

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

Information Processing and Management

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

Identifying and predicting the desire to help in social question and answering

Zhe Liu^{a,*}, Bernard J. Jansen^b^a IBM Almaden Research Center, San Jose, CA, USA^b Social Computing Group, Qatar Computing Research Institute, Doha, Qatar

ARTICLE INFO

Article history:

Received 3 September 2015

Revised 12 April 2016

Accepted 2 May 2016

Available online xxx

Keywords:

Social Q&A

Question routing

Online information seeking

Social networks

Twitter

Predictive model

ABSTRACT

The increasing volume of questions posted on social question and answering sites has triggered the development of question routing services. Most of these routing algorithms are able to recognize effectively individuals with the required knowledge to answer a specific question. However, just because people have the capability to answer a question, does not mean that they have the desire to help. In this research, we evaluate the practical performance of the question routing services in social context by analyzing the knowledge sharing behavior of users in social Q&A process in terms of their participation, interests, and connectedness. We collect questions and answers over a ten-month period from Wenwo, a major Chinese question routing service. Using 340,658 questions and 1,754,280 replies, findings reveal separate roles for knowledge sharers and consumers. Based on this finding, we identify knowledge sharers from non-sharers a priori in order to increase the response probabilities. We evaluate our model based on an analysis of 3006 Wenwo knowledge sharers and non-sharers. Our experimental results demonstrate knowledge sharer prediction based solely on non-Q&A features achieves a 70% success rate in accurately identifying willing respondents.

© 2016 Published by Elsevier Ltd.

1. Introduction

Social networking sites have been widely adopted for online communication. Besides using these sites for relationship formation and maintenance (Zhang, Jansen, & Chowdhury, 2011), many people also rely on social networking sites (SNS) for seeking information (Jansen, Sobel, & Cook, 2011; Morris, Teevan, & Panovich, 2010). Although not intentionally designed for questioning and answering, people enjoy expressing their information needs in natural language questions on SNS (Liu & Jansen, 2012; Paul, Hong, & Chi, 2011; Shah, Oh, & Oh, 2008), a behavior referred to as social questioning and answering (social Q&A). By making use of social interactions online (Evans & Chi, 2008), social Q&A techniques provide individuals with simpler and more personalized search experiences over conventional information retrieval methods. Due to such advantages, social Q&A sites have attracted researchers' attention and has motivated the creation of models and tools to facilitate the social information seeking process (Jansen, Zhang, Sobel, & Chowdhury, 2009).

Among the proposed methods are several question routing algorithms that mostly involve expert finding techniques to solve the problem of nonguaranteed responses in a social context. Many studies claimed that by routing questions to a larger audience base can effectively reduce the number of unanswered inquiries. However, we argue that even after finding people with required knowledge, we still do not know if they have the desire to help. In other words, we believe that it is

* Corresponding author.

E-mail addresses: liuzh@us.ibm.com, lztoby@gmail.com (Z. Liu), jjansen@acm.org (B.J. Jansen).

very important to validate the feasibility of the question routing mechanism by measuring its practical effectiveness in real world circumstances. In addition, it would be also beneficial to develop methods to determine if an individual is willing to share his/her knowledge via responding to other's questions.

To address these issues, we analyze questions and answers posted during a 10-month period on Wenwo, a Chinese question routing service based on microblogging sites. We evaluate the real world performance of Wenwo from three perspectives, which are user's (a) participation (b) interest, and (c) connectedness. These measurements are chosen intentionally as they have been adopted as means to explore the patterns of user engagement in question answering communities (Adamic, Zhang, Bakshy, & Ackerman, 2008; Gyongyi, Koutrika, Pedersen, & Garcia-Molina, 2008; Shah, Oh, & Oh, 2009).

Next, we develop a predictive user engagement model based on a number of non-Q&A features, including: user profile, posting behavior, language style, and social activities. We did not rely on any Q&A related feature in this study in order to avoid the cold-start problem. Through our model, we find that less popular but more interactive individuals are more willing to respond to others in social Q&A environments. In addition, we also note that the psycho-linguistic characteristics of an individual's microblogging posts, such as their usage of verbs, pronouns, and cognitive expressions, are also indicative of their roles in the social Q&A sites. This research is beneficial because it provides a more in-depth understanding of the social Q&A process, especially the characteristics of people who are willing to engage in knowledge sharing. The findings can also be viewed as design guidelines for future question routing systems based on both capability and desire of the potential respondents.

In the next section, prior studies on social Q&A related to this work are presented. Research questions are proposed in Section 3 and followed by our data collection process. We examine the knowledge sharing behaviors and patterns among strangers in social Q&A in Section 5. Based on our results, we build a classifier of potential knowledge sharer in Section 6. We conclude in Section 7 with some future work.

2. Literature review

2.1. Social Q&A

Defined by Morris et al. (2010), social Q&A is the process of discovering information online with the assistance of social resources. It lies between the boundaries of technical and human-powered information seeking models. The Social Q&A technique outperforms the traditional information-seeking methods (e.g. search engine and online databases, etc.) for both more personalized search experience and results. Studies investigating motivations for participation in the social Q&A process suggested that people primarily search socially due to their trust in friends over strangers (Liu & Jansen, 2012; Morris et al., 2010; Yang, Morris, Teevan, Adamic, & Ackerman, 2011), weak beliefs on search engine performances (Morris et al., 2010), as well as non-urgent information needs (Teevan, Collins-Thompson, White, Dumais, & Kim, 2013).

2.2. User engagement in knowledge sharing

Although the perception of an individual's information needs in the social Q&A process is critical, it is equally important to measure the level of user engagement in knowledge sharing, since social Q&A services thrive on users' active involvement (Shah et al., 2009). Studies measured individual's knowledge sharing on Q&A sites mainly from three different perspectives: user engagement, user interest, and user connectedness. In terms of user engagement, Nam, Ackerman, and Adamic (2009) analyzed the Knowledge-IN website and found a significant separation between asker and answerer roles, with very little within-category reciprocity. While analyzing user engagement in both Yahoo!Answers and Google Answers, Shah et al. (2008) found that the majority of the population in Yahoo!Answers participated in both posting questions and answers, whereas in Google Answers, there were many one-time consumers and a small number of contributors. Gazan (2007) divided questioners into two types according to their involvement in follow-up discussions on Answerbag: Seekers and Sloths. Seekers tended to interact with others about their questions, while Sloths post their question and interact no further. Paul et al. (2011) noted that the majority of questions posted on Twitter received no response. When sending questions to strangers for help, Nichols and Kang (2012) indicated that less than half of the questions received answer.

Concerning user's topical interests in the process of Q&A, Liu and Jansen (2013) studied the questions posted on Sina Weibo, the largest Chinese microblogging site. They noticed a relatively higher response rate for questions posted on Sina Weibo than questions posted on Twitter. They also found that the question's topic could effectively affect its response rate. For instance, they noticed that questions on the topics of Entertainment, Society, Computer, etc., received fewer responses as compared to questions from the other categories. Lampe, Gray, Fiore, and Ellison (2014) analyzed a set of public status updates posted to Facebook and stated that mobilization requests got more responses than other kinds of posts. Adamic et al. (2008) analyzed a set of question-answer pairs from Yahoo!Answers and found different distributions of questioner/answerer overlap across topical categories.

Lastly, from the perspective of measuring user connectedness in the process of Q&A, Gyongyi et al. (2008) analyzed the interactions between individuals on Yahoo!Answers by analyzing undirected bipartite graphs generated by asker and answers under each topical categories. Their results showed that the vast majority of users were connected within a single large community on Yahoo!Answers. Zhang, Ackerman, and Adamic (2007) did the same kind of study by using a bow-tie structure analysis. They found that more than half of the users on Yahoo!Answers usually only ask questions without

answering. [Adamic et al. \(2008\)](#) performed ego network analysis on a set of question-answer pairs from Yahoo!Answers and found “discussion users” and “answer users” within different Q&A communities.

Besides the above-mentioned literature measuring user engagement, there are also a stream of studies that take cognizance to issues germane to enhancing user engagement. [Chua and Banerjee \(2015a\)](#) have developed a framework to explain why some questions were responded to while the others do not. Their framework suggested that the answerability of questions depends on asker meta data (e.g. the asker’s popularity, participation, asking time), as well as the way the questions were phrased (e.g. the level of details, specify, clarity). [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. [L. Yang et al. \(2011\)](#) analyzed the unanswered questions on Yahoo!Answers using both content and heuristic-based features. They declared that questions to be either too long or too short were less likely to be answered. While studying the intrinsic motivations for user engagement in the health Q&A process, [Oh \(2011\)](#) listed ten stimuli, including: enjoyment, efficacy, learning, personal gain, altruism, community interest, social engagement, empathy, reputation, and reciprocity. She found that altruism had the most significant effect on individual’s willingness to help others, followed by enjoyment and efficacy.

