Question and Answering Made Interactive

An Exploration of Interactions in Social Q&A

Zhe Liu and Bernard J. Jansen
College of Information Sciences and Technology
The Pennsylvania State University
University Park, Pennsylvania 16802
zul112@ist.psu.edu, jjansen@acm.org

***Abstract*— With the advancement of Web 2.0 techniques, social question and answering has become a new venue for individuals to seek for information online. Although it has been investigated by a number of works lately, so far still little has been known about how people interact with each other in order to satisfy their information needs in social Q&A. With the aim to understand the patterns of user interactions in the social Q&A context, as well as factors that may affect such kind of back-and-forth communications, in this work we collect over 1,000 question and answering dialogues from Sina Weibo. Statistical analyses including ANOVA, Pearson's correlation, linear regression and independent t-test are performed in order to answer our proposed research questions. Our results demonstrate the importance of studying the interactions in social Q&A given that about half of our collected question-answer pairs are of interactive nature. From the quantity perspective, we observe that questions within more complicated topics, such as "Healthcare" and "Education" generate more interactions. Significantly positive correlation is also noticed between social tie strength and the number of interactions. By manually annotating all interactive answers, we also indicate the importance of weak ties in providing high quality answers and interactions. Based on our results, we proposed potential implications for future design and implementations.**

***Index Terms*— Social question and answering, social Q&A, social search, social networks, information seeking, information exchange, Weibo**

I. INTRODUCTION

The huge rise in the popularity of social networking sites (SNSs) in the past decade has made them a preferable platform for people to catch up on news and real-time events that happened to their social ties. Everyday hundreds of millions of new pieces of content are posted by SNS users and shared with both their immediate networks and the larger Web community [1, 2]. This results in a relatively new data source with communicative structures that allowed collaborative information seeking to emerge in the recent years and informed the proposal of the concept of social question and answering (Q&A).

Most often social search is referred as the behavior of asking natural language questions to one's friends or followers in a network [7]. However, to better clarify what we mean when we are using the term of social Q&A, in this study, we give it a formal definition of: an elicitation for information communicated to a network of others via a web-based service in order to get a response that addresses the elicitation. Examples of social Q&A include: Anyone knows how to fix blinking monitor?; Can anyone recommend any good places to go for afternoon tea in central London?; #healthadvice Twitter I need help - how can I kick a cold/flu illness quickly?

Q&A on SNSs is different from the traditional information seeking techniques (e.g. search engine and online databases etc.) with more interactive search experience. Through back-and-forth exchange of messages on a question, social Q&A allows the questioners and the answerers to ask and respond further questions synchronously in real time. In order to understand the quantity and quality of interactions in social Q&A, we collect 1,003 information seeking posts along with their responses from Sina Weibo to explore the interaction characteristics and patterns demonstrated in user's social Q&A behaviors.

Considering the social nature of Q&A on SNSs, in this study we also introduced the concept of two types of answers in social Q&A: the informational and the conversational answers. By manually annotated nearly 3,000 answerer-generated replies into either informational or conversational reply types based on pre-defined coding criteria. We investigate the types of answers at different levels of questioner-answerer interactions, as well as across different topics and social ties.

Our findings suggest that questioner-answerer interactions are of huge importance to social Q&A. The number of interactions varied significantly across different topical categories and social tie strengths. Weak ties are important to answering one's question given the limitation to accessing more diverse solutions from the strong ties. Through our further analysis from the perspective of social Q&A answer types, we found that about half of the answers received are of social intentions, which generates more turns of inter-personal interactions as compared to those informational answers. Design implications based the above findings are also included in this study.

II. RELATED WORK

Due to the rapidly increasing popularity of SNS, more and more people choose to post their information needs to those virtual communities accessible to all their online friends. According to [3], almost 15% of everyday tweets contain information needs. Much of the prior studies in social Q&A investigate factors that motivate people to seek information via

978-0-7695-4998-9/13 $26.00 © 2013 IEEE
DOI 10.1109/SOCIETY.2013.16

social platforms [4], highlighting elements such as inter-personal trust and relatively easier access to timely information on social medias [5]. Besides, compared with traditional information retrieval, questions asked on SNS demonstrated more personalized information needs [6], as can be proved in [7] who found that the majority of the questions on SNS were recommendations and opinions.

In addition to study the motivations of social search, other research has been conducted to understand the taxonomy of questions asked on SNS. Through an analysis of 100 question tweets, [8] found that both factual and impersonal opinion questions were asked in social Q&A in order to satisfy one's information needs. Based on this result, a taxonomy of social Q&A questions was proposed from both the audience and the information needs perspectives. [9] conducted similar study using Broder's [10] proposed taxonomy of traditional search (transactional, navigational, and informational). A social search model of user activities before, during, and after search was presented based on question classifications and proved the value of social interactions in information seeking tasks. [11] evaluated the value of microblogging platform good place for asking questions. By analyzing question tweets, the authors found that due to the chatty nature of Twitter, rhetorical questions were the most popular form of questions asked, followed by questions seeking for factual knowledge.

