SocialQ&A: An Online Social Network Based Question and Answer System

Haiying Shen, Guoxin Liu, Nikhil Vithlani
The Department of Electrical and Computer Engineering
Clemson University, Clemson, SC 29631, USA
Email: {shenh, guoxinl, nvithla}@clemson.edu

***Abstract*—Question and Answer (Q&A) systems play a vital role in our daily life for information and knowledge sharing. Users post questions and pick questions to answer in the system. Due to the rapidly growing user population and the number of questions, it is unlikely for a user to stumble upon a question by chance that (s)he can answer. Also, altruism does not encourage all users to provide answers, not to mention high quality answers with a short answer wait time. The primary objective of this paper is to improve the performance of Q&A systems by actively forwarding questions to users who are capable and willing to answer the questions. To this end, we have designed and implemented SocialQ&A, an online social network based Q&A system. SocialQ&A leverages the social network properties of common-interest and mutual-trust friend relationship to identify an asker's friends who are most likely to answer the question. We describe the architecture, algorithms and user interface of SocialQ&A, and analyze the Q&A behavior of real users and questions from a small-scale real-world SocialQ&A system. We also conducted comprehensive large-scale simulation to evaluate SocialQ&A in comparison with other methods. Our results suggest that social networks can be leveraged to improve the answer quality and asker's waiting time.**

Keywords*—*Question and answer systems, Social networks, Information search

I. INTRODUCTION

The Internet is an important source of information, where the amount of data is vast and constantly growing. Users rely on search engines to find specific information in this knowledge base. Search engines such as *Google* and *Bing* use keywords provided by the users to perform searches. Recently, industrial research and development activities, such as Microsoft and Facebook's social-featured Bing search endeavor, try to combine search engines and online social networks for higher search performance. As previous research has indicated [1, 2], search engines perform well in indexing web pages and providing users with relevant content to their search but are not suited for non-factual questions such as "Which is the best local auto shop?". To address this particular class of non-factual questions, many Question and Answer (Q&A) systems such as *Yahoo! Answers*, *Baidu Zhidao*, *Stack-Exchange*, *Quora* and *Ask* have been developed. Since their inception, Q&A systems have proved to be a valuable resource for sharing expertise and consequently are used by a large number of Internet users. For example, *Yahoo! Answers* was launched at the end of the year 2005 and attracted more than 10 million users in February of 2007 [3], and hit 200 million users in December of 2009 [4, 5]. Q&A systems also preserve all questions and answers, thus acting as a repository for information retrieval. They are not only important for sharing technical knowledge, but also as a source for receiving advice and satisfying one's curiosity about a wide variety of subjects [6].

With a vast population in a Q&A system, a large number of questions are posed online every day. For example, there are 823,966 questions and answers posed to *Yahoo! Answers* per day [4]. Then, when a user intends to answer a question, (s)he may be overwhelmed by the plethora of questions. Moreover, simply relying on altruistic users to provide answers cannot encourage all users to provide answers and to answer questions quickly. To locate appropriate answer providers, current Q&A systems allow users to choose tags (i.e., interest categories) for their questions. However, it may not be easy to determine the appropriate tag(s) for a question such as "how is the computer organization class at our university?".

As a result, current Q&A systems may not meet the requirement of providing high quality answer with a short answer wait time, though users wish to receive satisfactory answers quickly. This is confirmed by the study in [5]. It found that for *Yahoo! Answers*, only 17.6% of questions were answered satisfactorily; for the remaining 82.4%, one fifth of the questions remained unanswered. For *Baidu Zhidao*, 22.7% of questions were successfully answered, and 42.8% of the unresolved questions were not answered at all. Thus, there is an increasing need for an advanced Q&A system that can decrease the number of unanswered questions, enhance the answer quality and decrease the response time.

To meet this need, we propose SocialQ&A, an online social network based Q&A system, that actively forwards questions to those users with the highest likelihood (capability and willingness) of answering them with expertise and interest in the questions' subjects. The design of SocialQ&A is based on two social network properties. First, social friends tend to share similar interests (e.g., lab members majoring in computer systems) [7]. Second, social friends tend to be trustworthy and altruistic due to the property of "friendship fosters cooperation" [8]. Accordingly, SocialQ&A favors routing queries among friends and identifies a question's potential answerers by considering two metrics: the interest of the friend towards the question and the social closeness of the friend to the asker/forwarder. Thus, the answer receivers have high probability of providing high-quality answers in a short time. The contributions of this work are as follows:

1) *The design of SocialQ&A*. SocialQ&A is composed of three components: *User Interest Analyzer*,

978-1-4799-9964-4/15/$31.00 ©2015 IEEE

Question Categorizer, and *Question-User Mapper*. *User Interest Analyzer* associates each user with a vector of interest categories. *Question Categorizer* associates a vector of interest categories to each question. Then, based on user interest and social closeness, *Question-User Mapper* identifies potential answerers for each question.

2) *Comparative trace-driven experiments.* We conducted comprehensive large-scale simulation to evaluate SocialQ&A in comparison with other methods. Our results suggest that SocialQ&A improves the quality of answers and reduces the wait time for answers.
3) *The development of a real-world SocialQ&A.* We have prototyped the SocialQ&A system with user interfaces, and conducted a real-world small-scale test with real users from India, the United Kingdom, and the United States for a period of approximately one month.
4) *The analysis of the data from real SocialQ&A.* We have analyzed the features of the questions posted, the questioning and answering activities of users, the quality of answers, and the wait time for answers. Analytical results show the benefits of SocialQ&A in enhancing answer quality and wait time.