2.3. Question routing in social Q&A

The concept of question routing refers to directing newly posted questions to potential answerers. According to the previous studies ([Pal, Chang, & Konstan, 2012](#); [Paul et al., 2011](#); [Zhou, Lyu, & King, 2012](#)), the appropriateness of potential answerers was mainly measured based on their expertise as demonstrated in their previously answered questions. Numerous algorithms have been proposed to address the problem of question routing on community Q&A portals, which are sites specifically designed for asking questions. [Chen and Nayak \(2011\)](#) incorporated question category into their question routing model to sift out irrelevant questions according to the profile of an answerer. [Zhou et al. \(2012\)](#) considered the problem of question routing as a classification task, developing a number of features that capture different aspects of questions, users, and their relations. [Guo, Xu, Bao, and Yu \(2008\)](#) recommended questions to potential answerers by discovering latent topics and interests. [Luo, Wang, Zhou, Pan, and Chen \(2014\)](#) built a system that can route questions to individuals based on their ability, willingness, and readiness.

Different from the above-mentioned studies, [Pan, Luo, Chi, and Liao \(2013\)](#) identified potential answerers by only leveraging users’ social activities. Similarly, [Srba, Grznar, and Bielikova \(2015\)](#) proposed a question routing method to analyze user’s non-Q&A traces extracted from many different sources. Both studies proved that using only non-Q&A features could be effective in predicting one’s desire and ability to help, as well as avoiding cold-start problems.

Even though the reviewed studies claimed that question routing service could effectively reduce the number of non-responded questions, few of them actually examine its real-world performance. Through online experiments, [Nichols and Kang \(2012\)](#) explored the feasibility of users responding to questions sent by strangers. They found that fewer than half of the people answered questions posted by strangers. However, their work failed to indicate the characteristics of those responders. [Mahmud, Zhou, Megiddo, Nichols, and Drews \(2013\)](#) further confirmed the absence of knowledge sharing in social Q&A process, indicating less than 5% of questions sent to strangers were answered. Given the low response rate reported in study of [Paul et al. \(2011\)](#), we believe that more comprehensive evaluation of factors affecting an individual’s desire to contribute knowledge on social Q&A site, will assist in achieving SNS’s power in social information seeking. Considering the lack of study and the importance of identifying potential responders in terms of their willingness to respond in the social Q&A process, we believe that understanding what kinds of users are more willing to respond to questions on SNS.

3. Research objectives

Inspired by the above-mentioned literature, we propose two research goals in this study. One measures the practical effectiveness of question routing service in social Q&A process, and the other identifies those with the desire to help from all the potential answerers. To be more specific, our research objectives are as follows:

RO1: *Measuring the practical performance of question routing service in social Q&A context.*

For this research objective, we seek to examine to what extent the question routing services can contribute to the social Q&A process. We are particularly interested in evaluating the impact of question routing services on user engagement, which according to previous literature ([Adamic et al., 2008](#); [Gyongyi et al., 2008](#); [Shah et al., 2009](#)), can be further characterized as individuals’ engagement, interests, and their social connectedness with other users in the Q&A process.

RO2: *Detecting individuals with the desire to engage in knowledge sharing by using their non-Q&A characteristics.*

Our second research objective explores the distinction between active knowledge sharers and non-sharers in social Q&A process. By active knowledge sharers, we refer to individuals with high willingness to answer other’s questions. To measure the differences, we introduced non-Q&A features from four different aspects, including: SNS profile, posting behaviors, language style, and social activities. We choose features from those perspectives given their power in characterizing users on SNS ([Chen, Hsieh, Mahmud, & Nichols, 2014](#); [Lee, Mahmud, Chen, Zhou, & Nichols, 2014](#); [Liu, Kliman-Silver, & Mislove, 2014](#); [Liu & Jansen, 2013](#); [Pennacchiotti & Popescu, 2011](#)). Q&A-related attributes such as the difficulty of the question asked, and the knowledge and expertise level of the answerer are not considered in this study. In achieving the second research objective, we are able to contribute to the design of future question routing services or tools, which can better direct questions toward individuals who are likely to provide answers.

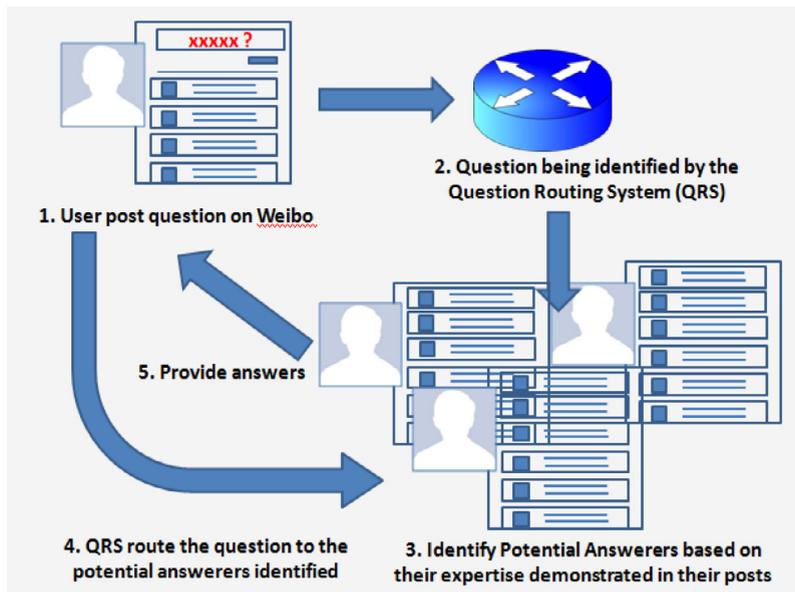


Fig. 1. Illustration of the question routing process of Wenwo.

4. Data collection

To measure the effectiveness of real world question routing services, we collected data from the social Q&A platform, Wenwo (微问). Wenwo is a third party question routing application of Sina Weibo, which is China's largest microblogging site. Sina Weibo had more than 600 million registered accounts by September 2014 (Bai, 2015), accounting for 93.60% of the total Internet users in China. Each month, over 2 billion statuses were posted on Sina Weibo. At the time of the study, Weibo essentially adopted the operating concept and provided very similar functions as Twitter.

Wenwo operates in a different manner compared to those traditional community Q&A portals, such as Yahoo!Answers and Baidu Knows. In those traditional community Q&A sites, people ask questions and then passively wait for the potential helpers to see the questions and to respond. In contrast, in Wenwo, individuals can either post questions directly to the site, or they can post their questions on Sina Weibo by mentioning @微问 (@Wenwo). After receiving the questions, Wenwo identifies a number of potential respondents based on their expertise and experience as demonstrated on their Weibo profiles, using machine-learning techniques. By routing questions to those "qualified" respondents, Wenwo effectively attempts to increase the probability of obtaining high quality responses. A graphical demonstration of the question routing procedure of Wenwo is shown in Fig. 1.

Another difference between Wenwo and many other community Q&A sites is that, in addition to presenting the answers received, in most cases, Wenwo also identifies up to 10 individuals to whom the question has been routed but who have not answered. This feature allowed us to distinguish who responded to a question and who did not, with which we are able to build a classifier with both positive and negative instances, and this is also the reason why we choose Wenwo as the data source to study knowledge sharer and non-sharer behaviors. Fig. 2 is a screenshot of Wenwo with major sections highlighted.

Since Wenwo limits the number of questions that one can view to only popular questions or questions routed to him / her, we decomposed our data collection process into two steps: (a) identify the questions and (b) automatically collect the identified questions along with their answers using a web-based crawler. To identify the questions asked or extracted by Wenwo, we adopted the method of searching Sina Weibo with the key phrase "I just posted a question on [Wenwo]" (我刚刚在【微问】提了一个问题). We selected this key phrase because once someone successfully posts a question to or on Wenwo, as a marketing strategy, the service will generate an automatic post to the asker's Sina Weibo timeline, using the templated phrase "I just posted a question on [Wenwo]". Using this key phrase and the Sina Weibo API, we collected 340,658 questions posted during a ten-month period from January 24, 2013 to October 18, 2013, along with the URLs linking to their Wenwo pages. Then, with a web-based crawler, we collected the question category, posting time, as well as all respondents and non-respondents for each identified questions. In total, we collected 1,754,280 answered and 585,359 unanswered records. All questions and answers are maintained in Chinese for later analysis.