Besides the above literatures on social Q&A questions, there are also studies conducted focusing on the answers. Given the low response rate demonstrated in social search [12], many researchers turned their research focus to professional Q&A sites, such as Yahoo! Answers, Answerbag etc, to explore the secret of high quality and high quantity response. [14] explored patterns demonstrated in knowledge sharing activities by clustering forum categories according to their content characteristics. A strong association was found between user's entropy (the broadness of user's focus) and the rating of the answers. Through a comparative field study, [15] investigated predictors of answer qualities across multiple online Q&A sites. The comparison indicated that fee-driven Q&A sites, such as Google Answers, received higher quality answers than other online but free sites. Extended from the above study, [13] provide an overview of the structural characteristics of activities happened on Q&A sites, followed by an explanation of both intrinsic motivations and extrinsic factors that affected the user participation in online Q&A environment.

There are also other literatures studying the quality of answers in addition to response quantities. A user study was performed in [16] to evaluate the effectiveness of social Q&A. By asking the participants to post a variant of the question to their Facebook status, the authors found that about half of the subjects received responses from their network before completing their search, which demonstrate the feasibility of using SNS for information seeking purpose. By performing statistical analysis based on real tweets, Paul et al. [17] noted that the majority of questions received no response on Twitter. They also found that among those few interrogative tweets with answers, the response rate is strongly related with some of the characteristics of the question askers, such as the size of their networks. Liu and Jansen [18] conducted analytic studies on Weibo posts and summarized eight factors that affect the response rate, including the number of followers, the number of @mentions, the usage of hashtags, etc.

To date, there is very limited work investigating the questioner-answerer interactions. However, interaction as the lifeblood of online social communities [4], mediated human's information seeking process by helping them to reconceptulize their information needs in the context of social Q&A [19]. By analyzing two online communities, [19] demonstrated the "mediation" and "expansion" [20] effects of community members, regarding their role in helping the information seekers to understand their information needs. In another work, Gazan further divided the answerers into synthesists and specialists [21] and the questioners into seekers and sloths [22], based on the different roles that they are playing in the process of social Q&A with the synthesists rated higher across all topics and specialists preferred within niche communities. Although findings from above studies are informative, there still lacks a comprehensive understanding regarding questioner-answerer interactions in social Q&A.

III. RESEARCH QUESTIONS

To address the gap, the goal of this study is to learn more about user's interactive information exchange behaviors in social Q&A. We organize this work by addressing both the quantity and quality perspectives of questioner-answerer interactions in social Q&A.

A. Quantity Aspect

First, from the quantity point of view, we propose the following research questions:

RQ1.1: *Does the overall interaction level between the questioner and the answerer differ with respect to the topic of question asked?*

Topic represents the information domain of a question [15]. It has been investigated in many previous studies as an important dimension in both professional and social Q&A. In both [15] and [18], topic has been used to successfully predict the quantity of the answers received in social Q&A. In addition, as indicated in [23] synchronous Q&A should leverage the topical knowledge of both the questioner and the answerer in order to proceed the question answering. For instance, in [15], the authors found that entertainment-oriented questions received more replies than the other topical categories. However, in [18], the location specific questions attracted the most answers. Based on those findings and our own experience, we hypothesize that

> **H1.1:** *topics require only common sense familiarity could lead to more Q&A interactions than those demand professional knowledge.*

Here by saying topics require only common sense familiarity, we refer to topics including "Life", "Business" (e.g. "where to buy", "opens at what time" etc.), and "Entertainment". The rest three topical categories: "Healthcare", "Education" and "Technology" are defined as

the topics that demand professional knowledge. This classification is based on authors' own experiences. More rigorous classification is needed for future studies.

RQ1.2: *To what extent can social tie strength predict the interaction level between the questioner and the answerer?*

Ties are interaction links between nodes in a social network and represent the relationship between nodes in a social space [24]. Since strong and weak ties are compared in terms of their contributions to the information flow between two nodes [25], generally we can find that the social aspect predominantly conditions users' interactions on SNSs, as proved [26]. Consistent with previous results, in this work we assume that people are more willing to share their knowledge with their strong relationships [27]. So we hypothesize that

H1.2: *stronger social ties could lead to more Q&A interactions than those weak social ties.*

Quality Aspect

Second, from the quality point of view, in order to measure the quality of user responses, we first define two types of answers: informational answers and conversational answers, *Informational answers are those in which respondents provide direct information that is on-topic and valid. Conversational answers are comments motivated by the desire to chat, which are irrelevant to the question.* For instance, for question "Anybody knows where I can find a decent custom suit?" an informational answer would be "The tailor's shop near the Agriculture University", while a conversational answer could be expressed as "Why do you need a custom suit?"

Based on the two types of answers, three more corresponding research questions are addressed in this study:

RQ2.1: *Does the number of two types of answerers, informational and conversational, differ with respect to the topic of question asked?*

Contrary to our assumptions made for RQ1.1, regarding this research question, we hypothesize that

H2.1: *topics require only common sense familiarity could lead to more conversational Q&A interactions than those demand professional knowledge.*

[15] proved our assumption given that the replies to those entertainment-oriented questions were poor in judged quality relative to other topics.