The rest of this paper is structured as follows. Section II presents a concise review of related work. Section III presents the details of the design of SocialQ&A and its user interface. Section IV measures the SocialQ&A's performance in comparison with other systems through trace-driven experiments. Section V analyzes a real trace obtained from the real-world SocialQ&A prototype. Section VI concludes this paper with remarks on our future work.

II. Related Work

The growing importance of Q&A systems demands an effort to better understand these systems and to improve them [9]. The works in [10–13] studied the influence of different factors (e.g., users' profiles, system interactions and community size) in the social networks on Q&A performance. These study results lay the foundation of SocialQ&A to leverage social network properties in the design. Note that the existing social network based on the asker-answerer relationship in current Q&A systems [13] is different from online social network based on the social relationship, which is used in SocialQ&A. The works in [14–17] concentrated on locating experts and authoritative users. Instead, SocialQ&A aims to find normal users that can answer questions including opinion-type questions. Some studies have been conducted to create reputation models in Q&A systems [18, 19] to increase the credibility of answers, and to determine the relationship between the reputation of the users and the quality of their provided answers [20]. SocialQ&A directly utilizes the social network property of mutual-trust friendship to motivate users to provide answers without relying on an additional reputation model.

Some research [21–23] categorizes questions into predefined categories, making it easier for users to locate previously asked questions and for experts to find questions they can answer. Quan *et al.* [22] proposed three new supervised term weighting schemes for question categorization, and evaluated each scheme using a trace from *Yahoo! Answers*. Song *et al.* [23] proposed a sequential process including topic-wise word identification and weighting, semantic mapping, and similarity calculation.

Text mining techniques also have been used to provide better answers [5, 24–27]. These categorization and text mining methods can be used in SocialQ&A to more accurately derive user interests and question interests. Li *et al.* [5] proposed a language model by combining expertise estimation and availability estimation, and later proposed category-sensitive language models [24] for expert identification, which helps route questions to available and capable experts. Zhou *et al.* [25] classified the questions using a variety of local and global features of questions and users' relationship in order to route a classified question to its potential answerers. Cao *et al.* [26] leveraged question category to enhance question retrieval in community-based Q&A systems. Guo *et al.* [27] proposed a topic-based model to identify appropriate answerers by calculating the similarities between questions' topics and users specialists.

Researchers also began to study social networks in the area of search engines [28, 2, 29]. Evans *et al.* [28] identified searching as a social activity and demonstrated that social interactions can help improve the search results. Morris *et al.* [2] discussed the growing trend towards posting queries as social network statuses instead of using web search engines. However, such question flooding to all of a user's friends may overburden the friends, leading to frustration. Horowitz and Kamvar [29] presented a social search engine, which finds the right person to satisfy a user's information need and provides trust based on social intimacy.

Different from previous Q&A system works, this work focuses on system design by leveraging social network properties and shows its promises for performance enhancement. SOS [30] is also a Q&A system based on a social network. However, SOS focuses on realizing a mobile Q&A system in a distributed manner and using knowledge engineering techniques. Also, it assumes that social closeness is already provided by users. Instead, SocialQ&A focuses on how to leverage social network properties in better identifying potential answerers with predefined interest categories and showing its benefits through the analysis on real users' Q&A activities.

III. SocialQ&A: An Online Social Network based Q&A System

A. The Rationale of SocialQ&A Design

A real-life social network is formed by regarding each person as a node and linking two nodes with a social relationship. This network is featured by social communities such as the football club and ECE department at a university. In real life, the people we rely on for answers to questions such as "how is the computer organization class at our university?" are usually those in our social communities. Persons in the same social community share common interests and trust each other on answering questions on their common interests, and are willing to answer the questions from community members.

An online social network connects friends with real-life relationship and online friendship, which shares similarity to the real-life social network. Friends in an online social network tend to share similar interests and trust each other [7, 31, 8]. Taking advantage of these properties, we design and develop SocialQ&A that incorporates an online social network to improve the quality of answers and decrease answer wait time. It forwards a user's questions to his/her social friends that have common interest and a close social relationship.

B. The Design of SocialQ&A

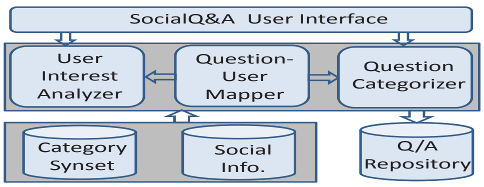


Fig. 1: The architecture of SocialQ&A.

Like all online social networks, the one in SocialQ&A has user profiles that record users' interests, education, hobbies and etc. Like *Yahoo! Answers*, SocialQ&A also predefines interest categories and subcategories. A total of 4 categories (music, movies, television, and books) and 32 subcategories (e.g., books: novel, drama) derived from *Yahoo! Answers* were used to implement SocialQ&A. We used these 4 categories as an example and will add more categories in our future work.

Figure 1 shows the high-level architecture of SocialQ&A and the interaction between the core components: *User Interest Analyzer*, *Question Categorizer*, and *Question-User Mapper*. *User Interest Analyzer* analyzes data associated with each user in the social network to derive user interests. *Question Categorizer* categorizes the user questions into interest categories based on the *Category Synsets*, which stores the synonyms of all categories' keywords from WordNet [32]. *Question-User Mapper* connects these two components by identifying potential answerers who are most likely to be willing to and be able to provide satisfactory answers. The data from user questions and answers is stored on *Q/A Repository* to serve subsequent similar questions. Below, we present each component and user interface.

	Rock	Classic	Action	Thriller	News	Shows	Story
User i	2	0	3	0	1	0	4

Fig. 2: User interest vector.