Fig. 2. Screenshot of Wenwo with major sections highlight.

5. Results evaluating the performance of Wenwo

5.1. User engagement in knowledge sharing

Based on an initial examination of the dataset, we notice that 339,878 out of the 340,658 questions in our collection received at least one answer, yielding a response rate of 99.77%. On average, each question received 5.14 answers. Compared to the relatively low number of answers received in many social Q&A settings, the dataset reveals the effectiveness of the question routing mechanism in real world circumstances.

In addition, we further analyze the roles that individuals played in Wenwo. In total, 715,907 unique Weibo users participated in the social Q&A process. Among them, 22,203 (3.10%) individuals both asked and answered questions, while 221,060 (30.88%) asked at least one question but provided no answer. In contrast, 472,644 (66.02%) users posted no question but replied at least once.

Fig. 3 shows the plotted distributions of the number of questions asked versus the number of answers provided. Surprisingly, we notice that there were more knowledge sharers (users who posted more answers than questions) than knowledge consumers (users who posted more questions than answers) on Wenwo. While comparing our results on knowledge contributors versus knowledge consumers with the findings presented in Shah et al. (2008)'s and Gyongyi et al. (2008)'s observations based on Yahoo!Answers, we note again the power of active question routing in social Q&A context. Further analysis on the distribution of answered questions revealed an uneven participation in Wenwo, where a small number of individuals contributed to a large proportion of questions and a large proportion of users only answered a few number of questions.

5.2. User interest in knowledge sharing

Next, we explore the topical distribution of answers received on Wenwo. To be more specific, we look at the contributions of each specific user in our dataset. We find that the average number of topics in which a user posted a question is 1.07, which is lower than the average number of topics a user responded (1.33). Among the results, 85.23% of the knowledge sharers answered questions under only one single topical category.

To investigate the distribution of knowledge sharers across topical categories, in Fig. 4, we plot the distribution of topics answered as a function of the total number of questions responded to by each individual user in our dataset. From the plot, we see that the number of topical categories answered by each individual increased along with the amount of questions to which they have responded. While analyzing the topical distribution of the top 1000 knowledge sharers, we find that on average, they answered 503.13 questions within 12.22 topics. The Spearman correlation coefficient between the number of topics answered and the total number of answers provided per user is 0.71, indicating a strong association between individual's responsiveness and diversity of interest.

To reveal the association between topical categories, we measure the extent to which individuals who answer questions in one category are also likely to do so in another. We adopt the Normalized Pointwise Mutual Information (NPMI) method

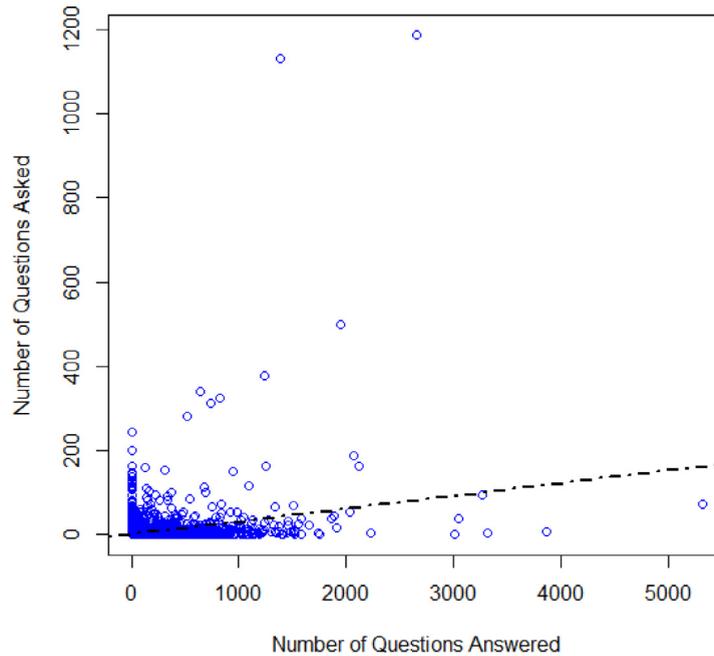


Fig. 3. Distributions of the number of questions posted and the number of answers provided.

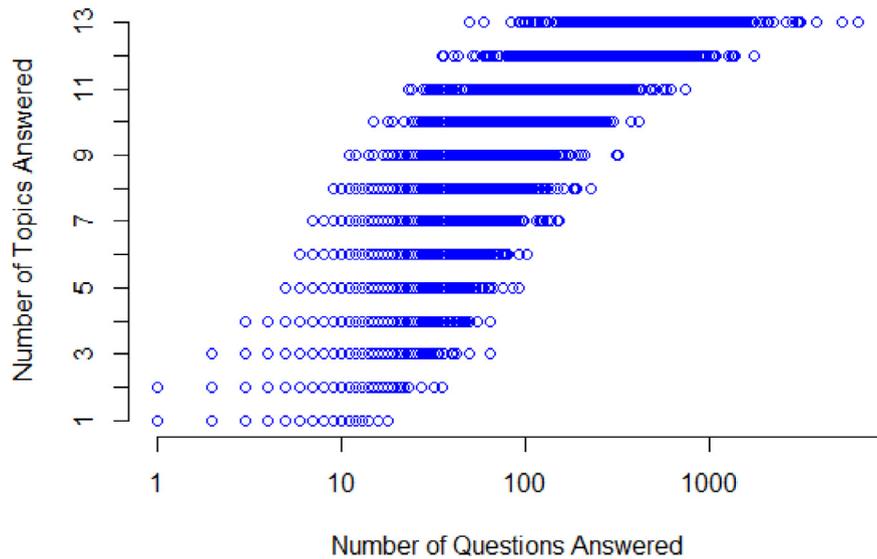


Fig. 4. Distribution of number of topics answered and questions answered.

to estimate the collocation strength between two different topics and expected findings, such as people who like to answer Healthcare related questions are likely to answer questions under the topical category of Vexation.

Proposed by Bouma (2009), NPMI is a bi-directional association measurement of the information overlap between two random variables, and it is a method that has been adopted to determine the degree of association between two events (Sousa, Sarmiento, & Mendes Rodrigues, 2010; Zhang & Pennacchiotti, 2013). Compared to Pointwise Mutual Information (PMI) (Church & Hanks, 1990), the results of NPMI are easier to interpret and at the same time are less sensitive to occurrence frequency. In our settings, for a pair of topics T_x and T_y , we calculate their association by relating the probability of questions within T_x and T_y being answered by the same users with the probabilities of questions within T_x and T_y being answered individually. To be more specific, we estimate NPMI as follows:

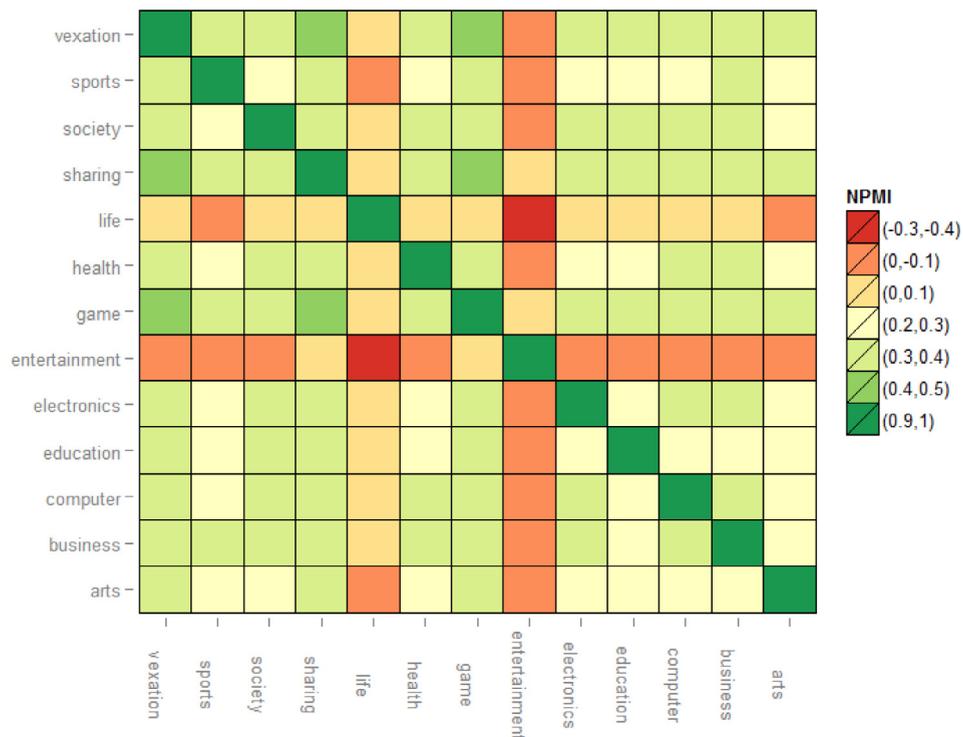


Fig. 5. The NPMI values between topic pairs, with associations displayed in seven level scale from $(-0.3, -0.4)$ to $(0.9, 1.0)$, lowest to highest association.