RQ2.2: *Does the number of two types of answerers, informational and conversational, differ with respect to their social tie strength?*

With this research question, we aim to calculate the mean strength of social ties of question answerers of two different types, providers of those informational answers and providers of those conversational ones. Krackhardt [33] pointed that in case of uncertainty, people choose to rely on their strong ties which constitute a base of trust that provide comfort in the face of uncertainty. Based on that, we hypothesize that

H2.2: *Stronger social ties could lead to more informational answers than those weak social ties.*

RQ2.3: *Does the level of questioner-answerer interactions differ with respect to the two types of answerers, informational and conversational?*

Given the interactive nature of SNSs, one extraordinary feature of social Q&A is that it allows the questioners and the answerers to seek for follow-up interactions in order to solve their problems [28]. So with this research question, we hypothesize that

H2.3: *informational answers trigger more interactions than those conversational answers.*

IV. METHOD

A. Data Set

In this study, we collected data from China's largest microblogging site, Sina Weibo. Launched in 2009, Weibo attracted nearly 30 million users within only three years, with a current average of 10 million Weibo statuses published every day [30], at the time of the study. Weibo essentially adopts the same operating concept and provides very similar functions to its users as Twitter.

The main reason to select Sina Weibo over Twitter as the data source in our study is because of its more user-friendly replying mechanism. Rather than mixing the replying tweets together with the @replies, as Twitter does, Weibo's threaded comment feature obviously makes the responding process not only simpler, but also more organized. It can be seen in Figure 1, with Weibo user's replying and displaying actions all done in one section. Weibo's reply function makes the social feedback process much easier for its users, and thus could be a better source for studies focusing on the behaviors of social response.

In addition, another consideration that we chose Weibo in this study is because of its richer content. Due to the fact that Chinese characters are logograms rather than phonograms, the same number of Chinese characters can convey more information than English letters. Therefore, with the same 140-character limit, Weibo users can post much more elaborated questions and answers compared to Twitter users.

Using Weibo Search API, we collected 1,003 Weibo questions (published between October 8th 2012 and October 15th 2012), together with all their replies in a week period. We adopted keywords "anybody knows" (请问有谁知道) plus question mark (?) to retrieve the posts with information seeking purpose. Modified snowball sampling method was conducted to identify those above keywords used for the task of question extraction.

B. Topical Categorization

We employed a categorization method by automatically submitting each of the collected questions into Baidu Zhidao (http://zhidao.baidu.com/) and retrieving their returned classifications. As the most famous professional Q&A site, Baidu Zhidao has a comprehensive hierarchical taxonomy of 14 main categories and a number of corresponding sub-categories within each main category (Figure 2). The most frequently occurred main category on the first returned page would be assigned to the question as its topic.

Fig. 1. Layout of Weibo replying

In order to keep our classification in this work meaningful, we manually reclassified all posts with the returned categories of "Region" and "Unknown". Besides, we merged two categories "Computer/ Network" and "Electronics" given that questions in those two categories could not be distinguished reliably even with human taggers. In total, six major categories were selected based on our dataset, which includes "Healthcare", "Business", "Entertainment", "Education", "Life" and "Technology".

Fig. 2. Layout of Baidu Zhidao

C. Tie Strength Generation

There are several past studies introducing methods of modeling social tie strengths between Facebook users [31, 36]. However, given the huge differences between Weibo and Facebook, in this work, we followed the method proposed by Chen et al. [37] to generate the social tie strength between questioners and answers.

In their method, Chen et al. introduced two types of scores, the communication score and the mutual friend score in order to measure the inter-user social tie strength. Communication score is defined as the logarithm of the number of @replies that both users A and B have posted between each other; while the mutual friend score refer to the average communication score of all the mutual bi-directional friends between A and B. The final tie-strength score for B is the sum of his/her communication score and mutual friend score. Although Chen et al.'s modeling of social tie strength has proved to be accurate through their following user studies, considering the different requirement of social information seeking and sharing. In this study, we modified Chen et al.'s method and proposed in this section our "customized" social tie strength calculation.

Suppose we are aiming to calculate the social tie strength of user B to user A in social Q&A, where A is the questioner and B is the potential answerer. First, instead of including all bi-directional replies between the individuals, in this work, communication score is estimated as the logarithm of the total number of responses that B contributed to A's most recent 20 posts. We made this modification given the fact that B whether or not answers A's question totally depends on B's preference towards A, but not rely on A's attitude towards B. We chose the most recent 20 posts rather than all published status considering the dynamic nature of interpersonal relationships on SNSs. In the same way, we computed the mutual friend score by first finding all mutual friends between A and B, then calculating the logarithm of the total number of responses that each mutual friend contributed to A's most recent 20 posts, and finally taking the average of all previously computed scores. We finally take the sum of the calculated communication score and the mutual friend score as the social tie strength for user B to user A.

So in our method, B would be considered a strong-tie of A if he/she replied many times under A's most recent 20 posts, or if a majority of B's mutual friends with A have done so.

D. Answer Types

According to our definition of informational and conversational answers, the first author of this paper manually classified all received replies (not include replies from the original questioners) into informational and conversational types. A second trained individual was then recruited to perform the identical annotation task on a randomly selected subset of question-answer pairs (250 questions together with their answers), without conferring with the first annotator. The resulting Kappa statistic of 0.86 indicates good interrater reliability.

E. Data Analysis

In this study, we first performed a descriptive analysis to gain a picture of the interaction patterns in social Q&A context. After that, in order to answer RQ1.1 (interaction differences across topics), we performed an ANOVA test. To better display the differences in interaction frequencies across topics, Tukey's HSD test was then carried out. We then conducted Pearson's correlation to explore the relationship between the strength of social tie and the number of interactions across topics (RQ1.2). To answer RQ2.1, we conducted cross tab analysis with chi-square test to examine the relationships between the types of answers, informational answers or conversational ones, across all six topics. Independent t-test was finally performed to answer our research question RQ2.2 (social tie strength differences between two types of answers)

and RQ2.3 (interaction differences between two types of answers).