1) User Interest Analyzer: User Interest Analyzer utilizes each user's profile information in the social network and user interactions (answers provided and questions asked) to determine the interests of the user in the predefined interest categories. This is because if a user asks or answers questions in an interest category, (s)he is likely to be interested in this particular category. As shown in Figure 2, the interests of user U_j are represented by a user interest vector V_{U_j}=<I_i, W_{I_i}> $(i = 1, 2...)$, where I_i represents an interest and W_{I_i} represents the weight (degree) of the user's interest in interest I_i. $W_{I_i} = 0$ indicates that the user does not have the corresponding interest. W_{I_i} is incremented by 1 for each appearance of the interest in the parsed information from a user's profile and interactions. The order of phrases does not necessarily represent the different preferences of a user. Thus, we count the frequency that an interest's synset appears in all phrases to indicate the user's perference on this interest, because the frequency represents a user's focus on an interest currently.

Algorithm 1 Pseudocode for the *User Interest Analyzer*.

Input: A user's profile, questions and answers
Output: The user's interest vector $V_{U_j} =< I_i, W_{I_i} >$

1: Parse the "interests" field to generate a token stream T_I
2: Parse the "activities" field to generate a token stream T_a
3: Use the inputs from the user's selection from the Music, Movie, Television and Book fields to generate token streams T_{mu}, T_{mo}, T_t and T_b
4: **for** each token stream T_x (T_x=T_I, T_a, T_{mu}, T_{mo}, T_t, T_b) **do**
5:   Check each token in the Synset
6:   **if** a matching interest category I_i exists **then**
7:     Update interest weight: W_{I_i}++ (e.g., W_{music}++)
8:   **end if**
9: **end for**
10: Keep updating W_{I_i} based on questions asked and answered and profile update
11: Periodically update W_{I_i} using $W_{I_i} = \alpha * W_{I_{i old}}$

Algorithm 1 shows the pseudocode for the *User Interest Analyzer*. When a user registers for SocialQ&A, (s)he is given the option of entering his/her interests and activities and to mark predefined interest categories to add to his/her interest list. SocialQ&A uses WordNet to parse these text fields to token streams (Steps 1-3). For every token, its matching interest category is located in the Synset and corresponding weight is updated (Steps 4-9).

For accurate user interest reflection, SocialQ&A keeps track of profile changes, the questions asked and answered by a user to update his/her interest vector. A user can indicate the interest tags for his/her questions. In the indicated tags and parsed interests, we use O_{I_i} to denote the number of occurrences of the interest I_i during the previous period. For an interest I_i, its weight is updated to $W_{I_{i new}} = W_{I_i} + O_{I_i}$, where $W_{I_{i new}}$ is the weight used in next period. In order to reflect users' current interests, the weight is periodically decayed by: $W_{I_i} = \alpha * W_{I_{i old}}$, where $W_{I_{i old}}$ is the weight used in last period

2) Question Categorizer: The primary task of *Question Categorizer* is to categorize a question into predefined interest categories based on the topic(s) of the question. We also allow users to input self-defined tags associate with questions, which are analyzed in question parsing. *Question Categorizer* generates a vector of question Q_i's interests, denoted by V_{Q_i}, using a similar algorithm as Algorithm 1. While processing a question, SocialQ&A uses WordNet to examine the tags and text of the question and generates a token string. The tokens are compared to SocialQ&A's Synset to determine the categories where the question belongs.

3) Question-User Mapper: Question-User Mapper identifies the appropriate answerers for a given question. The potential answer providers are chosen from the asker's friends in the online social network. Note that the changes in a user's friends in the online social network do not affect the performance of SocialQ&A as it always uses a user's current friends. To check the appropriateness of a friend (U_k) as an answer provider for a question, two parameters are considered: i) the interest similarity between the interest vectors of the friend and the question (denoted by Ψ_{I,U_k}); and ii) the social closeness between the friend and the asker (denoted by Ψ_{C,U_k}). The former represents the potential capability of a friend to answer the question, and the latter represents the willingness of a friend to answer the question.

We use $W_{I_j}^{U_k}$ to denote U_k's weight on interest I_j. For the asker's question with vector V_{Q_i},

$$\Psi_{I,U_k} = \sum_{I_j \in (V_{U_k} \cap V_{Q_i})} W_{I_j}^{U_k}. \quad (1)$$

In the online social network, a user's friends with more common interests, frequent interactions or common friends (i.e., higher social closeness) are more willing to respond to the user's question [30, 13, 2, 10]. Thus, to calculate Ψ_{C,U_k} between friend U_k and the asker, we consider three metrics: i) the similarity between their interest vectors (denoted by S, which is incremented by each matching entry); ii) their asking and answering interaction frequency (denoted by A); and iii) the number of their common friends, denoted by C. Given the asker's friend set $\mathcal{F}$, friend U_k's rates of S, A and C are calculated by:

$$P_{S_{U_k}} = \frac{S_{U_k}}{\sum_{i \in \mathcal{F}} S_i},\ P_{A_{U_k}} = \frac{A_{U_k}}{\sum_{i \in \mathcal{F}} A_i},\ P_{C_{U_k}} = \frac{C_{U_k}}{\sum_{i \in \mathcal{F}} C_i}. \quad (2)$$

Then, the social closeness of friend U_k is calculated as

$$\Psi_{C,U_k} = \gamma_S * P_{S_{U_k}} + \gamma_A * P_{A_{U_k}} + \gamma_P * P_{C_{U_k}}, \quad (3)$$

where γ_S, γ_A and γ_P are the weights of considering factors S, A and C, respectively. Since we make all metrics comparable by scaling them to $[0, 1]$, the weights represent the correlationship between each factor and the social closeness. We finally calculate the metric Ψ_{U_k} to measure the appropriateness of friend U_k as a potential answerer for U_j's question Q_i. That is:

$$\Psi_{U_k} = \beta * \Psi_{I,U_k} + (1-\beta) * \Psi_{C,U_k}\ (0 < \beta < 1), \quad (4)$$

where β is the weight for each consideration factor. In different circumstances, we can give different β values. Higher β value helps identify friends with higher capability to answer the question, while a lower β value helps identify friends with higher willingness to answer the question.