Eq. (1): Formula to estimate NPMI.

$$\text{NPMI}(T_x, T_y) = \frac{\ln\left(\frac{P(T_x, T_y)}{P(T_x)P(T_y)}\right)}{-\ln(p(T_x, T_y))} \quad (1)$$

To transform the probability distributions into observable frequencies, we define four variables U_x , U_y , U_{xy} , and U denoting the number of users who have answered questions within the topical category T_x ; the number of users who have answered questions within the topical category T_y ; the number of users who have answered questions under both topics; and the total number of users within our dataset. So the formula of NPMI can be rewritten as:

Eq. (2): Revised formula to estimate NPMI.

$$\text{NPMI}(T_x, T_y) = \frac{\ln\left(\frac{U_{xy}U}{(U_{xy}+U_x)(U_{xy}+U_y)}\right)}{-\ln\left(\frac{U_{xy}}{U}\right)} \quad (2)$$

The result of NPMI ranges from -1 to 1 , with the positive value indicating the association of two topics appearing together and the negative value indicating the association of not appearing together, and with 0 indicating statistically significant independence. We plot the calculated NPMI value for each topical pair within our dataset in Fig. 5.

As can be seen in Fig. 5, we discover that individuals answering questions under the topical categories of Life and Entertainment limit themselves to the current topics only and do not answer questions under the other topical categories. In contrast, we notice a number of topical pairs with positive associations, such as Game and Sharing (0.45), Sharing and Vexation (0.41), Health and Vexation (0.38), Computer and Game (0.38), etc.

5.3. User connectedness in knowledge sharing

In addition to our analysis on user engagement and interest, we also evaluate the impact of the question routing mechanism on user connectedness. In order to do that, we apply the bow tie structure analysis (Broder et al., 2000) to our dataset. The bow tie structure captures complex network structures. The key idea of the method is that a network can be viewed as a bow tie that is connected with four different components: Core, In, Out, and Tendrils/Tube, as shown in Fig. 6. The bow tie structure analysis has been used in previous studies analyzing the network structure of Yahoo!Answers (Chen & Nayak, 2011; Zhang et al., 2007).

In order to fit our Wenwo data into the bow tie model, we create a questioner-answerer graph by connecting users who asked questions with users who responded to these questions. Each node within the graph represents a user who has asked

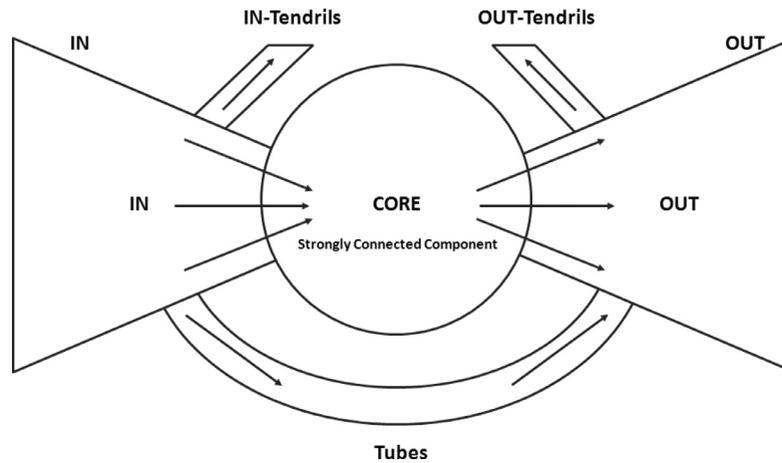


Fig. 6. Illustration of bow tie structure analysis.

Table 1

Connectedness statistics for questions within 13 topical categories in Wenwo.

Topical category	Num of nodes	Num of edges	Core (%)	IN (%)	OUT (%)	T&T (%)	Disconnect (%)	Mutual edges (%)
Life	242,707	472,664	0.52	23.03	6.00	56.58	13.87	0.21
Entertainment	294,473	462,439	0.32	20.02	7.15	60.02	12.48	0.09
Healthcare	49,923	106,711	0.66	30.37	1.52	62.70	4.75	0.40
Arts	27,790	40,516	0.03	2.43	0.29	39.43	57.82	0.16
Sports	19,719	22,948	0.02	1.54	0.10	63.16	35.18	0.07
Society	25,820	41,869	0.30	11.63	1.14	90.26	3.33	0.27
Business	24,024	40,042	0.22	14.01	0.76	74.34	10.89	0.17
Electronics	41,708	88,703	0.34	16.90	1.18	75.57	6.01	0.19
Education	49,514	73,637	0.23	9.21	1.11	72.68	17.01	0.16
Computer	45,098	77,479	0.35	13.79	1.09	71.80	12.97	0.19
Game	13,704	18,720	0.28	5.11	0.76	77.10	16.76	0.30
Vexation	19,359	32,639	0.59	12.12	1.88	73.73	11.68	0.44
Sharing	1364	1254	0.01	1.32	0.07	28.74	69.86	0.16

or answered a question, while each edge corresponds to the directed reply relationship between the questioner and the answerer. The graph contains a total of 715,907 nodes and 1,538,427 edges. The CORE component is the largest, most strongly connected component (SCC) of the questioner-answerer graph, in which any two users are mutually reachable by following the direct question-answering relationship. With the core component, we can detect the largest group of individuals who tend to help each other directly or indirectly on Wenwo. The IN component contains all nodes that are not part of the CORE but can reach it via directed paths. Users who always ask questions but rarely answer will primarily belong to the IN component. Similarly, the OUT component contains nodes that are reachable from the CORE via directed paths and, in our case, represents users who answer but infrequently ask. The Tendrils and Tube component (T&T) contains users who ask or answer only questions posted or responded by the users within the IN and OUT components.

For the Q&A graph, the CORE component contains 9552 nodes, which corresponds to 1.33% of all the users, which is quite different from the results reported in previous studies on Yahoo!Answers (Zhang et al., 2007). This indicates that the question answering process on Wenwo is not as social as one expects. A small proportion of users are connected on Wenwo through question and answering activities, while most of the users are quite segregated. In addition, to evaluating the reciprocal relationships between the questioners and the answerers, we also count the number of mutual edges in our created graph. We find that among all 1,538,427 edges, 9313 (0.60%) are mutually connected. We believe this indicates the well-separated roles played by the knowledge sharer and consumer in question routing environments, like Wenwo.

To test whether user connectedness correlate with topical interest, we measure the size of the largest SCC and the number of mutual edges within all 13 topical categories. Table 1 shows the results. Compared with the SCC measurement as reported in Adamic et al. (2008)'s work, we observe that the percentage of the nodes within the CORE for each individual category is much smaller than that for the whole dataset. None of the topical categories in Wenwo are well-connected, as all largest SCC contained less than 1% of the users within that topical category. We believe this might be due to the well-separated roles of knowledge sharer and consumer on Wenwo, as we discussed in an earlier section.

A further look at the results in Table 1 reveals that individuals who posted and answered questions under the topical category of Health, Vexation, and Life were relatively connected, with both a high percent of users within the largest SCC and a relatively large number of mutual edges. In other words, as compared to other topics, users focusing on those three

categories were more likely to answer each other's questions. This indicates the existence of connected communities within the social Q&A process under those topics. We assume that this might because of the common ground existed between individuals sharing either the same living background (as many of the Life questions were location specific, so only users from the same regions can answer those questions), or physical (individuals who experienced or known someone who had the diseases), or emotional conditions (individuals who had or known someone who suffered from the vexation). In contrast, users within the topical categories such as Arts, Sports, and Sharing were relatively less connected.