All statistical analyses were performed using SPSS with a pre-set significance level of 0.05 and all non-normally distributed variables were transformed into normal via the Box-Cox power transformation [32] using log10(variable+1).

V. RESULTS

A. Quantity Aspect

Through our descriptive analysis, we noticed that about half (409, 40.78%) of our collected posts received only 0 (329, 32.80%) or 1 reply (79, 7.89%). This is inconsistent with results as indicated in [17] which showed that the majority of questions received no response on Twitter. Among the 594 (59.22%) questions which received 2 or more than 2 replies, 523 (88.05%) of them contained at least one questioner-answerer interaction. In total, 4,901 replies were collected for all 1,003 question posts, indicating that the average reply per question is 4.89, which is relatively high. Among the 4,901 total replies, 4,572 of them are interactive ones, generating 2,206 interactions between the unique questioner and answerer pairs. 1365 (61.88%) of those interaction are less than two turns, 476 (21.58%) of them occurred for 2 or 2.5 turns, and 191 (8.66) for 3 or 3.5 turns. Around 96% of the interactions are less than 5 rounds in our collected dataset as can be seen from Figure 3.

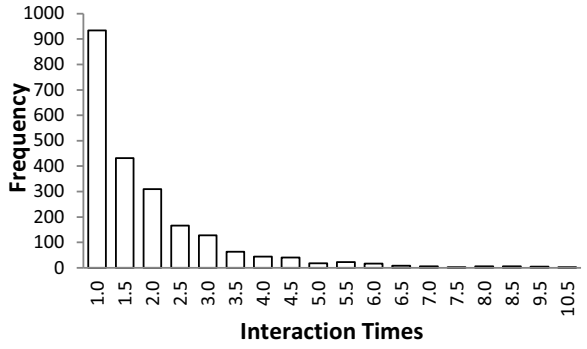

Fig. 3. Interaction frequency distribution

RQ1.1: *Does the overall interaction level between the questioner and the answerer differ with respect to the topic of question asked?*

H1.1: *topics require only common sense familiarity could lead to more Q&A interactions than those demand professional knowledge.*

From the ANOVA test, a significant main effect was observed for topics on the average times of questioner-answerer interactions ($F = 12.11$, $df = 5$, $p = 0.00 < 0.05$). As can be seen from Table I, Tukey's post hoc comparisons further indicated significant higher interaction times of the "Healthcare" and "Education" oriented questions, as compared to the ones from categories of "Business", "Entertainment" and "Life". On average the "Healthcare" related questions generated 2.95 times of back-and-forth interactions, while the average interaction times for questions from the "Education" category is 2.85. As compared to the questions under the topical categories of "Healthcare" and "Education", "Life" (2.56) and "Business" (2.46) related questions generated significantly lower times of questioner-answerer interactions. Among all six topical categories, "Entertainment" (2.25) and "Technology" related questions had the lowest number of average interaction frequencies (2.36).

Given the relatively more times of interactions generated by the "Healthcare" and "Education" topics, we can reject our H1.1. However, due to the relatively lower level of interaction times generated by those "Technology" oriented questions, we also failed to conclude that more complicated topics that require professional knowledge could lead to more Q&A interactions than those demand only common sense familiarity.

We assume that one possible explanation for this result is the fact that posts with higher complexity may demand more follow-up questions for clarification purpose. One example that supports our assumption is that:

Questioner (18:52PM): "Anybody knows any way to cure the migraine?"
Answerer (19:15PM): "Take Tianshu capsule."
Questioner (20:31PM): "Should I take the pill regularly or should I just take it when my head hurts?"
Answerer (22:12PM): "You should take it regularly. What causes your migraine? Nerve disorder?"
Questioner (22:20PM): "Not me, I am asking for someone else."
Answerer (22:21PM): "OK."
Questioner (22:24PM): "Oh, forget to ask is there any other way without taking the medicine?"

TABLE I. AVERAGE INTERACTION FREQUENCIES BY TOPICAL CATEGORIES

Topical Categories	Average Interaction Frequencies	SD of Interaction Frequency
Life $_a$	2.56	2.11
Entertainment $_b$	2.25	1.74
Business $_a$	2.46	1.63
Healthcare $_c$	2.95	1.98
Education $_c$	2.85	1.96
Technology $_{ab}$	2.36	1.77

Note: Topical Categories containing similar letters are non-significantly statistically different in average interaction frequency by Tukey's post hoc test results at $p < 0.05$.

RQ1.2: *To what extent can social tie strength predict the interaction level between the questioner and the answerer?*

H1.2: *stronger social ties could lead to more Q&A interactions than those weak social ties.*

Through our calculation of social tie strength, we found that the average social tie strength among all 2,206 interaction pairs is 0.51, with the standard deviation of 0.60 (Table II). Individuals interacted under the topic of "Healthcare" shares the strongest social tie, whereas, interactions within the "Business" category showed the weakest community structure.

