Information Processing and Management 54 (2018) 159–174

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Questioner or question: Predicting the response rate in social question and answering on Sina Weibo

Liu Zhe[a,*], Bernard J. Jansen[b]

[a] College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA 16802, United States
[b] Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar

ARTICLE INFO

Keywords:
Social Q&A
Social network
Information seeking
Weibo

ABSTRACT

With the noted popularity of social networking sites, people increasingly rely on these social networks to address their information needs. Although social question and answering is potentially an important venue seeking information online, it, unfortunately, suffers from a problem of low response rate, with the majority of questions receiving no response. To understand why the response rate of social question and answering is low and hopefully to increase it in the future, this research analyzes extrinsic factors that may influence the response probability of questions posted on Sina Weibo. We propose 17 influential factors from 2 different perspectives: the content of the question, and the characteristics of the questioner. We also train a prediction model to forecast a question's likelihood of being responded based on the proposed features We test our predictive model on more than 60,000 real-world questions posted on Weibo, which generate more than 600,000 responses. Findings show that a Weibo's question answerability is primarily contingent on the **questioner** versus the **question**. Our findings indicate that using appreciation emojis can increase a question's response probability, whereas the use of hashtags negatively influences the chances of receiving answers. Our contribution is in providing insights for the design and development of future social question and answering tools, as well as for enhancing social network users' collaboration in supporting social information seeking activities.

1. Introduction

The popularity of online social networking sites (SNS) has increased the use of these social platforms, such as Twitter and Facebook, for social question and answering (Q&A). As defined by Oh, Oh, and Shah (2008, p. 4), social Q&A is the process of "information seeking by asking natural language questions to other users in a network". Social Q&A can surpass web search methods in a number of ways: first, social Q&A reduces the cognitive load on the questioners by allowing them to frame information needs in a more natural way (Jansen, Spink, & Saracevic, 1998). Second, it generates more personalized and interactive answers, specially tailored to the questioner's individual contexts, especially in online communities (Lee & Brusilovsky, 2017). Third, the profile-based user accounts of SNS makes the platform non-anonymous, enabling information seekers to better trust the posted contents (Svensson, 2011). Fourth, compared to the traditional human-based question answering, social Q&A removes the temporal and geographical limits by allowing the questioners and the answerers to exchange information asynchronously online, facilitating many activities, such expert finding (Neshati, Fallahnejad, & Beigy, 2017). Social Q&A also changes the questioner-answerer relationship from one-to-one to one-to-many. In this way, the efficiency of question dissemination can be greatly improved. Social Q&A examples

* Corresponding author.
E-mail addresses: zul112@ist.psu.edu (Z. Liu), jjansen@acm.org (B.J. Jansen).

https://doi.org/10.1016/j.ipm.2017.10.004
Received 22 June 2015; Received in revised form 22 September 2017; Accepted 11 October 2017

from social media sites 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?*

Although social Q&A has a number of advantages relative to traditional information seeking and share methods, it also has the drawback of no guaranteed responses, as the SNS are designed as a platform for information dissemination rather than for Q&A. According to previous studies (Paul, Hong, & Chi, 2011), 18% of questions posted on Twitter ever receive a response, which is only 1/3 of the response rate of community Q&A sites, such as Yahoo!Answers.

To address the low response rate problem, in this research, we design a method to predict if a question posted on SNS will receive a response. We propose 17 influential factors from both the question and questioner's perspectives to build our model. We limit our scope to extrinsic factors only given the existence of literature focusing on factors from both social and cognitive perspectives in knowledge sharing (Chow & Chan, 2008; Lin, Hung, & Chen, 2009). To train and test our discriminative model, we retrieve more than 20,000 real-world questions from Weibo, a Chinese social media platform. We show that our model can effectively predict in advance questions that are unlikely to receive any response with a prediction accuracy of 74%. We also learn from the model factors that have a strong influence on question response rate, and provide interesting insights regarding the future design and develop of social Q&A tools.

We think our contributions are mainly in two folds: first, this research is one of a limited number of studies that address the problem of low response rate. Second, to the best of our knowledge, our work is one of the first studies of social Q&A on the Chinese site, Weibo, the largest Chinese microblogging site at the time of the study. Although focused on Weibo, we also expect our findings to inform behaviors on other social networking mediums as well. Third, we analyze a set of extrinsic factors that are likely to influence the question response rate in social Q&A. By identifying those factors, we could then incorporate them into the design of social Q&A tools or services to facilitate information sharing in variety of domains (c.f., Kim, 2017). For instance, by predicting the number of answers of questions yet to be answered, we could then route questions with low or no predicted response rate to more targeted people on the SNS for possible assistance to increase their response probabilities.

2. Literature review

2.1. Question asking

By making use of all possible online social interactions (Evans & Chi, 2008), social Q&A is consistent with many of the constructs of how people search for information (Jansen & Rieh, 2010), with linkages between social factors and search (Orso, Ruots, Leino, Gamberini, & Jacucci, 2017). Li, Si, Lyu, King, and Chang (2011) revealed that there were about 11% of general tweets containing questions, similar to 13% reported in Efron and Winget (2010), and 6% of tweets having information needs. In term of specific commercial information, Jansen, Zhang, Sobel, and Chowdury (2009) in their work examining Twitter as a mechanism for word-of-mouth advertising reported that 11.1% of the brand-related tweets were information-providing, while 18.1% were information-seeking.

In order to better understand the underlying motivation behind social Q&A, Morris, Teevan, and Panovich (2010) surveyed 624 social network users upon their reasons for choosing social networks as the platform for Q&A. Their results indicated that people search socially primarily due to their trust in friends relative to strangers. Other than that, weak beliefs concerning search engine performance and non-urgent information needs play a factor. To further examine factors influencing users' adoption of social Q&A, in their later work, Morris et al. (2010) conducted another study and confirmed that seeking information on social networks can provide more personalized answers and with higher answer quality. Using a survey-based method, Yang, Morris, Teevan, Adamic, and Ackerman (2011) found that the main reasons for engaging in social Q&A were that information obtained from one's social network was of higher quality and was more personalized as well. They also indicated that culture played a role in people's motivations to query their social networks. Evans, Kairam, and Pirolli (2010) conducted a study comparing the tactics of asking questions on social networking sites versus targeting friends. They found that participants received more responses via SNS but more in-depth answers via private channels. The authors also suggested the possibility of developing more supporting tools for online information seeking by combining different social strategies.

2.2. Question answering

In addition to the literature on question asking, there are also plenty of studies on the performance of question answering. Mamykina, Manoim, Mittal, Hripcsak, and Hartmann (2011) evaluated the answering performance on Stack Overflow and found that the response rate of Stack Overflow (92.6%) exceeded the rates reported for other community Q&A sites, such as Yahoo!Answers (88.2%) (Harper, Moy, & Konstan, 2009) and KiN (66%) (Nam, Ackerman, & Adamic, 2009). However, Paul et al. (2011) noted that the majority of interrogative tweets posted on Twitter received no response, with only 18.7% of questions received answers. Such low response rate in social Q&A was also reported in Brady, Zhong, Morris, and Bigham (2013)'s work. In their study, the authors claimed that of those who posted questions on Facebook, only 34% of them got many or all of their questions answered; and this number for the Twitter users was 33%. Nichols and Kang (2012) further confirmed this finding in their online experiment in which they sent questions to strangers for help. In their results, less than half of the questions received responses from strangers. In view of the unanswered question and the low response rate, a number of other studies were undertaken to analyze both the intrinsic and extrinsic motivations for question answering.

2.2.1. Intrinsic motivations for question answering

Intrinsic motivation is defined as "an innate, rather than derivative, propensity" to engage in an activity (Ryan, Connell, & Grolnick, 1992, p. 169). While investigating motivations behind the South Korean Q&A site, Naver, Nam et al. (2009) noted that the most important reasons for answering questions on Naver were the wish to help others, to learn new things, or participation as a hobby. Oh (2012) in her study investigating the motivations behind answering health-related questions presented ten motivations, including enjoyment, efficacy, learning, personal gain, altruism, community interest, social engagement, empathy, reputation, and reciprocity. She found that altruism was the most significant motivation affecting an individual's willingness to help others, followed by enjoyment and efficacy. Dearman and Truong (2010) surveyed 135 users of Yahoo!Answers to identify their reasons for *not* answering a question that they had read. Respondents revealed that the major reasons for not responding to a question include: the insincerity or meaningless nature of the question, the number of answers already received, the perception of how the questioner would interpret their response, and if the questions required too much effort to answer properly. Yang, Adamic, and Ackerman (2008) also stated that by offering money for a solution, a presented task could attract more views and more participation. In their study exploring why people help others, especially strangers, in electronic networks of practice, Wasko and Faraj (2005) also listed the factor of building a reputation as one of the significant motivation for active participation and knowledge contribution.