We also notice that among all topical categories, there were more users contained within the IN component than in the OUT component, especially for the topical categories of Health, Electronics, Computer, and Business. This is consistent with the nature of Wenwo, where users actively seek for help; however, compared with our previous results shown in Fig. 3, we notice that the majority of answers are provided by a small number of active answerers.

6. Predicting active knowledge sharers using non-Q&A features

So far, we have examined the knowledge sharing behaviors among relative strangers in Wenwo. From our analysis, we note the importance of identifying active knowledge sharers in social Q&A environments, as they not only provided the largest proportion of answers, but they are part of the largest component within the questioner-answerer graph. Therefore, we propose an experiment on predicting the active knowledge sharers in social Q&A environments.

Past studies focusing on this topic relied on obtaining a user's past Q&A records (Dror, Koren, Maarek, & Szpektor, 2011; Jurczyk & Agichtein, 2007; Luo et al., 2014); however, as suggested by Pan et al. (2013), answerer prediction relying only on historical Q&A records suffers from the risks of losing the potential contribution of those who are new to social Q&A services due to the cold-start problem. In order to be more effective in such cold-start recommendation conditions, in this study, we build our model based only on non-Q&A traces. We derive those non-Q&A features from users' Weibo accounts that are linked to their Wenwo profile. We assume that an individual's online social behaviors and patterns can reflect to some extent their internal traits and can thus be used to predict their potentiality in sharing their knowledge with others.

6.1. Active knowledge sharers vs. non-sharers

In order to identify the active knowledge sharers, we pick all users who have answered more than 50 questions in our collected dataset. Besides, given the unequal number of questions routed to those users, we further constrain our sample based on their response rate. Only individuals who have answered more than 90% of questions routed to them are included in our sample. We also manually tag for enterprise and marketing accounts and remove them from our sample considering the advertising nature of both their answers and their Weibo posts. In the end, we identify 1503 active knowledge sharers who answered 490,910 questions, with a mean answer rate of 326.62, and a standard deviation of 312.72.

To form the negative samples of non-sharers, we include individuals with response rates of less than 10%. Additionally, to ensure that those users were intentionally, repeatedly not answering questions being routed to them, we only select users with more than four unanswered questions. Again, we remove the enterprise and marketing accounts in order to be consistent with our sample of the active knowledge sharers. This left us with a total of 2980 non-sharers, based on which we randomly selected equal number of instances to produce a balanced dataset, which contains 1503 active knowledge sharers and 1503 non-sharers. In total, only 17 questions are answered by the 1503 non-sharers, with a mean of 0.01 and standard deviation of 0.12. The division thresholds of the active knowledge sharers and non-shares were chosen arbitrarily. We admit that this is a somewhat suboptimal cut-off, but we think the prediction model built based on these user sets can still be valuable to obtain a better understanding of the characteristics of active knowledge sharers in the social Q&A process.

6.2. Feature engineering

Given our second research objective is to predict active knowledge sharers in Wenwo based on no historical Q&A records, we introduce four non-Q&A feature classes based on individual's Weibo feeds including: (a) user profile, (b) posting behavior, (c) language style, and (d) social activities. Next, we introduce each of the four types of features in detail.

- Profile features

The profile features indicate the identity of users, as well as their activeness in virtual worlds. Features such as gender, number of followers, etc., have been investigated in a number of prior studies as indicators of individual's intrinsic characteristics. For instance, Ross et al. (2009) found a positive correlation between extraversion and the number of Facebook friends. Correa, Hinsley, and De Zuniga (2010) noticed that for females, there was a positive correlation between their openness and extraversion. Inspired from both studies, the profile features that we adopt in our model include: gender, whether or not is a verified account, number of followees, number of followers, longevity of the account, and posting frequency per day. The first four features can be retrieved directly from one's Weibo profile, while the posting frequency feature can be calculated by dividing the total number of status by the longevity of the account.

- Posting behavior features

Posting behavior features capture the way individuals use social networking services and how they interact with others. We believe that 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. Java, Song, Finin, and Tseng (2007) suggested that users who rarely post but follow those with many followers tend to be information seekers, while users who often post URLs in their tweets are most likely to be information providers. To verify these claims, we test 8 posting behavior features including: the percentage of retweets, the percentage containing at-mentions, URLs, hashtags, and emoticons, the percentage of original posts containing images or videos, and the average length of the posts. To quantify each of these features, we collect up to 450 of the most recent posts for each active and non-active user in our dataset. All posts are maintained in Chinese for later analysis.

- Language style features

Features under this category provide a comprehensive description of the psychological and linguistic characteristics of individuals on SNS, which help us better classify users based on their internal traits. To measure those language style features, Pennebaker and Francis (1999) created the Linguistic Inquiry and Word Count program (LIWC), which maps the relative word frequency to a set of psychological dimensions, such as linguistic dimensions (e.g., pronouns, tense), psychological constructs (e.g., positive motion), and personal concerns (e.g., leisure, death). Given its popularity, LIWC has been widely adopted as a psycho-linguistic measurement in several prior studies (Chen et al., 2014; Garimella, Weber, & Dal Cin, 2014; Tumasjan, Sprenger, Sandner, & Welpe, 2010)

To measure the linguistic features in Weibo posts, we use a simplified Chinese version of LIWC called TextMind (Gao, Hao, Li, Gao, & Zhu, 2013); however, unlike LIWC, TextMind defines 71 word categories (68 in LIWC 2001), each containing a number of corresponding words. With the help of TextMind, we count the number of words used under each of the 71 categories for every active and non-active user in our dataset. We assess the sufficiency of the data in each dimension and then decide to adopt all 71 categories for later analysis. We normalize the counts of every category with the total number of words appearing in all 450 posts.

- Social activities features

The features of social activities capture the social attractiveness of the questioner and his/her social engagement with followers based on past activities. We introduce a total of 3 social activity features, which are the average number of retweets, comments, and likes received. These features reveal the content value of the questioner, as well as his/her popularity among followers.

6.3. Classification algorithms and evaluation metrics

With the above features, we next build a binary classifier to automatically differentiate active knowledge sharers from the non-sharers. We train and test our model using a number of classification algorithms implemented in Weka (Hall et al., 2009), including: NaïveBayes, SVM (SMO) with linear kernel, and Logistic Regression, using 10-fold cross-validations. We choose these three models as our classifiers since both NaïveBayes and SVM have been proved to be effective in text classification tasks, and Logistic Regression enables us to demonstrate the relative importance of the different predictors in explaining individual's desire of knowledge sharing. Inspired by the study of Banerjee, Chua, and Kim (2015), we also include an ensemble classifier (Vote), which combines all three above-mentioned models together by voting through the average probability. For evaluation purposes, we use the traditional metrics, including: precision, recall, F1 and accuracy, as they have been adopted in many other studies (Castillo, Mendoza, & Poblete, 2011; Liu, Bian, & Agichtein, 2008). A majority vote baseline (50% accuracy) is applied to determine the baseline performance of our classifiers.

6.4. Classification results

For all four classification algorithms, we find that logistic regression achieves the best performance. A summary of the results is shown in Table 2. We see that the classifier using features from all four perspectives achieves a prediction accuracy of 70%, which is higher than the baseline accuracy of 50%. Considering only non-QA features have been adopted, we deem this method suitable to solve the cold start problem in predicting potential contributors in social Q&A sites. We see that the profile-based features show very limited effects on identifying potential answerers, whereas one's language style best predict whether or not he /she would be an active knowledge sharer in social context.

Table 3 summarizes the results of logistic regression using features from all four categories. Given the large number of features included in our model, we only list significant features with their correlation coefficient and p-value. We see from Table 3 that (in general) less popular and less social users on Weibo actually contribute more to answering stranger's questions in the social Q&A process. Next, we explain each of the significant features in more detail.

Table 2

Classification results using Naïve Bayes, SVM, and Logistic Regression.

