin, K. (2005). Collaborative information retrieval in an information-intensive domain. *Information Processing & Management*, *41*(5), 1101-1119.

Haythornthwaite, C., & Hagar, C. (2005). The social worlds of the Web. *Annual review of information science and technology*, *39*(1), 311-346.

Hertzum, M. (2008). Collaborative information seeking: The combined activity of information seeking and collaborative grounding. *Information Processing & Management*, *44*(2), 957-962.

Kim S. (2010). Questioners' credibility judgments of answers in a social question and answer site. *Information Research*, *15*(2), 5.

Morris, M. R., Teevan, J., & Panovich, K. (2010, April). What do people ask their social networks, and why?: a survey study of status message q&a behavior. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1739-1748. ACM.

Preece, J. (2000). Online Communities: Designing Usability, Supporting Sociability. Jonh Wiley & Sons. *Inc., New York, NY, USA*.

Reagle, J., & Rhue, L. (2011). Gender bias in Wikipedia and Britannica. International Journal of Communication, 5, 21.

Shah, C., Oh, J. S., & Oh, S. (2008). Exploring characteristics and effects of user participation in online social Q&A sites. *First Monday*, *13*(9).

Shah, C., & Pomerantz, J. (2010, July). Evaluating and predicting answer quality in community QA. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, 411-418. ACM.

Turner, T. C., Smith, M. A., Fisher, D., & Welser, H. T. (2005). Picturing Usenet: Mapping computer-mediated collective action. *Journal of Computer-Mediated Communication*, *10*(4).

Wang, G., Gill, K., Mohanlal, M., Zheng, H., & Zhao, B. Y. (2013, May). Wisdom in the social crowd: an analysis of quora. In *Proceedings of the 22nd international conference on World Wide Web*, 1341-1352. ACM.

Zhang, P. (2012). Information seeking through microblog questions: The impact of social capital and relationships. *Proceedings of the American Society for Information Science and Technology*, *49*(1), 1-9.