TABLE II. AVERAGE SOCIAL TIE STRENGTH FREQUENCIES BY TOPICAL CATEGORIES

Topical Categories	**Average Social Tie Strength**	**SD of Social Tie Strength**

Life	0.50	0.57
Entertainment	0.51	0.62
Business	0.32	0.46
Healthcare	0.72	0.69
Education	0.58	0.72
Technology	0.40	0.52
Overall	0.51	0.60

Overall, without separating the posts according to their topical categories, the Pearson's correlation showed positive but weak relationship between the strength of social tie and the number of interactions ($r = 0.16$, $p = 0.00$). Pearson's correlation coefficient is interpreted as follows: $r \leq 0.3$ were interpreted as a weak correlation, $0.3 < r < 0.5$ as a moderate correlation, and $r \geq 0.5$ as a strong correlation [29].

After conducting correlation analysis between interaction times and social tie strength for each topical category separately, we found that again the Pearson's correlation coefficient revealed weak-to-moderate but statistically significant positive associations between tie strength and questioner-answerer interactions across all six topics, except the category of "Healthcare" (Table III). In that sense, we failed to reject the null hypothesis under our RQ, and therefore conclude that stronger tie strength drives people to interact more in the process of information exchange, especially under the topic of "Healthcare" ($r = 0.50$, $p = 0.00$).

TABLE III. CORRELATION COEFFICIENT RESULTS

Topical Categories	r	P
Life	0.01	0.59
Entertainment	0.09	0.02*
Business	0.17	0.00*
Healthcare	0.50	0.00*
Education	0.24	0.00*
Technology	0.18	0.00*
Overall	0.16	0.00*

One possible explanation for this phenomenon is that "strong ties constitute a base of trust that can reduce resistance and provide comfort in the face of uncertainty" [33], so that under the case of solving more serious and complicated problems (such as Healthcare) people tend to rely more on their close contacts rather than casual acquaintances. In that case, if the social Q&A platform could transfer trust information or reputation from one connected community to another, then it would benefit the social Q&A process by allowing the questioners to exchange more information with the answerers.

B. Quality Aspect

Among all 4,572 of interactive replies, 1,893 of them were produced by the questioner him/herself. Only 2,679 of the replies were generated by the answerers. To our surprise, after eliminating 7 malicious answers, we found that among all 2,679 non-self-replied answers, 1,357 (50.65%) of them are conversational, whereas the rest (1,315, 49.09%) are informational. This result pointed out a future need to investigate the effectiveness of social Q&A regarding the received answer qualities.

Through our further analysis, we also found that among 523 questions which contain interactive replies, 420 (80.31%) of them received at least one informational non-self-response answer, whereas 395 (75.53%) out of the 523 questions contain at least one conversational response. This is a very interesting finding given that the overall number for the conversational answers received is larger than that of the informational ones, whereas the number of questions received conversational answers is less than the number of questions received informational responses. This indicates the existence of a few questions which may contain a high number of conversational answers as compared with the others. Our further analysis on topical categories, social tie strength and interaction levels would further investigate this issue.

RQ2.1: *Does the number of two types of answerers, informational and conversational, differ with respect to the topic of question asked?*

H2.1: *topics require only common sense familiarity could lead to more conversational Q&A interactions than those demand professional knowledge.*

From table IV, we can see that among all six topical groups, the category of "Life" contains the most questions, which indicates that people use SNSs to ask daily life questions, such as where to eat, how to get somewhere, and where to find something, etc. The least popular topic among all six categories is "Business", which contains only about ¼ of the questions as compared to the category of "Life".

When we cross tabbed the two types of answers by six topical categories, results clearly showed that there are significant differences among the types of answers received (chi-square = 64.06, df = 5, p=0.01). As can see from table IV, the topical categories of "Technology" (59.9%), "Life" (52.9%) and "Healthcare" (52.2%) attracted the most number of informational answers. Although without detailed qualitative analysis, through our observations we assumed that the high number of informational answers for "Technology" and "Healthcare" related questions may be due to two reasons. First, since those questions are what we called "questions require professional background", only people with the demanded knowledge tend to answer them seriously. Besides, given the complexity of those questions, as well as the lack of common ground between the questioners and the answerers, questions from the above two categories have higher probability of receiving follow-up questions for answer explanation or background clarification. This increases the chance for those questions to receive more informative answers rather than conversational replies.

We also found that among all six topics, two topical categories, "Entertainment" and "Education", received more conversational answers than informational ones. Among all the interactive answers received for "Entertainment" oriented questions, 63.5% (270) of all the answers were with the social intentions, whereas only 36.5% (155) of them are informative. Such high percentage of conversational answers became understandable when considering the entertaining nature that underlying behind those entertainment-related questions themselves. Similarly, only 38.3% (114) of the

received answers under the "Education" category is informational, the majority of its replies are only of social intentions. This finding indicated that although people choose SNSs as ideal platforms to ask "Education" related questions, the results are not as relevant as they might expected. This could be due to young people being more inclined than adults to use the internet for socializing [34]. However, further studies are needed to fully understand this phenomenon and to access whether or not SNSs are good platform for Education purposes.

RQ2.2: *Does the number of two types of answerers, informational and conversational, differ with respect to their social tie strength?*

H2.2: *Stronger social ties could lead to more informational answers than those weak social ties.*

Tested on the whole dataset, the independent t-test demonstrated significant differences in social tie strength between informational and conversational answers ($t = -8.329$, $df = 2670$, $p = 0.00 < 0.05$). As can be seen from Table VI, the average social tie strength for the informational answerers is.1209, however, the average social tie strength for the conversational answerers is 0.17. Given that we can conclude weak ties affect in general social Q&A.