Classifier	Precision	Recall	Accuracy	F1
<i>Profile features</i>				
Naïve Bayes	0.57	0.56	0.55	0.55
SVM	0.58	0.57	0.57	0.56
Logistic Regression	0.61	0.61	0.61	0.61
Vote	0.62	0.61	0.61	0.61
<i>Posting behavior features</i>				
Naïve Bayes	0.60	0.59	0.59	0.59
SVM	0.62	0.62	0.62	0.62
Logistic Regression	0.63	0.63	0.63	0.62
Vote	0.63	0.62	0.62	0.62
<i>Language style features</i>				
Naïve Bayes	0.61	0.61	0.61	0.61
SVM	0.63	0.63	0.63	0.63
Logistic Regression	0.66	0.66	0.66	0.66
Vote	0.65	0.65	0.65	0.65
<i>Social activity features</i>				
Naïve Bayes	0.59	0.59	0.58	0.58
SVM	0.62	0.61	0.61	0.61
Logistic Regression	0.62	0.62	0.62	0.61
Vote	0.62	0.61	0.61	0.60
<i>Overall</i>				
Naïve Bayes	0.63	0.63	0.63	0.63
SVM	0.67	0.67	0.67	0.67
Logistic Regression	0.71	0.70	0.70	0.70
Vote	0.70	0.70	0.70	0.70

Table 3Results of Logistic regression with coefficients and *p*-values of each feature for prediction of willingness to provide response to question.

Feature	Coefficient	<i>p</i> -value
<i>Profile features</i>		
Num of Reciprocal Relations	0.03	0.00***
Num of Follower	−0.01	0.00***
<i>Posting behavior features</i>		
Percent of Mention	1.83	0.00***
Percent of URL	0.89	0.01**
Percent of Retweet	−0.67	0.01**
<i>Language style features</i>		
Pronouns (total pronouns, I, them, itself)	−111.96	0.02*
Personal Pronouns (I, them, her)	−618.83	0.00***
Second Person Pronouns (you, honorific of you)	−440.02	0.00***
I (I, me, mine)	−516.26	0.00***
We (we, us, our)	−515.66	0.00***
You (you, your, thou)	−489.72	0.00***
She he (she, her, him)	−497.23	0.00***
They (they, their, they'd)	−516.42	0.00***
Impersonal Pronouns (it, it's, those)	103.47	0.03*
Verb (common verbs)	11.10	0.01**
Fillier (blah, I mean, you know)	93.3	0.00***
Insight (think, know, consider)	32.73	0.01**
Cause (because, effect, hence)	40.53	0.01**
Tentative (maybe, perhaps, guess)	28.33	0.01**
Feel (feels, touch)	50.32	0.01**
Work (job, majors)	−10.09	0.03*
Achieve (earn, hero, win)	−17.72	0.03*
<i>Social activity features</i>		
Percent of Liked Post	2.38	0.00***
Percent of Retweeted Post	−2.28	0.00***
Percent of Total Interactions	−1.55	0.04*

*** *p*-value < 0.001.** *p*-value < 0.01.* *p*-value < 0.05.

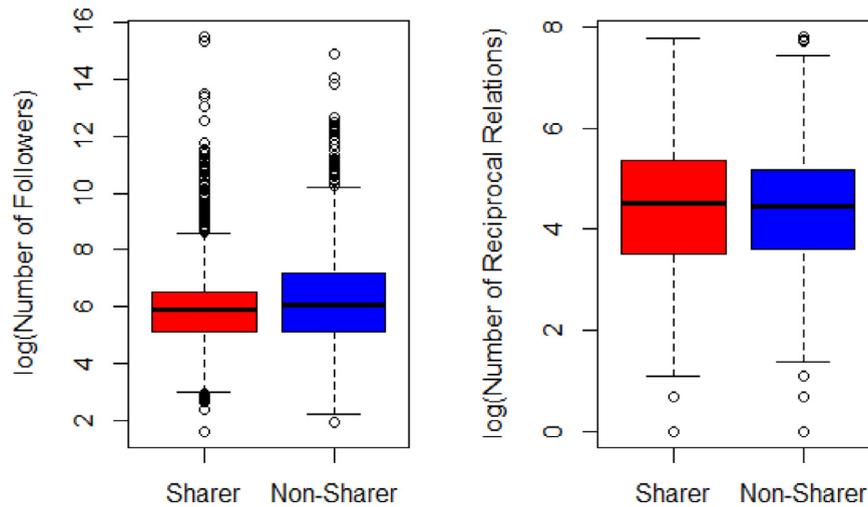


Fig. 7. Difference in log scaled number of followers and reciprocal relations for knowledge sharers and non-sharers.

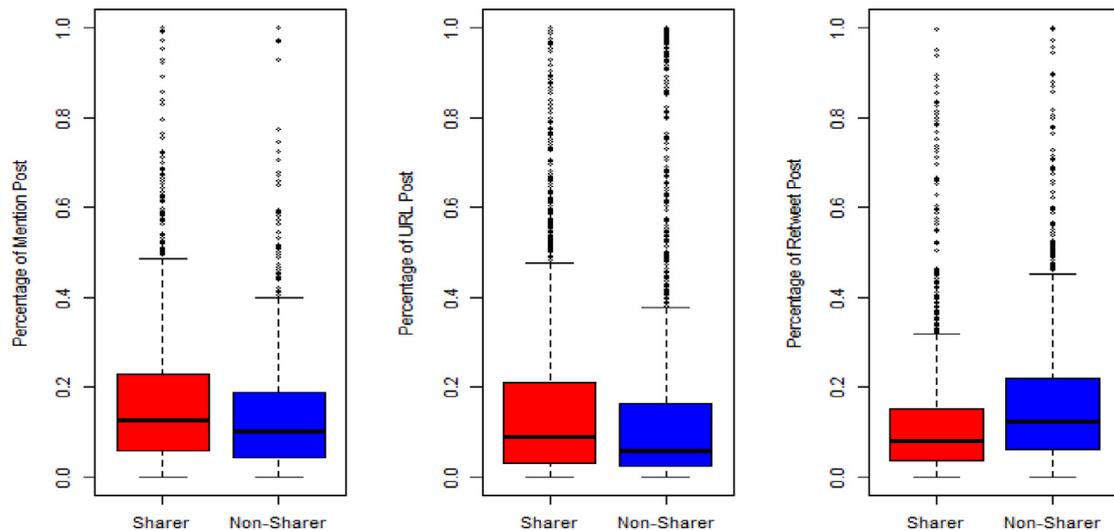


Fig. 8. Difference in percentage of mention, URL sharing, and retweet posts for knowledge sharers and non-sharers.

From the profile perspective, we find that a user's knowledge sharing behaviors are independent of gender and social status (verified or not). Individuals with fewer followers ($t = -0.07$, $p < 0.01$) but more reciprocal relations ($t = 6.82$, $p < 0.01$) were more likely to answer strangers' questions being routed to them. Fig. 7 shows the differences in the number of followers and reciprocal relations for active knowledge sharers and non-sharers in our dataset.

We notice that regarding the posting behaviors, non-sharers tended to retweet more than contributors ($t = -1.82$, $p = 0.04$); whereas, knowledge sharers interacted more with others by adopting a larger amount of @mentions in their post ($t = 6.87$, $p < 0.01$). Compared with non-sharers, knowledge sharers are more willing to spread cross-platform information with other, given that a slightly higher percentage of their tweets contained URLs, although such difference is not statistically significant ($t = 0.76$, $p = 0.45$). We demonstrate those differences in boxplots as shown in Fig. 8.

With the psycho-linguistic features, we observe that users' contributing behaviors are significantly associated with a set of LIWC measures. First, we find that posts of non-sharers contain more personal pronouns ($t = 4.41$, $p < 0.01$), such as I, we, you she, he, etc., indicating the more mundane nature of their posts; however, sharers use more impersonal pronouns ($t = 1.38$, $p = 0.05$), such as it, it's, those, etc. As suggested by Ravid, van Hell, Rosado, and Zamora (2002), there is a positive correlation between the usage of impersonal pronouns and expository text. In addition, we also observe more verbs ($t = 6.36$, $p < 0.01$) and fillers ($t = 3.61$, $p < 0.01$) in knowledge sharer's Weibo posts. Prior studies on personality traits implied that high extraverts use more verbs to make their descriptions more lively (Oberlander & Gill, 2004). Third, active knowledge sharers also used more words to describe their underlying cognitive mechanisms ($t = 2.61$, $p < 0.01$), including cause, insight, and tentative expressions. Fourth, the knowledge sharing behavior is negatively associated with the use of work ($t = -3.49$, $p < 0.01$) and achieve words ($t = -1.57$, $p = 0.01$), which according to Schwartz (1992)'s theory of human values reflect high self-enhancement.