2.2.2. Extrinsic motivations for question answering

In contrast to intrinsic motivation, extrinsic motivation is defined as the motivation that comes from outside the individual (Porter & Lawler, 1968). While analyzing unanswered fact-seeking questions on Yahoo!Answers, Shah, Radford, Connaway, Choi, and Kitzie (2012) constructed a typology with four main categories: unclear, complex, inappropriate, and multiple questions to code reasons of failures for these questions using a grounded theory approach. They found that a significant proportion of the failed questions were either too complex or overly broad, indicating the difficulties on crafting coherent questions. Paul et al. (2011) observed that distinct question types lead to different response rates. For instance, they found that some rhetorical questions received a relatively large number of replies as compared to personal and health-related questions. In addition, the response rate was strongly related to some of the question asker's characteristics, such as the size of their networks. Lou, Fang, Lim, and Peng (2013) discovered that the rewarding mechanism used by Baidu Knows had significant positive effect on the quantity of knowledge contribution, although such impact was not found on the quality of answers provided.

Liu and Jansen (2013) studied the social Q&A responses posted on Sina Weibo. They found that the topic of a question could effectively affect its response rate. For instance, they noticed that questions of the topics of *Entertainment, Society, Computer*, etc. received fewer responses as compared to questions from the other categories. Lampe, Gray, Fiore, and Ellison (2014) studied a set of public status Facebook updates and found that mobilization requests got more responses than other kinds of posts. Additionally, they noted that the type of support requested impacts the response speed. Most related to our work, L. Yang et al. (2011) analyzed the not-answered questions on Yahoo!Answers using both content and heuristic-based features. They noted that for both very long and short questions, the more objective questions were more likely to be unanswered. In addition, the authors found that question topic and posting time also had a significant effect on question response rate.

From these previous works, it is apparent that people seek to engage with social Q&A; however, the low response rate for questions posted is a serious drawback. Prior work also has done work on the extrinsic aspects of the questions and their effect on response rate. There has been some work on the intrinsic motivations of those who respond to questions, along with those who do not respond. However, many of these studies, especially the extrinsic motivation work, are focused on survey data. With prior works as the foundation, our research aims to find methods to address the response rate issue by investigating the extrinsic factors that may affect question response probability in social Q&A.

3. Research questions

Given the relatively low response rate in social Q&A, we need a more comprehensive evaluation of what drives individual intentions on helping others in information seeking. Specifically, we aim to answer two research questions:

RQ1: What are the extrinsic factors likely to influence the question response rate in social Q&A?

RQ2: Can we use the extrinsic features to predict whether a question posted on SNS will receive any answer?

4. Methods

4.1. Data collection

To study the response rate in social Q&A, we collected data from China's largest microblogging site, Sina Weibo. Each month, more than two billion statuses are posted on Weibo. Since it's launching 2009, Sina Weibo had attracted over 530 million users by June 2014, which accounted for 83.4% of the total Internet users in China, at the time of the study. Although Sina Weibo essentially adopts similar operating concepts and functions as Twitter, it differs from Twitter in some perspectives. First, Sina Weibo adopted the use of threaded comments, which Twitter used until its re-design in 2013. As can be seen in Fig. 1, comments to the same post on Weibo are shown in a chronological thread. The threaded comments design adds a dimension of interactivity and engagement to microblogging, thus makes it a better source for studies focusing on the behaviors of social response. Second, 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 elaborate questions and answers compared to Twitter users.

Fig. 1. The Layout of Weibo interface highlighting the replying aspect of threaded answers.

Considering the huge percentage of non-information-seeking questions posted on SNS (Li et al., 2011), we adopted a keyword-based method for data collection. Initially, using Weibo Search API, we randomly collected 1000 question posts containing the keyword "?". We used Ansj,[1] a widely-adopted Chinese word segmentation tool, to split each post into unigram, bigram, and trigram words. Two Chinese native speakers were recruited to manually screen the top ranked n-grams for common question cue words. Potential cue words, which appeared more than 50 times in the collected dataset, were added to the keywords pool for the next round of search. This process is continued until no newer keyword is discovered. To extract serious information-seeking questions only, we next removed question particles, (such as "吗", "呢" and "吧" which all have no practical meanings but converts statements into interrogative posts), from the keywords pool. This left us a total of three keywords, including: "anybody can tell" (谁能告诉), "please recommend" (求推荐), and "almighty Weibo" (万能的微博, a common Weibo terminology used to express one's desperate needs for help and information). We collected all available Weibo posts using each of the three keywords. This resulted in 126,071 information-seeking questions in total from 106,367 unique Weibo accounts were collected during a five-month period from May 1 to September 30, 2014. To control the potential effect of the response time on the number of answers received, we intentionally collected the questions and answers up to a week after they were posted. Given the seven-day gap and that 97% of replies to questions happen within an hour (Yang & Counts, 2010; Yang & Wei, 2009), we believed the response period would collect the majority of responses. We also evaluated the response quality for the collected dataset to make sure that at least one response received is related to the question, in other words, not all of the responses are just interacting with the questioner. To do that, we randomly sampled 100 questions with responses from our dataset and manually labeled those responses into informative or interactive. It turns out that about 86% of the questions received at least one informative response.

In preparing our data set for analysis, we noticed that there were many questions with identical content. After a detailed examination of numerous identical questions, we discovered that the duplication was due to two major reasons. First, when people ask broad questions, such as "Anybody can me tell when is the game?", "Please recommend a good movie?", and "Almighty Weibo, what should I do?", they tend to use common words and expressions. A common characteristic of these broadly asked questions is that they tend to be short in length and with nearly identical questions posted by other users. Second, and interestingly, we noticed that the duplication of specific questions can be interpreted as an indication of either spam accounts (which generate posts automatically by copying others) or advertisement campaigns (which contain branding message disguised as non-information-seeking questions). To preserve the quality of the study's data, we decided to delete the later types of duplicated questions. We deleted all duplicated questions with a length greater than 30 characters in our collected dataset based on pure text match after removing all the @ mentions, hashtags, emojis and links. The length was selected based on an analysis of the dataset. We admit that removing duplicated questions in this way may inadvertently exclude information-seeking questions that are not spam or advertising campaigns. However, by doing so we could ensure the validity of the data collected by eliminating the spam and advertisement posting.

To ensure our model's generalizability, we also removed questions from individuals with more than 10,000 followers or less than 20 followers, as we deemed these accounts as outliers. These deletions left us with 62,106 unique questions from 57,699 users with more than 600,000 responses. For each question in our dataset, we collected its content, posting time, the number of comments, and

[1] https://github.com/NLPchina/ansj_seg.

questioner profile via the Sina Weibo API. We also obtained up to 135 latest messages from each questioner to measure their posting styles as well as their social attractiveness and engagement.

4.2. Feature engineering

To predict the response rate of questions posted on Sina Weibo, we introduced features extracted from two different perspectives: the question-based features and the questioner-based features. Next, we introduced each of the two types of features in detail.

4.2.1. Question-based features

The question-based features depict the content and syntactic characteristics of a question, either Weibo-dependent or Weibo-independent, including (1) Informativeness, (2) Attractiveness, (3) Urgency, (4) Politeness, (5) Posting Period, and (6) Topical Category.

Informativeness The informativeness of a question evaluates whether the question contains enough information for the potential respondents to understand what is being asked. In this study, we propose two features to measure the informativeness of a question.

− ***Question Length*** Intuitively, the length of a question reflects how detailed one's information needs are being explained. In that sense, the longer a question is, the better the respondents can understand the questioner's needs and provide assistance. So, we assumed that the question length in both characters and clauses levels are positively correlated with the response rate.

− ***Question Uniqueness*** This feature measures a question's uniqueness as compared to others collected in our dataset. Previous studies suggested that the amount of unique information in a question positive impacted the question's informativeness (Kitzie, Choi, & Shah, 2013). The researchers claimed that the unique information presented may help an answerer interpret the questioner's information needs with a higher level of specificity, thus improving the answer's overall quality. Inspired by Kitzie et al. (2013), we measured the uniqueness of a specific term i in question j using the tf-idf weight, where tf_{ij} is defined as the frequency of term i in question j, normalized by the length of question; and idf_i is the logarithm of the total number of questions collected divided by the number of questions that containing term i. The uniqueness of the question j is then calculated as the average tf-idf value of all terms used within that question.

Attractiveness The attractiveness features assess the eye-catching qualities of a question post. Studies have demonstrated that eye-catching contents, such as posts containing images or URLs, spread over longer time spans than posts with plain text (Sun, 2010; Zhao, Zhu, Qian, & Zhou, 2013). So, we assumed that questions with attractive contents would have a higher probability of receiving answers. Within the context of social Q&A, three variables have emerged as good indices of measuring the attractiveness of a question, specifically: whether a question contains at-mention (e.g., @), hashtag (e.g., #), or emoji (e.g., ☺).

− ***Mention*** The mention feature of Weibo enables users to directly reference others by putting an @ symbol before their screen names. According to Huberman, Romero, and Wu (2009), this feature is widely adopted by Twitter users with about 25.4% of all daily tweets being directed ones. Mention is a strong predictor of information diffusion (Yang & Counts, 2010), as well as a significant factor in enlarging a post's attractiveness and helping initiate responses and conversations (Vega, Parthasarathy, & Torres, 2010). The effect's presence of the Mention was shown in both Comarela, Crovella, Almeida, and Benevenuto (2012) and de Souza, de Magalhães, de Costa, & Fechine's (2012) studies.