After analyzing the overall social tie strength across different answer types, we also tested that within each individual topical category separately. As demonstrated in Table VI, we can see that all conversational answers across topics are generated by relatively strong ties as compared to the average tie strength of those informational answerers. However, the results of independent t-test for each topic indicated different findings.

With independent t-test, we noticed that topical categories "Life" ($t = -5.87$, $df = 992$, $p = 0.00 < 0.05$), "Business" ($t = -2.30$, $df = 280$, $p = 0.01 < 0.05$), "Health" ($t = -3.07$, $df = 337$, $p = 0.04 < 0.05$) and "Education" ($t = --5.35$, $df = 296$, $p = 0.04 < 0.05$) showed significant differences in social tie strength between informational and conversational answers. Although the rest two categories "Entertainment" ($t = -1.57$, $df = 423$, $p = 0.12 > 0.05$) and "Technology" ($t = -1.08$, $df = 332$, $p = 0.50 > 0.05$) also revealed stronger social tie strength on conversational answerers, such difference is not statistically significant. We try to explain this pattern of behaviors from the standpoint of Granovetter's [30] weak tie theory, attributing it to weak ties advantages in accessing more diverse solutions. According to Granovetter's theory, strong ties may provide more credible and readily available information sources for people. However, due to their constraints on the breadth and the non-redundancy of information, strong ties also limit people's opportunities to access more diverse solutions. In the context of social Q&A, weak tie theory can be explained as the higher probability for weak ties to know the answer, given the common background that the strong tie may have as questioners.

Given the significantly different contributions between strong ties and weak ties to "Life", "Business", "Health" and "Education" oriented questions, it would be better to get questions under those categories shown to as many weak ties as possible, rather than limited it to only close friends. Only in this way the system can benefit the questioner with more informative answers.

TABLE IV. CROSS-TAB RESULTS ON TOPICS AND QUESTION TYPES

			Topic						
			Life	Entertainment	Business	Healthcare	Education	Technology	Total
Question Type	Conversational	Count	468	270	139	162	184	134	1357
		% within tag	34.5%	19.9%	10.2%	11.9%	13.6%	9.9%	100.0%
		% within topic	47.1%	63.5%	49.3%	47.8%	61.7%	40.1%	50.8%
		% of Total	17.5%	10.1%	5.2%	6.1%	6.9%	5.0%	50.8%
	Informational	Count	526	155	143	177	114	200	1315
		% within tag	40.0%	11.8%	10.9%	13.5%	8.7%	15.2%	100.0%
		% within topic	52.9%	36.5%	50.7%	52.2%	38.3%	59.9%	49.2%
		% of Total	19.7%	5.8%	5.4%	6.6%	4.3%	7.5%	49.2%
Total		Count	994	425	282	339	298	334	2672
		% within tag	37.2%	15.9%	10.6%	12.7%	11.2%	12.5%	100.0%
		% within topic	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	37.2%	15.9%	10.6%	12.7%	11.2%	12.5%	100.0%

TABLE V. INDEPENDENT T TEST RESULTS ON SOCIAL TIE STRENGTH BETWEEN ANSWER TYPES ACROSS TOPICS

Topic	Answer Type	Mean social tie strength	SD social tie strength	t	df	p
Life	Informational	0.12	0.14	-5.87	992	0.00*
	Conversational	0.18	0.16			
Entertainment	Informational	0.13	0.16	-1.57	423	0.12
	Conversational	0.16	0.15			
Business	Informational	0.08	0.12	-2.30	280	0.02*
	Conversational	0.11	0.14			
Healthcare	Informational	0.17	0.16	-3.7	337	0.00*
	Conversational	0.22	0.18			
Education	Informational	0.10	0.13	-5.35	296	0.00*
	Conversational	0.20	0.18			
Technology	Informational	0.11	0.14	-1.08	332	0.28
	Conversational	0.13	0.14			
Total	Informational	0.12	0.14	-8.33	2670	0.00*
	Conversational	0.17	0.16			

TABLE VI. INDEPENDENT T TEST RESULTS ON INTERACTION TIMES BETWEEN ANSWER TYPES ACROSS TOPICS

Topic	Answer Type	Mean interaction times	SD interaction times	t	df	p
Life	Informational	2.06	1.66	-5.53	992	0.00*
	Conversational	2.79	2.37			
Entertainment	Informational	1.74	1.41	-3.54	423	0.00*
	Conversational	2.30	1.80			
Business	Informational	2.05	1.44	-4.02	280	0.00*
	Conversational	2.72	1.72			
Healthcare	Informational	2.42	1.90	-3.53	337	0.00*
	Conversational	3.17	2.02			
Education	Informational	1.92	1.59	-5.69	296	0.00*
	Conversational	3.07	2.01			
Technology	Informational	1.95	1.48	-2.87	332	0.00*
	Conversational	2.58	2.03			
Total	Informational	2.04	1.62	-10.04	2670	0.00*
	Conversational	2.75	2.10			

RQ2.3: *Does the level of questioner-answerer interactions differ with respect to the two types of answerers, informational and conversational?*

H2.3: *informational answers trigger more interactions than those conversational answers.*

Overall, the independent t-test demonstrated significant differences in interaction levels between informational and conversational answer types ($t = -10.04$, $df = 2670$, $p = 0.00 < 0.05$). As can be seen from Table VI, the average interaction times for the informational answerers is 2.04, which is significantly lower than the interaction time for those conversational answerers, with the average interaction times of 2.75. This indicated that given the interactive design of SNSs, people tend to interact more with each other when talking about things irrelevant to their information needs in the social Q&A environment.