Algorithm 2 Pseudocode for the *Question-User Mapper*.

Input: Interest vectors of a user, his/her friends and question
Output: A list of potential answer providers

```
for each friend U_k in the friend set of U_j do
    Compute Ψ_{I,U_k} based on Eq. (1)
    Compute P_{S_{U_k}}, P_{A_{U_k}} and P_{C_{U_k}} based on Eq. (2)
    Compute Ψ_{C,U_k} based on Eq. (3)
    Compute Ψ_{U_k} based on Eq. (4)
end for
Order the friends in descending order of Ψ_{U_k}
Notify the top N friends
```

SocialQ&A then orders an asker's friends in the descending order of their Ψ_{U_k} values, and routes the question to the top N friends. N is a tradeoff between system overhead and response efficiency. If N is larger, the system overhead is larger, but the answer response efficiency is improved; and vice versa. Algorithm 2 shows the pseudocode of the *Question-User Mapper*. If no one responds during a specific time period, SocialQ&A can try the nodes in 2-hop social distance from the asker, and then in 3-hop social distance, until the nodes in Time-To-Live (TTL)-hop social distance have attempted. A question receiver can forward the question if (s)he cannot answer it. The question-user mapper algorithm is called while asking or forwarding questions. When forwarding a question, the asker's information is replaced by the forwarder's information. The *Question-User Mapper* can be executed in either a centralized manner or a decentralized manner [30]. In the centralized manner, the centralized server selects the potential answerers for each question and sends the question to them. In the decentralized execution, each node autonomously determines the potential answerers for the question initialized or received by itself to send the question. If there are not enough N selected friends through the *Question-User Mapper*, the remaining answerers are randomly selected from all users having such interests.

4) User Interface: SocialQ&A allows users to register and modify user information, add/remove friends, ask/answer/forward questions and check question notifications. Consider a hypothetical user named Mike. When Mike registers, he is required to provide essential information about himself, such as his personal information, area of study/expertise, his current interests, and his involvement in other activities. Users are also encouraged to describe their interests in terms of a few predefined categories, such as movies, books, television, music. *User Interest Analyzer* uses the registration information to determine Mike's interests.

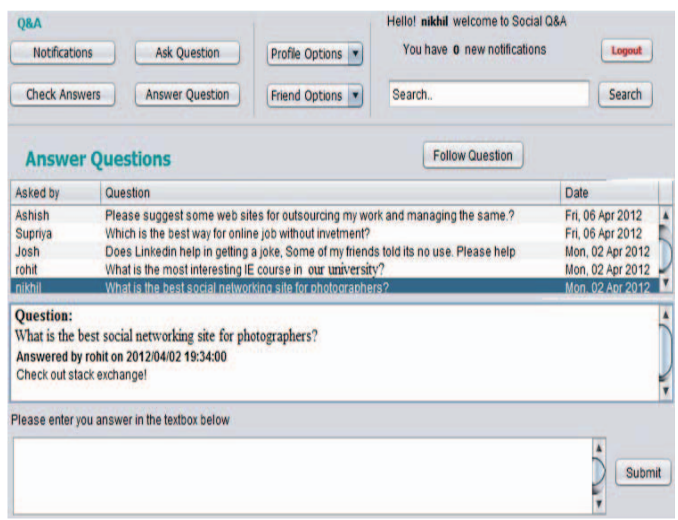
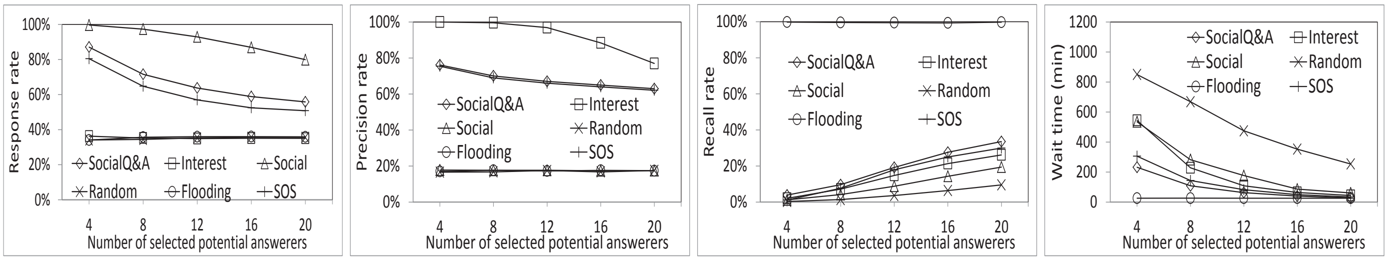


Fig. 3: Interface of a question and answer thread.

As shown in Figure 3, the main user interface includes a social platform (including Profile and Friend Options), a Q&A domain (including Ask Question, Answer Questions and Check Answers), and a notification module (including events from both social platform and Q&A domain). Whenever there is a user profile or friendship change of Mike, the interest similarity S and friend closeness C are updated accordingly by *Question User Mapper*.