− ***Hashtag*** Hashtagging is the way Weibo and Twitter categorize posts according to specific keywords or topics; hashtags' abilities to group conversations and information diffusion has been well studied. Rossi and Magnani (2012) investigated hashtag-based conversations and found that hashtags help break through the social network's restricting structure and make conversations based on non-reciprocal following relationships possible. Based on these findings, we inferred that users can enlarge the visibility of their questions by adopting hashtags, and thus have an increased possibility of getting responses.

− ***Emoticon*** Emojis are graphic representations of facial expressions that Weibo users can embed in their post. Previous literature (Lo, 2008; Walther & D'Addario, 2001) suggested that in compensation for the lack of nonverbal cues, people tend to use more emojis in computer-mediated communications to enable them to better maintain their social presence and to be more engaged in social interactions. Emojis can also be used to draw attention from the recipients (Kriplean, Toomim, Morgan, Borning, & Ko, 2012). We expect a positive correlation between the emoji use and question response rate.

Urgency To measure the urgency of a question, we included three measurements: (a) urgency words, (b) repeated punctuations, and (c) repeated interjections. Urgency words signify the requirement of immediate action or attention (Hellier, Edworthy, Weedon, Walters, & Adams, 2002). Individuals are more likely to respond to urgent warnings created by using urgent words and high signal intensity, as indicated in Baldwin's work (2011). Likewise, repetition regardless of length limitations indicates its importance in communicating one's social meanings to the others (Kalman & Gergle, 2010; Suh, Hong, Pirolli, & Chi, 2010). Repeated punctuations and interjections on SNS are ways questioners used to emphasize their anxieties, whereas emojis are used to attract others' attentions in the absence of verbal communication.

To extract the urgency features, we analyzed all questions collected by comparing them with predefined lists of urgency words (e.g., "urgent" (急), "wait on line" (在线等)), repeated punctuations (e.g., ! !, ??, !?), and 23 Chinese interjections (e.g., "啊", "呀", "哇", "吧", all have no actual meaning). All predefined patterns identified above were based on our observed frequencies of their occurrence based on our analysis of user behavior on Weibo and on the use of urgency observed in prior work (Hellier et al., 2002). We used Ansj to segment words and to remove punctuation. We expected a positive correlation between the usage of urgency expressions and the question response rate.

Politeness As a motivation for pro-social behaviors, prior literature (Bartlett & DeSteno, 2006; Tsang, 2006) has affirmed the significant effect of politeness on social exchange. It is believed that gratitude expressions can enhance the helper's feelings of self-

efficacy and social worth (Bartlett & DeSteno, 2006), and thus it encourages them to engage in pro-social behaviors (Grant & Gino, 2010). We hypothesized that questions with gratitude expressions have higher probabilities of receiving responses. To measure the politeness of a question, we adopted a predefined list containing 14 gratitude (e.g., "thanks", "feel grateful") and reciprocation words (e.g., "pay it back", "return the favor").

Posting Period Previous studies addressed the significant variation in Internet traffic at different times of the day (Beitzel, Jensen, Chowdhury, Grossman, & Frieder, 2004; Jansen, Liu, Weaver, Campbell, & Gregg, 2011). Beitzel et al. (2004) indicated a big change search queries' frequency distribution in a day. Paul et al. (2011) later showed the influence of posting time on response rate, given that tweets posted during peak hours could easily get buried in content streams. This leads us to investigate posting time's effect on question response rate. We divided the posting timestamp of the collected questions into four categories of equal durations, including: *nights (0:01 a.m.–6:00 a.m.), morning (06:01 a.m.–12:00 p.m.), afternoon (12:01 p.m.–6:00 p.m.) and evening (6:01 p.m.–12:00 a.m.)*, as indicated from previous studies (Demirbas, Bayir, Akcora, Yilmaz, & Ferhatosmanoglu, 2010; Wakamiya, Lee, & Sumiya, 2011). As a categorical predictor, posting time period was dummy coded for later analysis.

Topical category Given that expertise is usually context dependent, we assume that the response rate of questions might be also distinct across topical categories. Focusing on the topical categorization of interrogative tweets, prior studies Efron and Winget (2010) and Liu and Jansen (2012) presented significant differences on the number of questions posted across categories. Another study Yang and Wei (2009) showed such topical variance in people's knowledge sharing behavior, indicating that certain categories tend to attract more answers than the others.

To examine this assumption, we employed a categorization method by automatically submitting each of the collected questions to Baidu Zhidao[2] and retrieving the returned classifications. Baidu Zhidao contains 14 main categories, including: *Computer and Internet, Life, Health, Sports, Electronics, Business, Education and Science, Society, Culture and Arts, Game, Entertainment, Personal Vexation, and Region*. Under each main category, there are also a number of sub-categories. The most frequently occurring main category of the top five returned results would be assigned to the question as its topic. For questions containing more than five clauses, we shortened them using the sub sentence containing the question keyword with one clause before and one clause after.

Baidu Zhidao could not categorize 2361 of the 62,106 questions, so we recruited two human annotators to manually label these questions. Additionally, we also asked the annotators to recode 3116 questions with the returned topic of "Region" and "Resource Sharing" into the other 12 categories, given the non-exclusive nature of both categories. From the results of our human coders, 4946 out of 5477 questions (90.30%) received agreement on their topics from the two coders. We merged questions with the topic of *Computer and Internet* and *Electronics* into a new category called *Technology*, given that most of the questions asked under both categories require some technical skills from the potential respondent. We also merged the topical category *Game* into *Entertainment* given its entertainment nature. Finally, we categorized all collected questions into 10 topical categories, including: *Technology, Life, Health, Sports, Business, Education and Science, Society, Culture and Arts, Entertainment, and Personal Vexation*.

4.2.2. Questioner-based features

The questioner-based features describe the author of a question via his/her profile characteristics, as well as past behaviors. Below are the questioner-based features adopted in this study, including (1) Activeness, (2) Posting Style, and (3) Historical Interactions.

Activeness The activeness of a Weibo user can be measured by two means: the number of followers and the number of posts per day.

−***Number of followers*** The number of followers has been investigated as a possible indicator of a user's activeness in spreading information to effective readers (Cha, Haddadi, Benevenuto, & Gummadi, 2010). Given that questions posted on one's SNS can often only be seen by one's followers/friends, we assume that the more followers a questioner has, the more responses he/she could expect to receive. Consistent with our assumption, Paul et al. (2011) concluded that the probability of receiving a response is intrinsically associated with the questioner's number of followers. de Souza et al. (2012) also supported the relationship through their computational investigation.

−***Posting frequency*** Besides a large number of followers, a high posting frequency can also, to some extent, indicate an active SNS user (Reynolds et al., 2010). Thus, it might have a positive impact on the response rate of social Q&A as well; however, based on past literature, posting too much mundane information on one's social network may cause negative physiological consequences to audiences and lead to information fatigue (Gelter, 2010). People may become uninterested in reading the messages posted by followees who disclose too much information (Oulasvirta, Lehtonen, Kurvinen, & Raento, 2010). Both Sibona and Walczak (2011) and Kwak, Lee, Park, and Moon (2010) demonstrated that people "unfriend" those who post too frequently about unimportant topics or mundane personal details. To calculate one's posting frequency, we divided the total number of posts by the total number of days on Weibo from the first day of registration.

Posting Style Every user has his/her own style of posting on SNS. Some prefer to retweet, while others like to engage in dialogues. Wallsten (2008) suggested that users can be grouped into different types based on the way they post on blogs. He showed that blogs were complex platforms that contain a mix of opinion statements, mobilization attempts, requests for audience feedback, and links to information produced by others. Liu and Ingmar (2014) further indicated that the way people tweet reflected their behavior toward ideological friends and foes. Inspired by Wallsten (2008), we measured one's posting style on Weibo from four different dimensions, including (a) their rate of retweet, (b) mention, and (c) URL sharing.

Historical Interactions The features of historical interactions captured the social attractiveness of a questioner. We introduced three metrics for quantifying a questioner's historical interactions on Weibo: (a) the average number of retweets, (b) comments, and (c) likes received.

[2] http://zhidao.baidu.com/.

4.3. Data analysis

For RQ1, we identified the associations between the proposed features and the probability of a new question being answered on Weibo through logistic regression. Before we conducted the regression analysis, multicollinearity was checked by examining the bivariate correlations across all predictor variables. As none of the values in the bivariate correlation matrix exceeded the recommended value of 0.7 (Slinker & Glantz, 1985), our data suggested a lack of multicollinearity among independent variables in the multiple linear regression models. Given the skewed distribution of all our numerical variables, we used a log-transformation to normalize the data. We dummy coded all categorical predictor variables. We employed SPSS for the analysis. The *p*-value was set at .05 to be statistically significant.

In addition to looking into each of the individual features, we also assessed the predictive power of the whole model to answer the second research question. We treated the task of response rate prediction as a classification problem, ie, categorizing questions as with high probability of being answered or with a low probability of being answered. To build the classification model, we experimented a number of different algorithms including: logistic regression, J48 decision tree, support vector machine (SVM) and random forest, as implemented in Weka with default settings and 10-fold cross-validation. We balanced the data set to achieve an equal number of positive and negative instances. For evaluation purpose, we used the traditional metrics, including: precision, recall, F1, and accuracy. Majority induction was adopted as the baseline model to interpret our classification results. With the balanced dataset, we have a baseline accuracy of 0.50.