A further analysis of the interaction differences between informational and conversational answers across topics was then conducted by following the same method as we have introduced under RQ2.2. Table VI shows the results, in which all conversational answerers engaged in more interactions across all six topics. Although all significant, when comparing the distinctions between informational and conversational interactions within groups, we found that the topical category "Education" contains the biggest difference. The conversational interactions triggered by the "Education" related questions were on average about 1.60 times of the informational interactions. This again demonstrated the drawback of using SNSs to ask "Education" related questions.

VI. DISCUSSION AND IMPLICATIONS

Questioner-answerer interaction plays an important role in actual Q&A experience. People interact back-and-forth with each other in the Q&A process in order to confirm the intention of the questioners or to clarify the background context of the questions being asked, etc. Through an analysis of the real world microblogging questions, together with their answers, our study showed that there were a huge amount of interactions happened in the context of social Q&A. About half of the questions posted on SNSs contained questioner-answerer interactions. The distribution of the communication patterns roughly followed a power law distribution with most of the interactions happened once or twice. Although interaction should be given attentions considering its popularity and influence in actual Q&A, currently there lacks a good understanding on the characteristics and patterns of back-and-forth information exchange in social information seeking. Most of the social search engines recently proposed rely only on the content of the question, the expertise of the potential

answerers or the inter-personal social relationships or etc. However, if those social information seeking tools can take user interactions into consideration and provide scaffolding structures for knowledge re-production, it could better help the information seekers to re-construct their information needs and to reduce the potential interactions needed for common ground building. Our analysis results of RQ1.1 further supported the demand of such kind of design implications, given the significantly higher number of interactions occurred within topics that require professional knowledge.

By evaluating the quality of the received answers, we found that one big challenge in social Q&A is the abundance of question-irrelevant replies, in this study, we called them conversational answers. Compared with those informational answers, the conversational replies are generated with the intention to chat or to social with one another. They actually do not provide any useful and relevant information to the information seekers. Ignoring those conversational answers when indexing all the replies can save a lot of both computational and storage resources and time. So how to automatically differentiate conversational answers from the informational ones would become a very important and influential topic for future studies.

One more challenge that our results demonstrated for social Q&A is the relatively higher number of conversational answers received under those "Education" related questions. This might be due to young people's preference to use SNSs for socializing. Given the huge potential of making SNSs a good platform for education purpose, how to control those conversations during the learning process should also be taking into consideration.

Last, we believe that our analysis for RQ 2.2 further emphasized the importance of weak ties in one's network, specifically for the task of social Q&A. We found from our results that most of the valid and relevant answers were provided by those weak ties, whereas strong ties tend to participate more in interactions with pure social intentions, given the chatty nature of SNSs. With our results of the social tie differences across topics as shown in Table IV, we believe that our study is of great value to the design and development of targeted question routing. For instance, according to our results, "Healthcare" oriented questions should be routed to weak tie rather than strong ties, since the latter may limit people's opportunities to access more diverse solutions. With this kind of consideration of routing questions to one's weak ties or even strangers, there should also be some modification to the current design of SNSs, given that with the current SNSs services, one's post can hardly be exposure to people outside of his/her network.

Given the above potential implications, we believe our study is of great value to the social Q&A field. Limitations of our current work include the small sample set as limited by our manually annotation method. In our future studies, we aim to find a method to automatically classify the answers into conversational and informational. In that way, we can re-conduct studies to prove the generalizability of this work. Besides, we need more rigorous qualitative studies in the future to further understand some of the findings in this work.

VII. CONCLUSION

With the objective to understand user cooperation in social Q&A context, we performed an analysis on inter-personal interactions from both the quantity and the quality perspectives. Using Weibo as data source, we performed statistical analysis on the variance of questioner-answerer interaction numbers and qualities across different topics. We also analyzed the relationship between interaction numbers and the social tie strength. Followed by that, investigations on the social tie and interaction differences between informational and conversational answers were also performed. Our results demonstrate the importance of interactions in social Q&A given a huge number of back-and-forth communications found in our dataset. However, when taking answer quality into consideration, the large proportion of those interactions contained conversational answers, which is irrelevant for the task of information seeking. In addition, findings from our studies suggested weak tie effect, as well as interaction differences in answering certain types of questions.

Based on our findings, we also proposed possible implications and challenges in the future works of social Q&A, including adopting inter-personal interactions into the development of social search engines and modifying the current design of SNSs by allowing questions to be exposed to more people, etc. We believe that our study offers valuable insights into the future development of social search systems or tools that can make good use of those features as introduced in this study.