When Mike submits questions, the user interface contacts *Question Categorizer* first, and then passes the question interests to *Question User Mapper*, which finally decides the top N answerers and sends out the questions. Another feature provided by SocialQ&A is the option to forward and follow questions so that the forwarder will also receive answers from



(a) Response rate (b) Precession rate (c) Recall rate (d) Wait time

Fig. 4: Effectiveness of Q&A systems.

the answerers. Further, SocialQ&A provides search function to enable users to search in the repository of previous posted questions and answers.

Unlike previous Q&A approaches, SocialQ&A exploits the users' profile information and interests, in addition to the user's social network and Q&A activities to determine potential answer providers. Additionally, the interest information of all users in the system is continuously updated based on their actions. SocialQ&A is also distinguished by routing questions only to potential answer providers rather than flooding to all friends, thereby reducing overhead and frustration to users.

IV. Simulation Performance Evaluation

We conducted trace-driven experiments on the PlanetLab [33] to evaluate the performance of SocialQ&A in both searching effectiveness and efficiency for potential answerers. PlanetLab is a real worldwide testbed, consisting of 1295 nodes, on which you can deploy and test your network applications. We used the *Yahoo! Answers* question/answer trace data and Facebook user profile (interests and activities) trace from [30]. The Yahoo! Answers trace includes 9419 questions posted in the Entertainment & Music Movies section, and the Facebook trace includes 1000 users with profiles and friendship information. We tested 1000 users using randomly selected 200 PlanetLab nodes worldwide, each simulating 5 users. We randomly assigned users' profile and associated relationship in the trace to simulated users. From the question trace, we randomly selected 100 questions with keywords which can be mapped to the category of movies. We also used this set of questions to simulate those in each of other three interest categories. All questions were randomly assigned to users having the same interest. The TTL was set to 3, since according to [34], people are influenced by other people who are at most at 3-hop social distance. The rating range in *Yahoo! Answers* is [1, 5].

A successful response to a question includes answering or forwarding the question, in which if a question receiver has an answer to the received question, (s)he replies to it; otherwise, (s)he forwards the question. Intuitively, each potential answerer willing to answer the question should have at least one very high score for S, A or P in Equation 3. Thus, we give equal weights to all factors as $\gamma_S = \gamma_A = \gamma_P = 1$. The social closeness between friends ranges in [0,1.2]; if an asker's friend has social closeness larger than 0.6, (s)he is willing to respond to the asker's question. If we set this willingness threshold to be larger, there will be fewer successful responses in both our method and comparison methods, and vice versa. The probability that other friends respond to the question was randomly chosen from $\{10\%, 20\%, 30\%\}$. The question query rate was set to one question per minute. These parameters are adjustable parameters and their changes will not affect the relative performance differences between the systems in comparison. The distribution of response time to a question follows the trace [30]. We use BA to denote the Best Answer set of a question existing in the system, and use RA to represent the Retrieved Answer set in the system. We define the *precision rate* as $|RA \bigcap BA|/|RA|$ to represent the received answers' quality, and define the *recall rate* as $|BA \bigcap RA|/|BA|$ to denote the received answers' completeness.

Recall that SocialQ&A considers both interest similarity and social closeness. We compare SocialQ&A with i) SocialQ&A only considering interest similarity (denoted by Interest), ii) SocialQ&A only considering social closeness (denoted by Social), iii) random friend selection (denoted by Random), iv) flooding method selecting all friends [30] (denoted by Flooding), and v) SOS [30]. These systems were implemented in a distributed manner; that is, each node selects its friends to send/forward questions autonomously. Unless otherwise specified, β in Eq. (4) was set to 0.6.

A. Performance with Varying Number of Selected Answerers

We calculated the response rate as the number of all successful responses divided by the total number of question receivers. Figure 4(a) shows that the response rate follows Social>SocialQ&A>SOS>Interest≈Random≈Flooding. In SocialQ&A and Social, users choose friends with higher social closeness who are most willing to answer questions, so they have a higher response rate than others. SOS does not consider the potential willingness of friends with many common interests when calculating social closeness. Thus, its response rate is lower than SocialQ&A and Social, but higher than the other three methods without social closeness consideration. In SocialQ&A, users may choose friends with high interest similarity but lower social closeness. Thus, it generates a lower response rate than Social. We also see that the response rate of SocialQ&A, Social and SOS decreases as the number of selected answerers increases, since friends with lower social closeness are more likely to drop questions. This result implies that SocialQ&A's incentive works well when the set of answerers selected is small.

Figure 4(b) shows the average precision rate of each system, which follows Interest>SocialQ&A≈SOS>Social ≈Random≈Flooding. This is because Interest, SocialQ&A and SOS choose answerers with interest consideration, while Social, Random and Flood do not. By considering interest and social closeness simultaneously, SocialQ&A and SOS have lower precision rate than Interest, and their precision rates decrease as the number of selected answerers increases

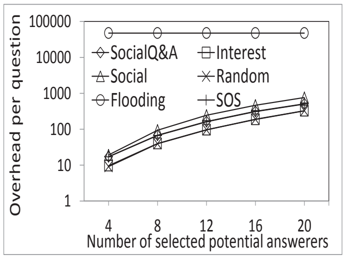
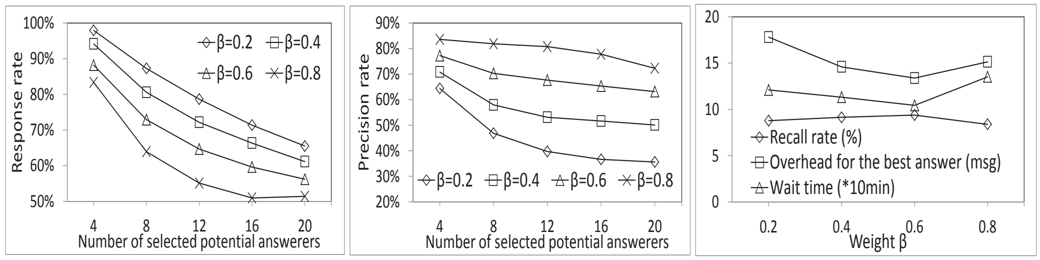


Fig. 5: System overhead.