5. Results

5.1. Descriptive statistics

Overall, the 62,106 information-seeking posts garnered 607,497 responses. At least one answer was provided to 61.68% of the Weibo questions. In general, the response rate of Weibo questions followed a long tail distribution as shown in Fig. 2. We noticed that the response rate of our data set was relatively higher than a previous study (Paul et al., 2011). We thought this might due to the keyword-based method adopted for question collection. While using keywords to retrieve questions, we could exclude many conversational questions with relatively low response rates. Therefore, our identification criteria of a question may be tighter than in the one used in previous work (Paul et al., 2011).

5.2. Data analysis

To answer our first research question, we model the question response rate using a logistic regression implemented in SPSS. The results are summarized in Table 1 with the coefficient, odds ratio, and the *p*-value associated with each feature.

From Table 1, we noticed that all proposed features, taken together, accounted for roughly 40% of the total variance. Given that all features studied in this work are, by design, extrinsic factors that may have an impact on people's social Q&A behavior without the inclusion of any intrinsic determinant, the relatively lower R-square value in our model is reasonable.

In addition to an overall assessment of the prediction model, we also evaluated our proposed features by examining the *p*-values,

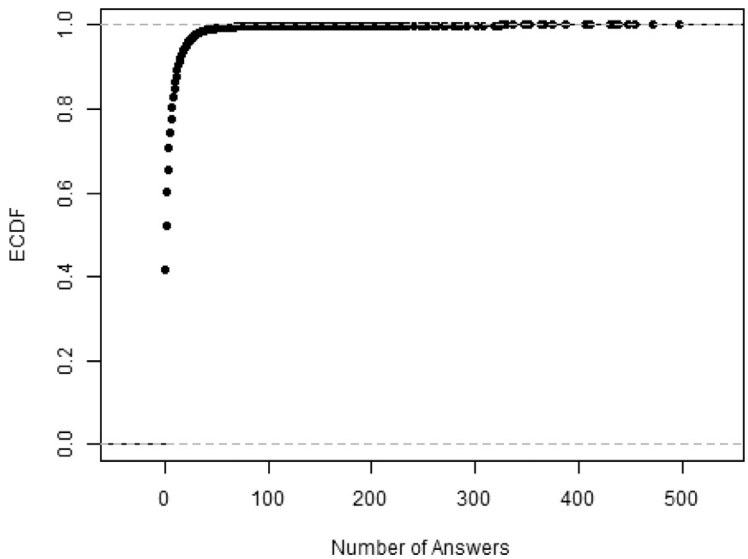

Fig. 2. Empirical cumulative distribution function of question answer numbers on Weibo.

Table 1
Results of logistic regression with coefficient, odds ratio, and *p*-value.

Feature	Coefficient	Odds ratio	*p*-value
Question-based Features			
Informativeness			
Question length	0.07	1.00	.10
Question uniqueness (average tf.idf)	0.30	1.36	.00***
Attractiveness			
Question contains at-mention	0.69	2.00	.00***
Question contains hashtag	−0.37	0.69	.00***
Question contains emoticon	0.13	1.14	.00***
Urgency			
Question contains urgency words	0.26	1.30	.00***
Question contains repeated punctuations	0.00	1.00	.99
Question contains repeated interjections	−0.06	0.95	.36
Politeness	0.09	1/10	.10
Posting Time Period			
0:00–06:00AM - 18:01–23:59PM	−0.06	0.94	.32
06:01–12:00PM - 18:01–23:59PM	0.06	1.06	.08
12:01–18:00PM - 18:01–23:59PM	−0.05	0.95	.02*
Topical Category			
Sports - Entertainment	−0.07	0.93	.38
Healthcare - Entertainment	0.34	1.41	.00***
Business - Entertainment	−0.12	0.89	.15
Education - Entertainment	−0.11	0.90	.02*
Culture - Entertainment	−0.02	0.98	.75
Vexation - Entertainment	−0.24	0.79	.00***
Life - Entertainment	0.16	1.17	.00***
Technology - Entertainment	−0.03	0.98	.41
Society - Entertainment	0.07	1.07	.26
Questioner-based Features			
Activeness			
Number of followers	0.01	1.00	.00***
Posting frequency	0.00	1.00	.43
Posting Style			
Percentage of retweet	1.78	5.92	.00***
Percentage of at-mention	−0.15	0.86	.01**
Percentage of URL sharing	−0.39	0.68	.00***
Historical Interactions			
Percentage of posts received retweet	1.06	2.88	.00***
Percentage of posts received comment	5.64	279.84	.00***
Percentage of posts received like	0.86	2.35	.00***

*** *p*-value < .001,
** *p*-value < .01,
* *p*-value < .05

coefficients and odds ratios associated with each of the predictor variables. We found no statistical evidence to support our hypotheses on posting frequency, question length, and question politeness.

Findings from the multivariate logistic regression provided a general indication of the associations between hypothesized predictors and the question response rate; however, in order to better understand those associations, we conducted more in-depth analyses for each individual feature. Relevant patterns were documented and are discussed in the following section.

5.3. Feature analysis

5.3.1. Question-based features

5.3.1.1. Informativeness

5.3.1.1.1. Question length. In Fig. 3, we plotted the density distributions of the question length for both responded and non-responded questions. We noticed that the length distribution of the responded questions skewed slightly more towards longer questions than non-responded questions, even though the majority of both types of questions contained about 40 characters ($\text{Mean}_{\text{responded}} = 40.59$, $\text{Mean}_{\text{nonresponded}} = 38.80$). This indicated that individuals prefer to answer questions with more specificity.

5.3.1.1.2. Question uniqueness. From the results of the logistic regression, we discovered that questions with more unique words tend to result in higher response probability, reflected by the 0.30 coefficient. We observed that the average tf-idf score for responded questions is 0.23, which is slightly higher than that for non-responded questions (0.21). This observation corresponded with our findings on the question length, indicating that responded questions are more specified and personalized. This may because unique questions may assist the answerers in interpreting the questioner's information needs, thus improving the response's probability.

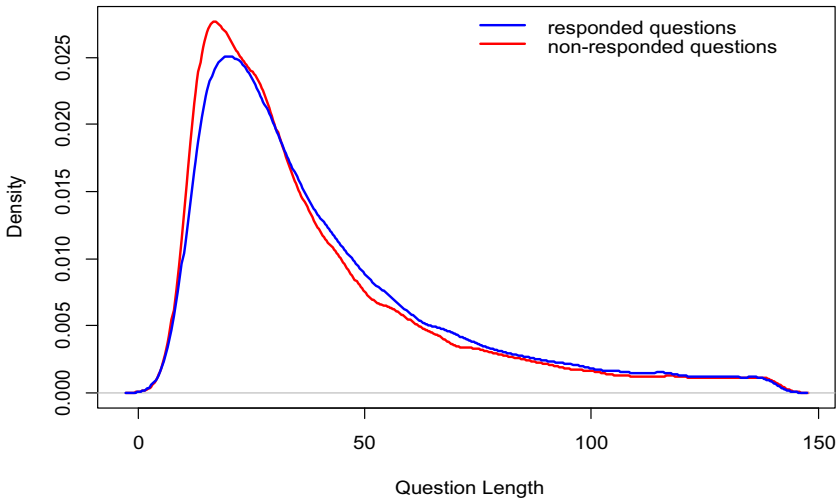

Fig. 3. Density plots for responded and non-responded questions.

5.3.1.2. Attractiveness

5.3.1.2.1. Emojis. While analyzing emojis usage in social Q&A, findings show that among the 24,898 questions with emojis adopted, 16,510 (66.31%) received at least one answer; however, only 58.53% of the questions without the adoption of emojis got responded. In other words, the adoption of emojis did attract more attentions and responses in social Q&A. In Table 2, we listed the top 10 emojis adopted in questions received at least one answer. While grouping the top 10 emojis into three sub-categories according to their underlying intent, namely, the *frustration* emojis, the *appreciation* emojis, and the *other* emojis, we noticed that half of the most adopted emojis were used to express frustration or sadness in not receiving adequate or timely support. In addition, we also found that compared with adopting frustrating emojis, using appreciation emojis significantly increased the response probability in social Q&A, even though such gratitude was not clearly identified at the textual level.

5.3.1.2.2. Mentions. A cross-tabular table was constructed as shown in Table 3. As shown, 76.28% of questions with @ mentions were answered by at least one person, whereas only 59.13% of questions without @ mentions got their reply. This indicated that mentioning a specific user in a question increased the probability of receiving a response.

5.3.1.2.3. Hashtags. To our surprise, results shown in Table 1 indicated that the adoption of hashtag might actually decrease the probability of receiving an answer. While through a more detailed analysis we found that given no set rules for hashtag creation,

Table 2
Top 10 emojis adopted in responded questions for social Q&A on Weibo.