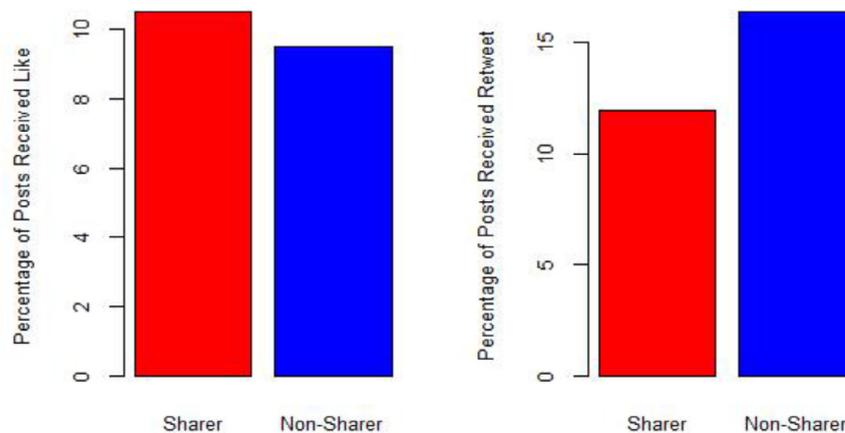


Fig. 9. Difference in percentage of posts received like, retweet, or any type of interaction for knowledge sharers and non-sharers.

Lastly, from the social activity perspective, we see that posts of active knowledge contributors tend to receive more ($t = 1.64$, $p = 0.09$), although not significantly more, likes, but less retweet ($t = -8.34$, $p < 0.01$) than non-sharer's posts. In other words, sharers are better at maintaining weak social relationships than non-sharers. Fig. 9 illustrates such differences.

7. Discussion and impacts

In this research, we evaluate the performance of question routing services from three different perspectives, including user engagement, interests, and social connectedness with others. First, our results prove the effectiveness of question routing systems in attracting knowledge shares. Questions posted on Wenwo receive a much higher response probability than questions posted on many other platforms, even though a large proportion of those answers are contributed by a small number of individuals. Second, the respondents' topical interest increases along with their response frequencies with more contributing individuals answering questions under more topical categories. Third, through a network analysis, we find that users participated in the question routing process seemed less connected than users in traditional Q&A settings. A more detailed analysis within each topical category further indicates the existence of connected communities under topics of Vexation, Life, and Healthcare. We believe this might be because of the common ground existed between individuals sharing either the same living background, or physical and emotional conditions under those topics.

In addition to our assessment of the question routing system, we also build a classification model based on a set of non-Q&A features. The classifier demonstrated that using only non-Q&A features, we can predict a potential contributor on social Q&A sites, performing much better than the baseline of 50%. By analyzing each significant predictor in our regression model, we notice that less popular but more interactive individuals on Weibo actually contribute more in the social Q&A process. Users with more original and URL sharing posts answer more questions than those who retweeted a lot. From the psycho-linguistic perspective, we identify that individuals use less pronouns and achievement-related words, but more verbs and cognation-related expressions, tend to be more active in answering other's questions.

Our contribution is two-fold: first, we evaluate the effectiveness of the question routing scheme in the social Q&A process. Although previous studies (Li & King, 2010; Zhou et al., 2012) have suggested the adoption of question routing scheme in Q&A sites, such as Yahoo!Answers and even SNS, to the best of our knowledge, this is one of the first studies that actually evaluates the performance of question-routing services in a real world setting. As our results suggest, there are both pros and cons to using a question routing scheme under social Q&A context, as it effectively increased the response rate of questions posted; however, it also decreased the social interactivity between users on SNS. Second, while identifying proper individuals for question routing, our study adopts a different perspective than most prior work (Liu, Croft, & Koll, 2005; Pal et al., 2012). Instead of assessing an individual's capabilities with respect to their answering probabilities, we propose our predicting model from the perspectives of the individual's desire in helping others when facing questions that fit their expertise because we believe that capability is not the only factor that determines one's answering behavior in the social Q&A process. We believe our prediction model can be applied to community Q&A sites, such as Yahoo!Answers, Quora, and Baidu Knows, given that the language style features itself can provide very effective separation between knowledge sharers and non-sharers.

In analyzing our results, we are aware of certain limitations that may restrict the ability to generalize our conclusions. One limitation is that our study is based on only one social Q&A service, Wenwo, so it may not be representative of the question answering behaviors demonstrated in other social Q&A sites, especially the cross-country services, such as Twitter and Facebook; however, in their study analyzing the Q&A behaviors on Baidu Knows, China's largest Q&A site, Yang and Wei (2009) demonstrated similar characteristics as those reported for Yahoo!Answers (Adamic et al., 2008; Zhang et al., 2007). Therefore, we expect the same carryover with Wenwo, although we certainly plan to apply our analysis meth-

ods on various platforms. There is also a limitation in the methodology of defining sharers and non-sharers on Wenwo. The thresholds that we used in this study was determined arbitrarily according to our observation of the collected data. It is a matter of debate whether these thresholds have any bias on our classification results, as well as has any validity in terms of discriminating sharers and non-sharers on other platforms. Another limitation is that only non-Q&A features are adopted in this study to avoid the cold start problem.

In future work, we would like to further improve our classifier by involving Q&A based features, such as previous answers provided, best answers provided, etc. Additionally, we would like to research incentive mechanisms, which was currently adopted in Wenwo, to motivate non-sharers to share their knowledge with others. Besides, inspired by some early work in response evaluation (Chua & Banerjee, 2015b; Kitzie & Shah, 2011; Mamykina, Manoim, Mittal, Hripcsak, & Hartmann, 2011), we think that another interesting extension of the current work could be examining the quality and the promptness of answers received via question routing services in social context.

8. Conclusion

To explore the knowledge contributing behaviors among strangers in the social Q&A process, we experiment with Wenwo, a third party social Q&A application based on Sina Weibo. In order to understand the effectiveness of question routing in social Q&A environments, we analyze the Q&A activities on Wenwo collected over a ten-month period. One of the problems with such question routing system was that, even if we know a person who had the expertise and experience to answer a question, their willingness to help was still indeterminate. To solve this problem, we propose a predicative model automatically identifying active knowledge sharers using non-Q&A features from four dimensions: profile, posting behavior, language style, and social activities. We adopt only non-Q&A features in our model considering the cold start situations. We find that from what an individual posts (e.g. mention, retweet, URL sharing, etc.) and how he/she posts it (e.g., usage of pronouns, achievement-related words, feeling-related words, etc.), one could predict with a high degree of accuracy if the user is willing to answer a stranger's question that fits his/her expertise.

We believe that this study provides a theoretical understanding of the knowledge contribution behaviors and patterns in the social Q&A process, as it offers both the advantages and bottlenecks of the adoption of a question routing scheme under social Q&A context. In addition, we view our work as a crucial step toward designing and implementing more accurate question routing mechanisms in social contexts, given its consideration to an individual's desires to answer questions, in addition to their capabilities.