(a) Response rate (b) Precision rate (c) Effectiveness

Fig. 6: Performance vs. varying β values.

due to the same reason as in Figure 4(a). SocialQ&A and SOS both consider the interest similarity, so they produce similar precision rates. This implies that they have higher answer quality when the number of selected neighbors is small. Combining the results in Figures 4(a) and 4(b), we see that SocialQ&A performs the best regarding both response rate and answer quality.

Figure 4(c) shows the average recall rate of each system. We see that the recall rates of all systems except Flooding increase when the number of selected potential answerers increases. As more potential answerers are selected, more answers are provided, which increases the probability of receiving the best answers. We also see that the recall rate follows Flooding>SocialQ&A>SOS>Interest>Social >Random. Flooding sends a question to all friends, thus produces the highest recall rate. Since SocialQ&A and SOS consider both interest and willingness to respond, they produce much more high-quality answers than in other systems except Flooding. SocialQ&A has a higher recall rate than SOS due to its higher response rate than SOS as shown in Figure 4(a). We see that Interest has a larger recall rate than Social, especially when the number of selected neighbors is large. This is because Social ensures high willingness to respond but cannot guarantee the answer quality, while Interest provides high answer quality. As Random does not consider interest and social closeness, it generates the lowest recall rate. This figure indicates that SocialQ&A can recall more answers than other systems without flooding questions.

We define the *wait time* of a question as the time interval between the time when a question is asked and the time when the first best answer of this question is received. Figure 4(d) shows the average wait time for all questions. It follows Flooding<SocialQ&A<SOS<Interest<Social<Random and all methods have shorter response time when there are more answerers selected due to the same reason as in Figure 4(c). This figure indicates that SocialQ&A leads to shorter wait time for answers than other methods except Flooding. However, Flooding generates low precision rate and also high overhead which will be shown in Figure 5.

We define the *overhead* of a question as the number of forwarding messages generated for the question. Figure 5 shows the overhead per question. It follows Social>SocialQ&A> SOS>Interest≈Random due to the same reason as in Figure 4(a). A higher response rate leads to more forwarding messages. SOS, Interest and Random produce lower overhead because many message receivers do not respond. Flooding generates the highest overhead which is nearly constant because of its broadcasting feature. We also see that the overhead increases as the number of selected answerers increases in other methods as more queries are sent or forwarded.This figure indicates that by more accurately routing a question to users that are capable and willing to answer the question, SocialQ&A generates relatively low overhead and high response rate.

B. Performance with Varying Weight Values

Figure 6(a) shows the response rate of different β values versus different numbers of selected answerers. It shows that the response rate decreases when β increases. Because a larger β gives less weight on the social closeness, which reflects the users' willingness to respond to a question. Figure 6(b) shows the precision rate of different β values versus different numbers of selected answerers. It shows that the precision rate increases as β increases because a larger β gives more weight on the interest similarity, which reflects users' capability to answer questions. The results show that β controls the tradeoff between the response rate and precision rate.

We define *overhead for the best answer* as the the number of forwarding message generated for a question before the first best answer is received. Figure 6(c) shows the recall rate, wait time and overhead for the best answer when 8 potential answerers were selected in SocialQ&A. As β increases, the wait time and overhead first decrease then increase, and the recall rate first increases and then decreases at the point of $\beta = 0.6$. This is caused by the increasing weight of interest closeness and decreasing weight of social closeness. This figure shows that when β equals 0.6, SocialQ&A has the shortest wait time, largest recall rate and lowest overhead, thus reaching the optimal tradeoff between efficiency, effectiveness and cost.

V. Real-World Experimental Result Analysis

Section IV shows the outperformance of SocialQ&A compared with other systems by measuring its potential answerer searching effectiveness and efficiency. In this section, we show the performance of SocialQ&A under real environment. This section presents analysis on the usage of our deployed real-world SocialQ&A system over a period of approximately one month starting at the beginning of March, 2012. Since it is not a mature and commercial software, and it is hard to attract users to use it, we call for volunteers to use it within this month. A total of 124 users registered and used SocialQ&A, 163 questions were posted and 282 answers were posted in response. The distribution of the users. Approximately 35 users were from the United States, 70 users were from India, and 1 user was from the United Kingdom. This small-scale testing shows the potential benefits in both the searching effectiveness and efficiency and interesting usages of SocialQ&A to a certain degree. Evaluation on a large-scale user base remains as our future work.

Why do people like bass low frequencies in music?	I want to start photography can anyone suggest me a good camera?
What are the effects of music on plants?	What is a good car renting service in the washington DC area?
Please can anybody suggest me the book for multimedia Systems?	What is better programming language PHP or python?
What is RTP?	skype or yahoo which is better....
Best movies of 2012...	How do I make my playlist private on youtube?
Best character in The Lord Of the rings...???	What is a computer virus?
What is the best comedy show on TV right now?	What is our purpose of living life?
Who is the best actor in big bang	Why am I so happy today?

Fig. 7: A sample of questions from SocialQ&A.

A. *Analysis of Questions*

Recall that SocialQ&A uses four major categories: music, books, movies, and television. The left column in Figure 7 shows two example questions for each category. In this section, we further classified the questions based on the following four question types: Recommendation, opinion, factual and rhetorical. The right column in Figure 7 shows two example questions for each question type, respectively. Among all questions, we see that the users asked a large number of opinion-type questions. Approximately, 20% of the questions were recommendation type questions, 36% were opinion-type questions, 25% were factual-type questions, and 19% were rhetorical-type questions.