Emoticon	Frequency/percentage (%)	Responded questions (%)
Frustration		
😭	5938/23.85	65.08
😫	2162/8.68	61.74
😂	1374/5.52	64.63
😤	1009/4.05	65.54
😭	808/3.25	65.58
Appreciation		
🙏	1470/5.90	71.89
❤️	997/4.00	77.43
Others		
🤔	2219/8.91	64.80
😳	1775/7.13	70.00
😒	676/2.72	66.41

Table 3
Cross-tabular distribution of @ mention and responded type for social Q&A on Weibo.

	Without @ mention	With @ mention	Total
Non-responded	21,661 (40.87%)	2159 (23.72%)	23,820 (38.35%)
Responded	31,343 (59.13%)	6943 (76.28%)	38,286 (61.65%)
Total	53,004 (100%)	9102 (100%)	62,106 (100%)

95.62% of all the hashtags adopted in our collected questions appeared only once or twice. Among all 5128 questions with hashtag adopted, only 53.27% received an answer; whereas, 62.40% of the questions without hashtag got responded to by others. This indicates the hashtag's negative impact on response rate, which is consistent with previous findings (Liu & Jansen, 2013). Given hashtags' heavy usage in spam posts (Walther & D'Addario, 2001), we surmised that the negative impact of hashtag usage on question response rate might be because using hashtags can annoy readers and lead them to skip those questions.

5.3.1.3. Urgency. Analysis indicated that 70.87% of questions with urgency words received at least one answer. This percentage is much higher than the response probability of questions without urgency words (61.47%). This is consistent with Baldwin's (2011) observations that individuals are more likely to respond to demands created by using urgent words. However, such significant effect was not found when analyzing the adoption of repeated punctuations and interjections, which many associate with urgency.

5.3.1.4. Topical category. As shown in Fig. 4, the number of questions asked by Weibo users varied largely among all 10 topical categories. The most popular topic on Weibo was *Entertainment* (15,166, 24.42%). This is consistent with the entertaining nature of SNS. In contrast, the least popular topic was *Business* (916, 1.48%).

From Fig. 4, we further indicated that although the number of questions asked varied significantly across topical categories, the question response rate did not show huge differences, apart from a few exceptions. We found that questions regarding *Personal Vexation* had the lowest response probability, perhaps due to the ambiguity in user's information-seeking intents. In contrast, questions in the topical categories of both *Healthcare* and *Life* received the highest chances of being answered in our data set with their response probability as 0.69 and 0.68 respectively. Using Chi-square tests, such topical difference was proved to be significant for the question response rate at $p < .01$ ($\chi^2 = 945.89$, $p = .00$).

5.3.1.5. Posting period. We conducted an analysis on the temporal effect on social Q&A behaviors, as prior work has shown a temporal impact to online interactions (Zhang, Jansen, & Spink, 2009). As shown in Fig. 5, we noticed that people asked the most questions between *18:01 p.m. and 23:59 p.m.* (48.17%) and the fewest questions from midnight (3.15%) to the early morning (13.36%), which is consistent with the officially reported pattern of Weibo usage from Sina. Quite different from the distribution of question posting, the response probability cross temporal categories demonstrated completely different distributions with questions posted from *0:00 a.m. to 6:00 a.m.* (66.96%) and *6:01 a.m. to 12:00 p.m.* (67.12%) having the highest probability of being answered.

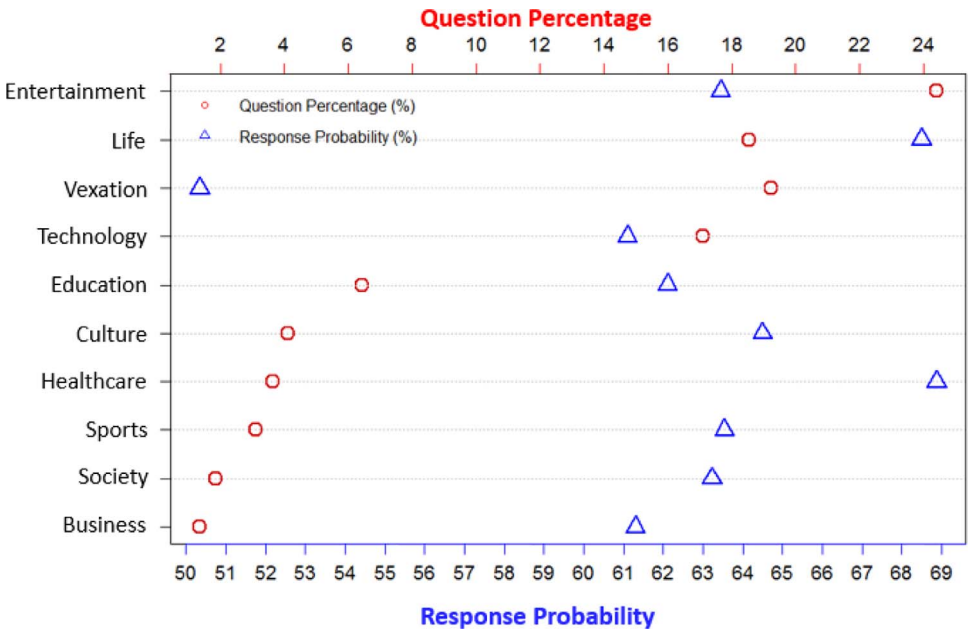

Fig. 4. Distribution of topical categories of Weibo questions and their response probability.

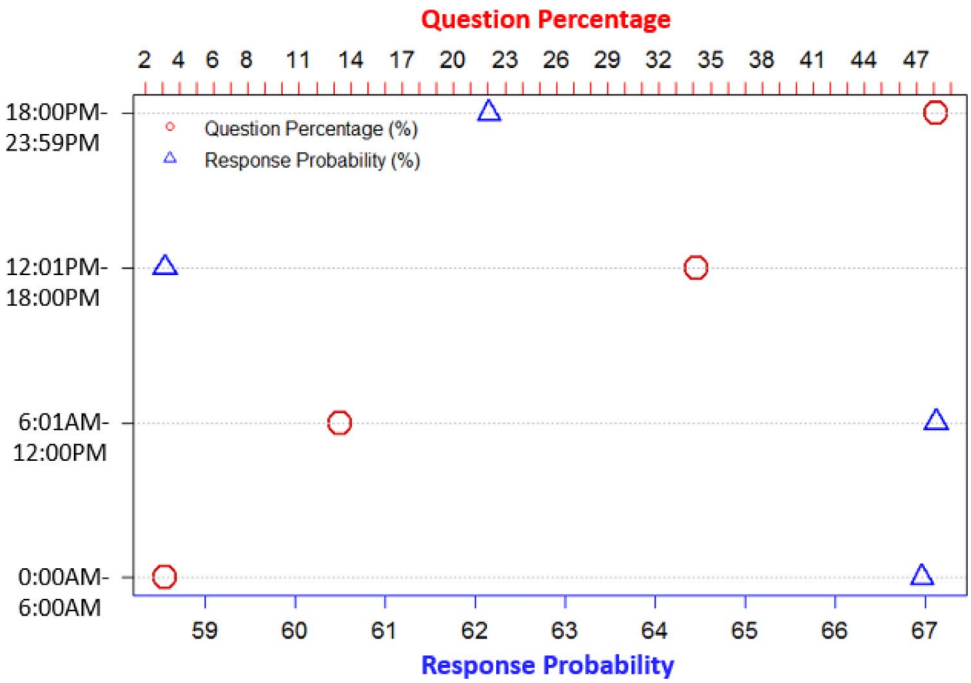

Fig. 5. Distribution of posting time period of Weibo questions and their response probability.

Again, chi-square tests validated the statistically significant difference on question response rate across temporal periods at the level of $p < .01$ ($\chi^2 = 230.525, p = .00$). Such distribution difference suggested a gap between question asking and answering, indicating that although people are more active in information seeking from *12:01 p.m. to 23:59 p.m.*, they actually have lower chances of getting their question answered during this period.

5.3.2. Questioner-based features

5.3.2.1. Activeness

5.3.2.1.1. Number of followers. Fig. 6 shows the cumulative distribution function (CDF) of the number of followers per questioner. Compared with individuals whose questions got no response, successfully responded questioners tend to have more followers. In particular, 80.00% of the non-responded questioners have less than 510 followers, which counted for 54.86% of the responded questioners.

5.3.2.2. Posting style. In Fig. 7, we plot the cumulative distribution function for three distinct posting styles of the questioners: the percentage of retweets, the percentage of mentions, and the percentage of posts containing URLs. We observed that in general questions from users who retweet more have a higher probability of being answered. The average percentage of retweets increased from 48.11% for non-responded questions to 55.96% for responded questions. In contrast, we noticed that individuals with more mentions and URL sharing had less chances of getting their questions answered. This is fairly intuitive since directed communication tends to limit the scope the information disclosure.

5.3.2.3. Historical interactions. According to the odds ratio from Table 1, we found that among all proposed predictors, the features of historical interactions have the most discriminative power on predicting question response probability. Specifically, the variable of percentage of posts received comment showed the most significant positive correlation with the question response possibility. However, the percentage of received likes fail to show a significant correlation with question response probability as the percentage of received comments or retweets.