VIII. REFERENCE

1. Twitter.numbers.
   http://blog.twitter.com/2011/03/numbers.html.
2. Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. *Twitter power: Tweets as electronic word of mouth*. Journal of the American society for information science and technology, 2009, 60(11), 2169-2188.
3. Efron, M. and M. Winget, *Questions are content: a taxonomy of questions in a microblogging environment*. ASIST 2010, 2010.
4. Gazan, R. *When online communities become self-aware in System Sciences*, 2009. HICSS'09.
5. Teevan, J., D. Ramage, and M.R. Morris. # *TwitterSearch: a comparison of microblog search and web search*. 4th ACM international conference on Web search and data mining. 2011.
6. Liu, Z. and B.J. Jansen, *Almighty Twitter, What Are People Asking for?* ASIST 2012, 2012.
7. Morris, M.R., J. Teevan, and K. Panovich. *What do people ask their social networks, and why?: a survey study of status message q&a behavior*. 28th international conference on Human factors in computing systems. 2010.
8. Efron, M., & Winget, M. *Questions are content: a taxonomy of questions in a microblogging environment*. Proceedings of the American Society for Information Science and Technology, 2010, 47(1), 1-10.
9. Evans, B. M., & Chi, E. H. *Towards a model of understanding social search*. In Proceedings of the 2008 ACM conference on Computer supported cooperative work, 2008, pp. 485-494.
10. Broder, A. *A taxonomy of web search*. In ACM Sigir forum, 2002, 36(2), pp. 3-10.

11. Paul, S. A., Hong, L., & Chi, E. H. *Is twitter a good place for asking questions? a characterization study*. In Proceedings of ICWSM'11, 2011.
12. Sysomos. Replies and Retweets on Twitter. http://sysomos.com/insidetwitter/engagement/
13. Raban, D., & Harper, F. *Motivations for answering questions online*. New Media and Innovative Technologies, 2008.
14. Adamic, L. A., Zhang, J., Bakshy, E., & Ackerman, M. S. *Knowledge sharing and yahoo answers: everyone knows something.* In Proceedings of the 17th international conference on World Wide Web, 2008, pp. 665-674.
15. Harper, F. M., Raban, D., Rafaeli, S., & Konstan, J. A. *Predictors of answer quality in online Q&A sites*. In Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, 2008, pp. 865-874.
16. Teevan, J., Morris, M. R., & Panovich, K. *Factors Affecting Response Quantity, Quality, and Speed for Questions Asked via Social Network Status Messages*. In Proceedings of ICWSM'11, 2011.
17. Paul, S. A., Hong, L., & Chi, E. H. *Is twitter a good place for asking questions? a characterization study*. In Proceedings of ICWSM'11, 2011.
18. Liu, Z. & Jansen, B. J. *Factors influencing response rate in social question & answering behavior*. In Proceedings of CSCW'13, 2013.
19. Lueg, C., *Beyond FAQs: From information sharing to knowledge generation in online travel communities*. From Information Provision to Knowledge Production, 2008, p. 105.
20. Lueg, C. *Mediation, expansion and immediacy: how online communities revolutionize information access in the tourism sector*, ECIS, 2006.
21. Gazan, R., *Specialists and synthesists in a question answering community*. Proceedings of the American Society for Information Science and Technology, 2006. 43(1): p. 1-10.
22. Gazan, R., *Seekers, sloths and social reference: Homework questions submitted to a question-answering community*. New Review of Hypermedia and Multimedia, 2007. 13(2): p. 239-248.
23. White, R.W. and M. Richardson. *Effects of expertise differences in synchronous social Q&A*. in Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval. 2012.
24. Wasserman, S. and Faust, K. *Social network analysis - methods and applications*, 1994, Cambridge University Press.
25. Friedkin, N. E. *Information flow through strong and weak ties in intraorganizational social networks*. Social Networks, 1982.
26. Sousa, D., Sarmento, L., & Mendes Rodrigues, E. *Characterization of the twitter@ replies network: are user ties social or topical?*. In Proceedings of the 2nd international workshop on Search and mining user-generated contents, 2010, pp. 63-70.
27. Hall, H. and G. Widen-Wulff, *Social exchange, social capital and information sharing in online environments: lessons from three case studies*. USE-2008, 2008.
28. Kim, S., Oh, J. S., & Oh, S. *Best-answer selection criteria in a social Q&A site from the user-oriented relevance perspective.* Proceedings of the American Society for Information Science and Technology, 2007, 44(1), 1-15.
29. Majeski, C. J., Johnson, J. A., Davison, S. N., & Lauzon, G. J. *Itch Severity Scale: a self-report instrument for the measurement of pruritus severity. British Journal of Dermatology*, 2007, 156(4), 667-673.
30. Granovetter, M. S. *The strength of weak ties*. American journal of sociology, 1973, 1360-1380.
31. Gilbert, E., & Karahalios, K. *Predicting tie strength with social media*. In Proceedings of the 27th international conference on Human factors in computing systems, 2009, pp. 211-220.
32. Box, G. E., & Cox, D. R. *An analysis of transformations*. Journal of the Royal Statistical Society, 1964, Series B (Methodological), 211-252.
33. Krackhardt, D., The strength of strong ties: *The importance of philos in organizations*. Networks and organizations: Structure, form, and action, 1992. 216: p. 239.
34. Morris, M. R., Teevan, J., & Panovich, K. *A comparison of information seeking using search engines and social networks*. In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, 2010.
35. Fukumoto, J. I., Kato, T., & Masui, F. Question Answering Challenge (QAC-1) An Evaluation of Question Answering Task at NTCIR Workshop 3. In Proceedings of the AAAI Spring Symposium, 2003, pp. 122-133.
36. Panovich, K., Miller, R., & Karger, D. *Tie strength in question & answer on social network sites*. In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, 2012, pp. 1057-1066
37. Chen, J., Nairn, R., Nelson, L., Bernstein, M., & Chi, E. *Short and tweet: experiments on recommending content from information streams.* In Proceedings of the 28th international conference on Human factors in computing systems, 2010, pp. 1185-1194.