Figure 8(a) shows the number of answers posted for each of the questions with at least one response. Figure 8(b) shows the distribution of the number of responses for these questions. Approximately, 47% of questions have just 1 answer, and 13% of questions have more than 4 answers. Most of the questions receiving only one answer are factual questions, since one answer is sufficed for such questions. Opinion-type questions tend to have more responses, as no answer is the final answer. For example, in the test, question “Should I buy a Windows laptop or MacBook?” received more answers than question “What is the capital of Oregon?”.

Potential benefit: SocialQ&A provides a platform for both factual and non-factual questioning, and there tend to be more answers for opinion based questions from social friends, which can be a better reference for the askers on non-factual questions. This feature of social network based Q&A systems is also indicated in [1, 2].

B. *Quality of Answers*

For every question asked, the asker was able to rate the answer on a scale of 1 to 10. Out of 282 answers posted, the users of SocialQ&A rated 233 answers. A single question may have multiple answers; hence, we calculated the average rating for each question and present the results in Figure 9(a). The average rating of all answers is 8.675, ignoring those that were not rated. The median is 9.29, the minimum is 1, and the maximum is 10. The result means that most answers provided in this test received high ratings.

We also analyzed the correlation between the question length and the question rating. Intuitively, long questions tend to be easier to understand. Thus, long questions help the answer provider determine what the asker is looking for, enabling him/her to provide a more accurate answer. Any question that was explained using more than one sentence is considered a long question. Our results show that longer questions have an average rating of 9.33, which is higher than the overall average rating.

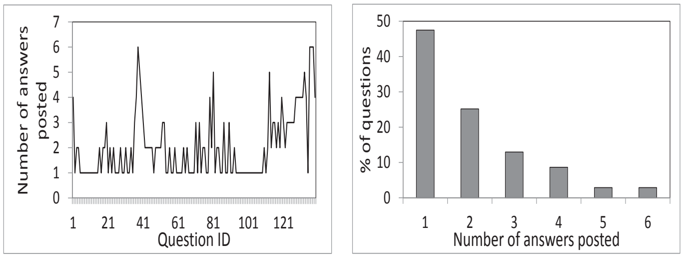
Another way to examine the answer quality is to find the maximum rating that an answer received for a particular question. The analysis of the maximum rating is meaningful because the highest rated answer provides the asker with the desired information and the other answers could be neglected. Figure 9(b) plots the percent of questions versus the maximum rating of each question. The average maximum rating over all questions is 9.05, the median is 10, the minimum is 1, and the maximum is 10. The results indicate that SocialQ&A provides satisfactory answers in most cases in this test.

The high answer ratings in SocialQ&A may be attributed to two factors: (1) since the answerer belongs to the asker’s immediate social network, (s)he is highly motivated to provide better quality answers, and (2) the question is mapped to the potential answer provider whose interests most closely match the topics of the question. The result of this analysis verifies the advantages of SocialQ&A by leveraging the previous studies [10–13] on the influence of social networks on Q&A performance to effectively identify potential answer providers that can provide high-quality answers. We expect that the answer quality would be further improved as more users join SocialQ&A, because more users will be willing to respond and the probability that an expert exists among users also increases.

Potential benefit: SocialQ&A can enhance the degree of satisfaction of askers on the answers especially for non-factual questions since answerers share interests with askers and are motivated to answer their questions, as indicated in [2, 10, 12, 13].

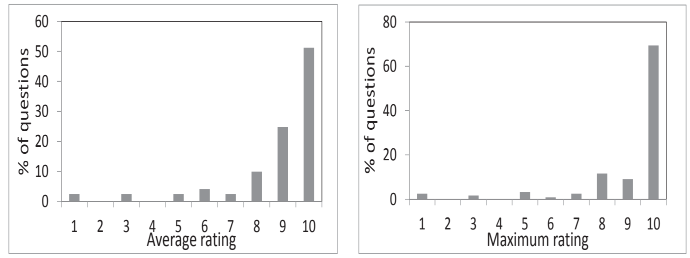
VI. Conclusion

Q&A systems are used by many people for purposes such as information retrieval, academic assistance, and discussion. To increase the quality of answers received and decrease the wait time for answers, we have developed and prototyped an online social network based Q&A system, called SocialQ&A. It utilizes the properties of a social network to forward a question to potential answer providers, ensuring that a given question receives a high-quality answer in a short period of time. It removes the burden from answer providers by directly delivering them the questions they might be interested in, as opposed to requiring answer providers to search through a large collection of questions as in *Yahoo! Answers* or flooding a question to all of an asker’s friends in an online social network. Our comprehensive trace-driven experiments and analysis results on the real-world Q&A activities from the SocialQ&A prototype show the promises of SocialQ&A to enhance answer quality and reduce answer wait time in current Q&A systems. In our future work, we will conduct tests on a large user base in the real-world experiment.



(a) # of answers received by each question (b) % of questions vs. # of answers received

Fig. 8: Analysis on the # of received answers.



(a) % of questions vs. the average rating (b) % of questions vs. the maximum rating

Fig. 9: Analysis on the answer rating.

SocialQ&A is mainly designed to answer a user's questions in his/her interests. However, a user may sometimes ask questions outside of his/her interests. We will explore a method to handle this case. We will also study how to protect the privacy of questions as requested by the askers.

ACKNOWLEDGEMENTS

This research was supported in part by U.S. NSF grants NSF-1404981, IIS-1354123, CNS-1254006, CNS-1249603, and Microsoft Research Faculty Fellowship 8300751.