5.4. Prediction experiment

Table 4 shows the results for the classification task. We observe that except J48, all three other classification methods performed competitively. Logistic regression, SVM, and random forest all achieved a classification accuracy of 0.74, which is much higher than the 0.50 baseline. Among the four classifiers, random forest slightly outperformed the other methods in achieving a higher AUC value.

Table 5 lists the classification results of the proposed random forest model using only features from the user's perspective, features from the question's perspective, as well as all features. In reviewing the findings, we noticed that adopting features from the user's perspective alone achieved a prediction accuracy of 0.73, which is almost the same as the model with all features included. The

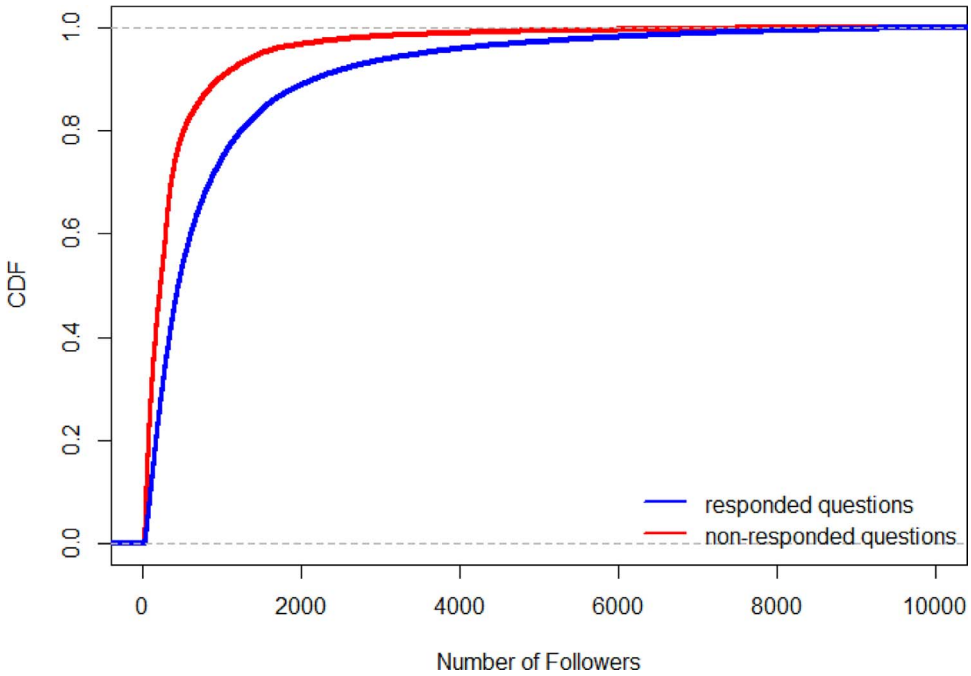

Fig. 6. Cumulative distributions functions of the number of followers for responded and non-responded questions.

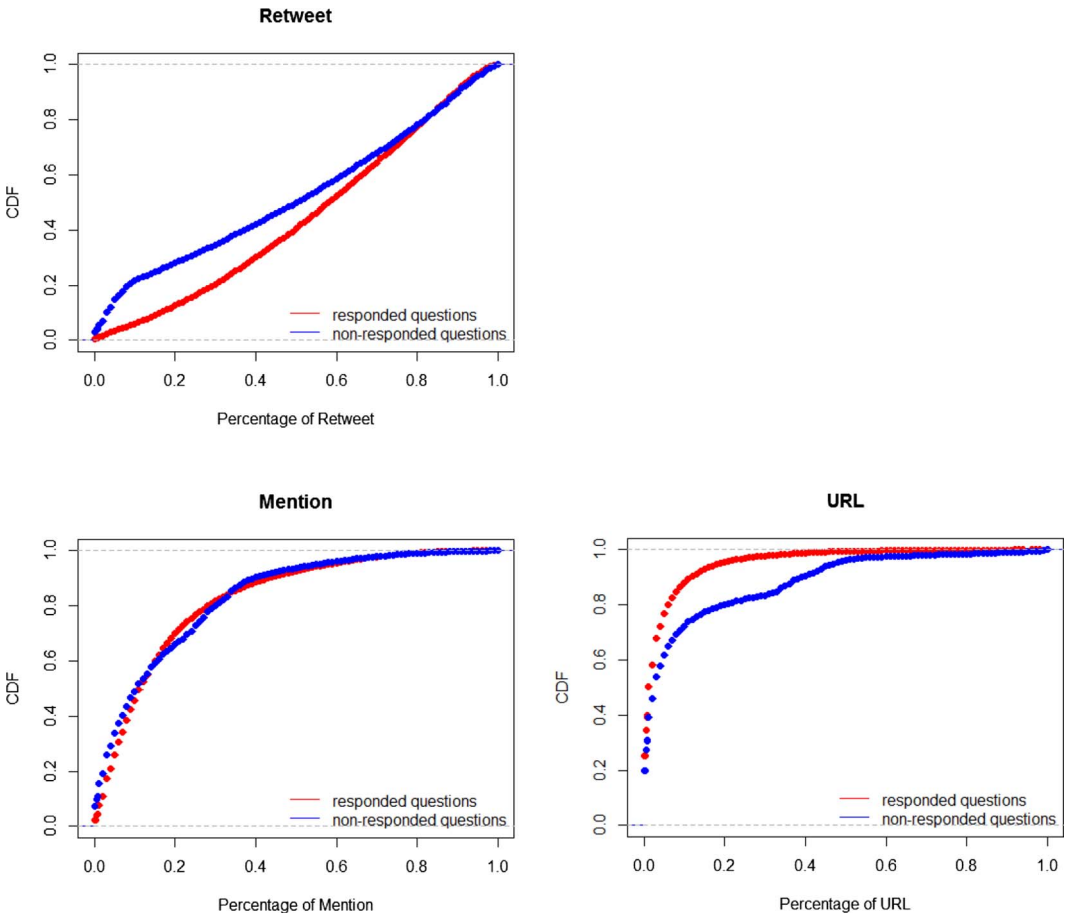

Fig. 7. Cumulative distribution functions of the percentage of retweets, mentions, and URLs for responded and non-responded questions.

Table 4
Classification results by four models using all proposed features.

Features	Precision	Recall	Accuracy	F1	AUC
Logistic regression	0.74	0.74	0.74	0.74	0.82
SVM	0.74	0.74	0.74	0.74	0.74
J48	0.70	0.70	0.69	0.70	0.69
Random forest	0.74	0.74	0.74	0.74	0.83

results indicated that whether a question would be answered on Weibo was related more to the **questioner** than to the **question**. This leads us to conclude the different nature of questions asked on SNS and community Q&A sites with questions posted on SNS required less expertise. In addition, consistent with our analysis results as shown in Table 4, with a further analysis of the predictive power of the features within the user's aspect, we found that features based on user's past social activities contributed the most to the discrimination of the responded and non-responded classes. This finding demonstrated the important role that social activities played in response rate prediction. In other words, more interactive users tended to receive higher probabilities to have their questions answered, regardless of question format, content, or style.

6. Discussion and implication

Although SNS has the potential of becoming a powerful platform for online information seeking, there have been situations in which questions were not being answered. To overcome the problem of the low answer rate, we sought to investigate factors that lead to higher response probabilities in social Q&A. It is noted that many prior studies on answer quantity in community Q&A (e.g., Yahoo!Answers, StackOverflow) were conducted from different perspectives than this research, often focusing on the composition of a question (Asaduzzaman, Mashiyat, Roy, & Schneider, 2013; Choi, Kitzie, & Shah, 2013; Dearman & Truong, 2010). For instance, Choi et al. (2013) reported that textual features, such as clarity, politeness, and specificity, etc., had a significant impact on question response probability on Yahoo!Answers. However, considering the social design and nature of SNS, in this study we also took into account the varied characteristics of different questioners. We found that, unlike in community Q&A, on SNS whether a question would be answered was determined more by the person who asked it than how the question was phrased. Moreover, we further noticed that by only observing the past interactions between the questioner and his/her followers, the response probability of a newly posted question could be very well predicted. These findings highlighted the important role that one's social context (e.g., social presence and social capital) played in social Q&A. More importantly, our results indicated the need of directing questions from those with low social presence and capital to a wider range of audience beyond the barriers of the follow/follower relationship. Even though the same conclusion has been reached by previous studies on predicting replies/retweets, our study provides more detailed analysis on individual's social capital and their network's knowledge sharing behaviors within social Q&A context.