References

- Adamic, L. A., Zhang, J., Bakshy, E., & Ackerman, M. S. (2008). Knowledge sharing and yahoo answers: Everyone knows something *Paper presented at the 17th international conference on world wide web*.
- Bai, J. (2014). Weibo user development report. Retrieved from <http://data.weibo.com/report/reportDetail?id=215>.
- Banerjee, S., Chua, A. Y., & Kim, J. J. (2015). Let's vote to classify authentic and manipulative online reviews: The role of comprehensibility, informativeness and writing style. *Paper presented at the science and information conference London*.
- Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. *Paper presented at the bi-annual conference of the german society for computational linguistics and language*.
- Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., et al. (2000). Graph structure in the web. *Computer Networks*, 33(1), 309–320.
- Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. *Paper presented at the 20th international conference on world wide web*.
- Chen, J., Hsieh, G., Mahmud, J. U., & Nichols, J. (2014). Understanding individuals' personal values from social media word use. *Paper presented at the 17th ACM conference on computer supported cooperative work & social computing*.
- Chen, L., & Nayak, R. (2011). Social network analysis of an online dating network. *Paper presented at the fifth international conference on communities and technologies*.
- Chua, A., & Banerjee, S. (2015a). Answers or no answers: Studying question answerability in stack overflow. *Journal of Information Science*, 41, 720–731.
- Chua, A., & Banerjee, S. (2015b). Measuring the effectiveness of answers in Yahoo! Answers. *Online Information Review*, 39(1), 104–118.
- Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1), 22–29.
- Correa, T., Hinsley, A. W., & De Zuniga, H. G. (2010). Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, 26(2), 247–253.
- Dror, G., Koren, Y., Maarek, Y., & Szpektor, I. (2011). I want to answer; who has a question?: Yahoo! answers recommender system. *Paper presented at the 17th ACM international conference on information and knowledge management*.
- 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*.
- Gao, R., Hao, B., Li, H., Gao, Y., & Zhu, T. (2013). Developing simplified chinese psychological linguistic analysis dictionary for microblog. *Paper presented at the 2013 international conference on brain and health informatics*.
- Garimella, V. R. K., Weber, I., & Dal Cin, S. (2014). From "I love you babe" to "leave me alone" - romantic relationship breakups on Twitter. *Paper presented at the 6th international conference on social informatics*.
- Gazan, R. (2007). Seekers, sloths and social reference: Homework questions submitted to a question-answering community. *New Review of Hypermedia and Multimedia*, 13(2), 239–248.
- Guo, J., Xu, S., Bao, S., & Yu, Y. (2008). Tapping on the potential of q&a community by recommending answer providers. *Paper presented at the 17th ACM conference on information and knowledge management*.
- Gyongyi, Z., Koutrika, G., Pedersen, J., & Garcia-Molina, H. (2008). Questioning Yahoo!Answers. *Paper presented at the first WWW workshop on question answering on the web*.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, 11(1), 10–18.
- Jansen, B. J., Sobel, K., & Cook, G. (2011). Being networked and being engaged: The impact of social networking on ecommerce information behavior. *Paper presented at the 2011 iConference*.
- 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.

- Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter: Understanding microblogging usage and communities. *Paper presented at the 9th WebKDD and 1st SNA-KDD 2007 workshop on web mining and social network analysis.*
- Jurczyk, P., & Agichtein, E. (2007). Discovering authorities in question answer communities by using link analysis. *Paper presented at the 16th ACM conference on information and knowledge management.*
- Kitzke, V., & Shah, C. (2011). Faster, better, or both? Looking at both sides of online question-answering coin. *Paper presented at the American society for information science and technology.*
- 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.*
- Lee, K., Mahmud, J., Chen, J., Zhou, M., & Nichols, J. (2014). Who will retweet this?: Automatically identifying and engaging strangers on twitter to spread information. *Paper presented at the 19th international conference on intelligent user interfaces.*
- Li, B., & King, I. (2010). Routing questions to appropriate answerers in community question answering services. *Paper presented at the 19th ACM international conference on information and knowledge management.*
- Liu, X., Croft, W. B., & Koll, M. (2005). Finding experts in community-based question-answering services. *Paper presented at the 14th ACM international conference on information and knowledge management.*
- Liu, Y., Bian, J., & Agichtein, E. (2008). Predicting information seeker satisfaction in community question answering. *Paper presented at the 31st annual international ACM SIGIR conference on research and development in information retrieval.*
- Liu, Y., Kliman-Silver, C., & Mislove, A. (2014). The tweets they are a-changin': Evolution of twitter users and behavior. *Paper presented at the 8th international AAAI conference on weblogs and social media.*
- Liu, Z., & Jansen, B. J. (2012). Almighty twitter, what are people asking for? *Paper presented at the American society for information science and technology.*
- Liu, Z., & Jansen, B. J. (2013). Factors influencing the response rate in social question and answering behavior. *Paper presented at the 16th ACM conference on computer supported cooperative work & social computing.*
- Luo, L., Wang, F., Zhou, M. X., Pan, Y., & Chen, H. (2014). Who have got answers?: Growing the pool of answerers in a smart enterprise social QA system. *Paper presented at the 19th International conference on intelligent user interfaces.*
- Mahmud, J., Zhou, M. X., Megiddo, N., Nichols, J., & Drews, C. (2013). Recommending targeted strangers from whom to solicit information on social media. *Paper presented at the 2013 international conference on intelligent user interfaces.*
- Mamykina, L., Manoim, B., Mittal, M., Hripscak, G., & Hartmann, B. (2011). Design lessons from the fastest q&a site in the west. *Paper presented at the SIGCHI conference on human factors in computing systems.*
- 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 2010 SIGCHI conference on human factors in computing systems.*
- 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 ICCHI conference on human factors in computing systems.*
- Nichols, J., & Kang, J.-H. (2012). Asking questions of targeted strangers on social networks. *Paper presented at the 15th ACM Conference on computer supported cooperative work & social computing.*
- Oberlander, J., & Gill, A. (2004). Individual differences and implicit language: Personality, parts-of-speech and pervasiveness. *Paper presented at the 26th annual conference of the cognitive science society.*
- Oh, S. (2011). 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.
- Pal, A., Chang, S., & Konstan, J. A. (2012). Evolution of experts in question answering communities. *Paper presented at the sixth international AAAI conference on weblogs and social media.*
- Pan, Y., Luo, L., Chi, C., & Liao, Q. (2013). To answer or not: What non-qa social activities can tell. *Paper presented at the 16th ACM conference on computer supported cooperative work & social computing.*
- 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.*
- Pennacchiotti, M., & Popescu, A. M. (2011). A machine learning approach to Twitter user classification. *Paper presented at the 5th international AAAI conference on weblogs and social media.*
- Pennebaker, J. W., & Francis, M. E. (1999). *Linguistic inquiry and word count (LIWC)*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ravid, D., van Hell, J. G., Rosado, E., & Zamora, A. (2002). Subject NP patterning in the development of text production: Speech and writing. *Written Language & Literacy*, 5(1), 69–93.
- Ross, C., Orr, E. S., Sisic, M., Arseneault, J. M., Simmering, M. G., & Orr, R. R. (2009). Personality and motivations associated with Facebook use. *Computers in Human Behavior*, 25(2), 578–586.
- Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, 25(1), 1–65.
- Shah, C., Oh, J. S., & Oh, S. (2008). Exploring characteristics and effects of user participation in online social Q&A sites. *First Monday*, 13(9) Retrieved from <http://firstmonday.org/article/view/2182/2028>.
- Shah, C., Oh, S., & Oh, J. S. (2009). Research agenda for social Q&A. *Library & Information Science Research*, 31(4), 205–209.
- Sousa, D., Sarmento, L., & Mendes Rodrigues, E. (2010). Characterization of the twitter@ replies network: Are user ties social or topical? *Paper presented at the 2nd International Workshop on Search and Mining User-generated Contents.*
- Srba, I., Grznar, M., & Bielikova, M. (2015). Utilizing non-QA data to improve questions routing for users with low QA activity in CQA. *Paper presented at the 2015 IEEE/ACM international conference on advances in social networks analysis and mining.*
- Teevan, J., Collins-Thompson, K., White, R. W., Dumais, S. T., & Kim, Y. P. o. p. A. (2013). Slow search: Information retrieval without time constraints. *Paper presented at the 7th annual symposium on human-computer interaction and information retrieval.*
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *Paper presented at the 4th international AAAI conference on weblogs and social media.*
- Wallsten, K. (2008). Political blogs: Transmission belts, soapboxes, mobilizers, or conversation starters. *Journal of Information Technology & Politics*, 4(3), 19–40.
- 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.*
- Yang, J., & Wei, X. (2009). Seeking and offering expertise across categories: A sustainable mechanism works for Baidu Knows. *Paper presented at the 3th International AAAI conference on weblogs and social media, San Jose, USA.*
- Yang, L., Bao, S., Lin, Q., Wu, X., Han, D., & Su, Z. (2011). Analyzing and predicting not-answered questions in community-based question answering services. *Paper presented at the 25th Conference on artificial intelligence San Francisco.*
- Zhang, J., Ackerman, M. S., & Adamic, L. (2007). Expertise networks in online communities: Structure and algorithms. *Paper presented at the 16th International conference on world wide web.*
- Zhang, M., Jansen, B. J., & Chowdhury, A. (2011). Business engagement on Twitter: A path analysis. *Electronic Markets*, 21(3), 161–175.
- Zhang, Y., & Pennacchiotti, M. (2013). Predicting purchase behaviors from social media. *Paper presented at the 22nd international conference on world wide web, Rio de Janeiro.*
- Zhou, T. C., Lyu, M. R., & King, I. (2012). A classification-based approach to question routing in community question answering. *Paper presented at the 21st international conference companion on world wide web.*