REFERENCES

[1] M. R. Morris, J. Teevan, and K. Panovich. A Comparison of Information Seeking Using Search Engines and Social Networks. In *In Proc. of ICWSM*, 2010.

[2] M. R. Morris, J. Teevan, and K. Panovich. What do People Ask Their Social Networks, and Why?: A Survey Study of Status Message Q&A Behavior. In *Proc. of CHI*, 2010.

[3] Z. Gyongyi, G. Koutrika, J. Pedersen, and H. Garcia-Molina. Questioning Yahoo! Answers. In *Proc. of QAWeb*, 2008.

[4] Yahoo!Answers Team. Yahoo! Answers BLOG. http://yahooanswers.tumblr.com/, [Accessed on 10/20/2014].

[5] B. Li and I. King. Routing Questions to Appropriate Answerers in Community Question Answering Services. In *Proc. of CIKM*, 2010.

[6] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. In *Proc. of WWW*, 2008.

[7] A. Mtibaa, M. May, C. Diot, and M. Ammar. Peoplerank: Social Opportunistic Forwarding. In *Proc. of Infocom*, 2010.

[8] E. Pennisi. How Did Cooperative Behavior Evolve? *Science*, 2005.

[9] M. L. Radford, C. Shah, L. Mon, and R. Gazan. Stepping Stones to Synergy: Social Q&A and Virtual Reference. *Proceedings of the American Society for Information Science and Technology*, 2011.

[10] M. Richardson and R. White. Supporting Synchronous Social Q&A Throughout the Question Lifecycle. In *Proc. of WWW*, 2011.

[11] R. W. White, M. Richardson, and Y Liu. Effects of Community Size and Contact Rate in Synchronous Social Q&A. In *Proc. of SIGCHI*, 2011.

[12] J. Teevan, M.R. Morris, and K Panovich. Factors Affecting Response Quantity, Quality, and Speed for Questions Asked via Social Network Status Messages. In *Proc. of ICWSM*, 2011.

[13] Z. Li and H. Shen. Collective Intelligence in the Online Social Network of Yahoo!Answers and Its Implications. In *Proc. of CIKM*, 2012.

[14] J. Zhang, M. S. Ackerman, and L. Adamic. Expertise Networks in Online Communities: Structure and Algorithms. In *Proc. of WWW*, 2007.

[15] J. Bian, Y. Liu, D. Zhou, E. Agichtein, and H. Zha. Learning to Recognize Reliable Users and Content in Social Media with Coupled Mutual Reinforcement. In *Proc. of WWW*, 2009.

[16] P. Jurczyk and E. Agichtein. Discovering Authorities in Question Answer Communities by Using Link Analysis. In *Proc. of CIKM*, 2007.

[17] M. Bouguessa, B. Dumoulin, and S. Wang. Identifying Authoritative Actors in Question-Answering Forums: the Case of Yahoo!Answers. In *Proc. of KDD*, 2008.

[18] L. Hong, Z. Yang, and B. D. Davison. Incorporating Participant Reputation in Community-Driven Question Answering Systems. In *Proc. of CSE*, 2009.

[19] W. Chen, Q. Zeng, W. Liu, and T. Hao. A User Reputation Model for a User-Interactive Question Answering System: Research Articles. *Concurrency and Computation: Practice and Experience*, 2007.

[20] Y. R. Tausczik and J. W. Pennebaker. Predicting the Perceived Quality of Online Mathematics Contributions from Users' Reputations. In *Proc. of SIGCHI*, 2011.

[21] A. Shtok, G. Dror, Y. Maarek, and I. Szpektor. Learning From the Past: Answering New Questions With Past Answers. In *Proc. of WWW*, 2012.

[22] X. Quan, W. Liu, and B. Qiu. Term Weighting Schemes for Question Categorization. *TPAMI*, 2011.

[23] W. Song, W. Liu, N. Gu, X. Quan, and T. Hao. Automatic Categorization of Questions for User-Interactive Question Answering. *Information Processing and Management*, 2011.

[24] B. Li, I. King, and M. R. Lyu. Question Routing in Community Question Answering: Putting Category in its Place. In *Proc. of CIKM*, 2011.

[25] T. C. Zhou, M. R. Lyu, and I. King. A Classification-Based Approach to Question Routing in Community Question Answering. In *Proc. of WWW (Companion Volume)*, 2012.

[26] X. Cao, G. Cong, B. Cui, C. S. Jensen, and C. Zhang. The Use of Categorization Information in Language Models for Question Retrieval. In *Proc. of CIKM*, 2009.

[27] J. Guo, S. Xu, S. Bao, and Y. Yu. Tapping on the Potential of Q&A Community by Recommending Answer Providers. In *Proc. of CIKM*, 2008.

[28] B. M. Evans and E. H. Chi. Towards a Model of Understanding Social Search. In *Proc. of CSCW*, 2008.

[29] D. Horowitz and S. D. Kamvar. The Anatomy of a Large-Scale Social Search Engine. In *Proc. of WWW*, 2010.

[30] Z. Li, H. Shen, G. Liu, and J. Li. SOS: A Distributed Context-Aware Question Answering System Based on Social Networks. In *Proc. of ICDCS*, 2012.

[31] M. Mcpherson. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 2001.

[32] G. A. Miller. WordNet: A Lexical Database for English. *Commun. ACM*, 1995.

[33] Planetlab. http://www.planet-lab.org/.

[34] N. A. Christakis and J. H. Fowler. *Connected: The surprising power of our social networks and how they shape our lives*. Hachette Digital, 2009.