While analyzing the effect of question composition on its response probability, our results demonstrated that questions posted during peak hours actually had lower chances of being answered. This led us to believe that the current timeline design of microblogging sites, such as Twitter and Weibo, was not optimal for the task of Q&A, as questions were mixed up with other non-information seeking posts and can be easily buried among the outdated ones. Possible improvements may include: highlighting those questions with information seeking intents in the timeline, or recommending potential answerers to the questioner and then directly send them the question by at-mentioning practices. The latter suggestion was inspired by our observation that the targeted questions (questions with @mention) received significantly more answers than the general ones. In addition, we also noticed the necessity of separating the interrogative questions from the rhetorical ones (Harper et al., 2009; Li et al., 2011) in social Q&A, as we found that questions within the topical category of *Personal Vexation* had the lowest response probability. One more interesting finding of the present study was that, contrary to our expectation, the adoption of the hashtag in a question actually hindered its chances of being answered. We thought this might because hashtag was useful for mapping against real-world events instead of random questions. Hashtags were originally developed as a tagging method for information filtering and promotion (Huang, Thornton, & Efthimiadis, 2010). It varies considerably in its activity and audience, with some hashtags being used widely while the others were adopted by only a few users (Lin, Margolin, Keegan, Baronchelli, & Lazer, 2013). To overcome the hashtag sparseness issue in social Q&A, the system can resort to hashtag recommendation algorithms and to suggest some commonly used hashtags with information-seeking intents (e.g., #lazyweb, #twoogle, #replytweet) to the questioners.

We believed that our findings are valuable to developers and researchers in the field of social information seeking, ranging from social scientists to social network users. Given the limited studies focusing on the response quantity of social Q&A, our predictive

Table 5
Classification results with the random forest model using questioner-based features, question-based features, and all features.

Features	Precision	Recall	Accuracy	F1	AUC
Question-based features	0.60	0.60	0.60	0.60	0.63
Questioner-based features	0.73	0.72	0.72	0.72	0.80
Question + questioner features	0.74	0.74	0.74	0.74	0.83

model can be adopted as the foundation for future studies with related focus. Social technologists could benefit from our study by understanding the factors that impact whether or not a question will be answered in social context. Taking advantage of our findings and suggestions can help social technologists know the bottlenecks of current systems in the social Q&A environment so that they can develop new tools to support user's questioning and answering behaviors on SNS. When asking questions on SNS, people can utilize the results of our study to match their conditions, checking to see if their question has the potential to receive an ideal number of replies. In summary, our work is of positive value to the research community. Our research aim is to build tools, such as an expert recommender, or a question routing mechanism based on our findings, to make the social Q&A process easy for users to use.

Limitations of our current work include the keyword method of extracting interrogative posts. Although those questions are representative given their high frequency of occurrence, only focusing on this method of extraction might affect the generalizability of work, especially when facing uncommonly asked questions. With only extrinsic features considered, the current model may not achieve as high of a prediction as is possible. So future research will aim at incorporating more intrinsic features, as adding more intrinsic and social predictors into the model will increase its predictability. The investigations on the relationships between social predictors and the response behavior could be another possible direction for later work.

7. Conclusion

With the objective to enhance user interactions and cooperation in a social Q&A context, we performed analysis on the relationships between extrinsic features and question response probability. Using Weibo as the data source and logistic regressions as a statistical method, our analysis results are highlighted as follows.

- We identified in total 17 extrinsic factors, extracted from the both the perspectives of question content and questioner characteristics, that have a significant effect on question response rate in social Q&A.
- We found that the question response probability was related more to who the questioner is (user activeness, posting style, and historical interactions) than how the question is asked (question informativeness, attractiveness, urgency, politeness, posting time, and topic).
- Among all extrinsic factors, we noticed that features based on the questioner's past social activities contributed the most to the discrimination of the responded and non-responded classes, and using it along can very well predict whether or not a new question from a specific user will be answered or not.
- To our surprise, we also observed that the usage of the hashtag in the question and URL in the daily posts had negative influences on the chances of receiving answers.

Through further analysis of those predictors, we noted design problems from two perspectives that may impact the usability of social Q&A. Since there are no official rules for hashtag usage from major SNS providers, hashtags might not be beneficial in the context of social Q&A, given that they do not increase the visibility of the question. Also, the asynchronous nature in social Q&A has been demonstrated while analyzing the response probabilities across question posting time periods.

Based on the present situation, developing corresponding tools and mechanisms, such as interrogative hashtag vocabularies, question recommendation, and intent identifiers could make the user's information seeking process easier and make the potential answerers more willing to share knowledge with others. Based on these possible implications proposed, 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, including the possible linkage with other factors, such as revenue generation (Ortiz-Cordova & Jansen, 2012).

References

Asaduzzaman, M., Mashiyat, A. S., Roy, C. K., & Schneider, K. A. (2013). *Answering questions about unanswered questions of stack overflow. Proceedings of the 10th IEEE Working Conference on Mining Software Repositories (MSR)*. IEEE97–100.

Baldwin, C. L. (2011). Verbal collision avoidance messages during simulated driving: Perceived urgency, alerting effectiveness and annoyance. *Ergonomics, 54*(4), 328–337.

Bartlett, M. Y., & DeSteno, D. (2006). Gratitude and prosocial behavior helping when it costs you. *Psychological Science, 17*(4), 319–325.

Beitzel, S. M., Jensen, E. C., Chowdhury, A., Grossman, D., & Frieder, O. (2004). Hourly analysis of a very large topically categorized web query log. *Paper presented at the 27th annual international ACM SIGIR conference on research and development in information retrieval (SIGIR)*.

Brady, E. L., Zhong, Y., Morris, M. R., & Bigham, J. P. (2013). Investigating the appropriateness of social network question asking as a resource for blind users. *Paper presented at the proceedings of the 2013 conference on computer supported cooperative work*.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, P. K. (2010). Measuring user influence in Twitter: The million follower fallacy. *Paper presented at the 4th international AAAI conference on weblogs and social media(ICWSM)*.

Choi, E., Kitzie, V., & Shah, C. (2013). A machine learning-based approach to predicting success of questions on social question-answering.

Chow, W. S., & Chan, L. S. (2008). Social network, social trust and shared goals in organizational knowledge sharing. *Information & Management, 45*(7), 458–465.

Comarela, G., Crovella, M., Almeida, V., & Benevenuto, F. (2012). Understanding factors that affect response rates in Twitter. *Paper presented at the 23rd ACM conference on hypertext and social media (HT)*.

de Souza, C. C., de Magalhães, J. J., de Costa, E. B., & Fechine, J. M. (2012). Predicting potential responders in Twitter: a query routing algorithm. *Paper presented at the 12th international conference on computational science and its applications (ICCSA) Salvador de Bahia, Brazil*.

Dearman, D., & Truong, K. N. (2010). Why users of yahoo!: Answers do not answer questions. *Paper presented at the proceedings of the SIGCHI conference on human factors in computing systems*.

Demirbas, M., Bayir, M. A., Akcora, C. G., Yilmaz, Y. S., & Ferhatosmanoglu, H. (2010). Crowd-sourced sensing and collaboration using Twitter. *Paper presented at the IEEE international symposium on a world of wireless, mobile and multimedia networks (WoWMoM)*.

Efron, M., & Winget, M. (2010). Questions are content: A taxonomy of questions in a microblogging environment. *Paper presented at the 2010 annual meeting of the*

American society for information science and technology (ASIS&T).
Evans, B. M., & Chi, E. H. (2008). Towards a model of understanding social search. *Paper presented at the 2008 ACM conference on computer supported cooperative work (CSCW)*.
Evans, B. M., Kairam, S., & Pirolli, P. (2010). Do your friends make you smarter?: An analysis of social strategies in online information seeking. *Information Processing & Management, 46*(6), 679–692.
Gelter, H. (2010). Friluftsliv as slow and peak experiences in the transmodern society. *Proceedings from Henrik Ibsen: The Birth of "Friluftsliv". A 150 Year International Dialogue Conference Jubilee Celebration. September 14–19, 2009, Norway*.
Grant, A. M., & Gino, F. (2010). A little thanks goes a long way: Explaining why gratitude expressions motivate prosocial behavior. *Journal of Personality and Social Psychology, 98*(6), 946.
Harper, F. M., Moy, D., & Konstan, J. A. (2009). Facts or friends?: Distinguishing informational and conversational questions in social Q&A sites. *Paper presented at the proceedings of the SIGCHI conference on human factors in computing systems*.
Hellier, E., Edworthy, J., Weedon, B., Walters, K., & Adams, A. (2002). The perceived urgency of speech warnings: Semantics versus acoustics. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 44*(1), 1–17.
Huang, J., Thornton, K. M., & Efthimiadis, E. N. (2010). Conversational tagging in twitter. *Proceedings of the 21st ACM conference on hypertext and hypermedia* (pp. 173–178). ACM.
Huberman, B. A., Romero, D. M., & Wu, F. (2009). Social networks that matter: Twitter under the microscope. *First Monday, 14*(1), 8.
Jansen, B. J., Liu, Z., Weaver, C., Campbell, G., & Gregg, M. (2011). Real time search on the web: Queries, topics, and economic value. *Information Processing & Management, 47*(4), 491–506.
Jansen, B. J., & Rieh, S. Y. (2010). The seventeen theoretical constructs of information searching and information retrieval. *Journal of the American Society for Information Science and Technology, 61*(8), 1517–1534.
Jansen, B. J., Spink, A., & Saracevic, T. (1998). Failure analysis in query construction: Data and analysis from a large sample of web queries. *Paper presented at the 3rd ACM conference on digital libraries*.
Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology, 60*(11), 2169–2188.
Kalman, Y. M., & Gergle, D. (2010). CMC cues enrich lean online communication: The case of letter and punctuation mark repetitions. *Paper presented at the AIS 5th Mediterranean conference on information systems (MCIS)*.
Kim, Y. (2017). Fostering scientists' data sharing behaviors via data repositories, journal supplements, and personal communication methods. *Information Processing & Management, 53*(4), 871–885.
Kitzie, V., Choi, E., & Shah, C. (2013). From bad to good: An investigation of question quality and transformation. *Paper presented at the 2013 annual meeting of the American society for information science and technology (ASIS&T)*.
Kriplean, T., Toomim, M., Morgan, J., Borning, A., & Ko, A. (2012). Is this what you meant?: Promoting listening on the web with reflect. *Paper presented at the 30th ACM SIGCHI conference on human factors in computing systems (SIGCHI)*.
Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Paper presented at the 19th international conference on World Wide Web (WWW)*.
Lampe, C., Gray, R., Fiore, A. T., & Ellison, N. (2014). Help is on the way: Patterns of responses to resource requests on Facebook. *Paper presented at the 17th ACM conference on computer supported cooperative work & social computing (CSCW)*.
Lee, D. H., & Brusilovsky, P. (2017). Improving personalized recommendations using community membership information. *Information Processing & Management, 53*(5), 1201–1214.
Li, B., Si, X., Lyu, M. R., King, I., & Chang, E. Y. (2011). Question identification on Twitter. *Paper presented at the 20th ACM international conference on information and knowledge management*.
Lin, M.-J. J., Hung, S.-W., & Chen, C.-J. (2009). Fostering the determinants of knowledge sharing in professional virtual communities. *Computers in Human Behavior, 25*(4), 929–939.
Lin, Y.-R., Margolin, B., Keegan, A., Baronchelli, A., & Lazer, D. (2013). # Bigbirds never die: Understanding social dynamics of emergent hashtag. *Proceedings of the 7th international AAAI conference on weblogs and social media*. Ann Arbor, MI, USA: AAAI.
Liu, Z., & Jansen, B. J. (2012). Almighty Twitter, what are people asking for? *Paper presented at the 2012 annual meeting of the American society for information science and technology (ASIS&T)*.
Liu, Z., & Jansen, B. J. (2013). Factors influencing the response rate in social question and answering behavior. *Paper presented at the 2013 ACM conference on computer supported cooperative work (CSCW)*.
Liu, Z., & Weber, I. (2014). Predicting ideological friends and foes in Twitter conflicts. *Paper presented at the 23rd international conference on World Wide Web companion*.
Lo, S.-K. (2008). The nonverbal communication functions of emoticons in computer-mediated communication. *Cyber Psychology & Behavior, 11*(5), 595–597.
Lou, J., Fang, Y., Lim, K. H., & Peng, J. Z. (2013). Contributing high quantity and quality knowledge to online Q&A communities. *Journal of the American Society for Information Science and Technology, 64*(2), 356–371.
Mamykina, L., Manoim, B., Mittal, M., Hripcsak, G., & Hartmann, B. (2011). Design lessons from the fastest Q&A site in the west *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 2857–2866): ACM.
Morris, M. R., Teevan, J., & Panovich, K. (2010). What do people ask their social networks, and why?: A survey study of status message Q&A behavior. *Paper presented at the 28th ACM conference on human factors in computing systems (SIGCHI)*.
Nam, K. K., Ackerman, M. S., & Adamic, L. A. (2009). Questions in, knowledge in?: A study of naver's question answering community. *Paper presented at the proceedings of the SIGCHI conference on human factors in computing systems*.
Neshati, M., Fallahnejad, Z., & Beigy, H. (2017). On dynamicity of expert finding in community question answering. *Information Processing & Management, 53*(5), 1026–1042.
Nichols, J., & Kang, J.-H. (2012). Asking questions of targeted strangers on social networks. *Paper presented at the ACM 2012 conference on computer supported cooperative work (CSCW)*.
Oh, S. (2012). The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments. *Journal of the American Society for Information Science and Technology, 63*(3), 543–557.
Oh, S., Oh, J. S., & Shah, C. (2008). The use of information sources by Internet users in answering questions. *Paper presented at the 2008 annual meeting of the American society for information science and technology (ASIS&T)*.
Orso, V., Ruots, T., Leino, J., Gamberini, L., & Jacucci, G. (2017). Overlaying social information: The effects on users' search and information-selection behavior. *Information Processing & Management, 53*(6), 1269–1286.
Ortiz-Cordova, A., & Jansen, B. J. (2012). Classifying web search queries in order to identify high revenue generating customers. *Journal of the American Society for Information Science and Technology, 63*(7), 1426–1441.
Oulasvirta, A., Lehtonen, E., Kurvinen, E., & Raento, M. (2010). Making the ordinary visible in microblogs. *Personal and Ubiquitous Computing, 14*(3), 237–249.
Paul, S. A., Hong, L., & Chi, E. H. (2011). Is Twitter a good place for asking questions? A characterization study. *Paper presented at the 5th international AAAI conference on weblogs and social media (ICWSM)*.
Porter, L. W., & Lawler, E. E. (1968). *Managerial attitudes and performance*.
Reynolds, W. N., Weber, M. S., Farber, R. M., Corley, C., Cowell, A. J., & Gregory, M. (2010). Social media and social reality. *Paper presented at the IEEE international conference on intelligence and security informatics (ISI)*.
Rossi, L., & Magnani, M. (2012). Conversation practices and network structure in Twitter. *Paper presented at the 6th international AAAI conference on web and social media (ICWSM)*.
Ryan, R. M., Connell, J. P., & Grolnick, W. S. (1992). When achievement is not intrinsically motivated: A theory of internalization and self-regulation in school. *167*,

(pp. 167–188).
Shah, C., Radford, M. L., Connaway, L. S., Choi, E., & Kitzie, V. (2012). "How much change do you get from 40$?"–Analyzing and addressing failed questions on social Q&A. *Proceedings of the American Society for Information Science and Technology, 49*(1), 1–10.
Sibona, C., & Walczak, S. (2011). Unfriending on Facebook: Friend request and online/offline behavior analysis. *Paper presented at the 44th Hawaii international conference on system sciences (HICSS)*.
Slinker, B., & Glantz, S. (1985). Multiple regression for physiological data analysis: The problem of multicollinearity. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 249*(1), R1–R12.
Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010). Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. *Paper presented at the 2nd international conference on social computing (SocialCom)*.
Sun, Y.-C. (2010). Developing reflective cyber communities in the blogosphere: A case study in Taiwan higher education. *Teaching in Higher education, 15*(4), 369–381.
Svensson, A. (2011). *Facebook – The social newspaper that never sleeps - A study of Facebook eWOM's persuasiveness on the receivers* Master Degree Thesis 2011:187.
Tsang, J.-A. (2006). Gratitude and prosocial behaviour: An experimental test of gratitude. *Cognition & Emotion, 20*(1), 138–148.
Vega, E., Parthasarathy, R., & Torres, J. (2010). Where are my tweeps?: Twitter usage at conferences. *Paper, Personal Information*, 1–6.
Wakamiya, S., Lee, R., & Sumiya, K. (2011). Crowd-based urban characterization: Extracting crowd behavioral patterns in urban areas from Twitter. *Paper presented at the 3rd ACM SIGSPATIAL international workshop on location-based social networks*.
Wallsten, K. (2008). Political blogs: Transmission belts, soapboxes, mobilizers, or conversation starters? *Journal of Information Technology & Politics, 4*(3), 19–40.
Walther, J. B., & D'Addario, K. P. (2001). The impacts of emoticons on message interpretation in computer-mediated communication. *Social Science Computer Review, 19*(3), 324–347.
Wasko, M. M., & Faraj, S. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly, 29*(1), 35–57 Special Issue on Information Technologies and Knowledge Management.
Yang, J., Adamic, L. A., & Ackerman, M. S. (2008). Crowdsourcing and knowledge sharing: Strategic user behavior on taskcn. *Paper presented at the proceedings of the 9th ACM conference on electronic commerce*.
Yang, J., & Counts, S. (2010). Predicting the speed, scale, and range of information diffusion in Twitter. *Paper presented at the 4th international AAAI conference on weblogs and social media (ICWSM)*.
Yang, J., Morris, M. R., Teevan, J., Adamic, L. A., & Ackerman, M. S. (2011). Culture matters: A survey study of social Q&A behavior. *Paper presented at the 5th international AAAI conference on weblogs and social media (ICWSM)*.
Yang, J., & Wei, X. (2009). Seeking and offering expertise across categories: A sustainable mechanism works for Baidu Knows. *Paper presented at the 3rd international AAAI conference on weblogs and social media (ICWSM)*.
Yang, L., Bao, S., Lin, Q., Wu, X., Han, D., Su, Z., et al. (2011). Analyzing and predicting not-answered questions in community-based question answering services. *AAAI*.
Zhang, Y., Jansen, B. J., & Spink, A. (2009). Time series analysis of a Web search engine transaction log. *Information Processing & Management, 45*(2), 230–245.
Zhao, X., Zhu, F., Qian, W., & Zhou, A. (2013). *Impact of multimedia in sina weibo: Popularity and life span. Semantic Web and Web Science*. Springer55–65.