Wisdom in the Social Crowd: an Analysis of Quora

Gang Wang, Konark Gill, Manish Mohanlal, Haitao Zheng and Ben Y. Zhao
Computer Science, UC Santa Barbara, Santa Barbara, CA 93106, USA
{gangw, konarkgill, manish, htzheng, ravenben}@cs.ucsb.edu

ABSTRACT

Efforts such as Wikipedia have shown the ability of user communities to collect, organize and curate information on the Internet. Recently, a number of question and answer (Q&A) sites have successfully built large growing knowledge repositories, each driven by a wide range of questions and answers from its users community. While sites like Yahoo Answers have stalled and begun to shrink, one site still going strong is Quora, a rapidly growing service that augments a regular Q&A system with social links between users. Despite its success, however, little is known about what drives Quora's growth, and how it continues to connect visitors and experts to the right questions as it grows.

In this paper, we present results of a detailed analysis of Quora using measurements. We shed light on the impact of three different connection networks (or graphs) inside Quora, a graph connecting topics to users, a social graph connecting users, and a graph connecting related questions. Our results show that heterogeneity in the user and question graphs are significant contributors to the quality of Quora's knowledge base. One drives the attention and activity of users, and the other directs them to a small set of popular and interesting questions.

Categories and Subject Descriptors

H.3.5 [**Information Storage and Retrieval**]: Online Information Services-Web-based services; J.4 [**Computer Applications**]: Social and Behavioral Sciences

General Terms

Measurement, Management, Design

Keywords

Q&A System, Online Social Networks, Graphs

1. INTRODUCTION

The Internet is a maelstrom of information, most of it real, and much of it false. Efforts such as Wikipedia have shown that collectively, Internet users possess much knowledge on a wide range of subjects, knowledge that can be collated and curated to form valuable information repositories. In the last few years, community question-and-answer (Q&A) sites have provided a new way for users to crowdsource the search for specific detailed information, much of which involves getting first-hand answers of specific questions from domain experts.

While these sites have exploded in popularity, their growth has come at a cost. For example, the first and still largest of these sites, Yahoo Answers, is showing clear signs of stalling user growth and stagnation, with traffic dropping 23% in a span of four months in 2011 [23]. In addition, the Google Answers service launched in 2001 was already shut down by 2006. Why is this the case? One of the prevailing opinions is that as sites grow, a vast number of low-value questions overwhelm the system and make it extremely difficult for users to find useful or interesting content. For example, ridiculous questions and answers are so prevalent on Yahoo Answers that a quick Google search for “Yahoo Answers Fail” turns up more than 8 million results, most of which are sites or blogs dedicated to documenting them.

Bucking the trend thus far is Quora, an innovative Q&A site with a rapidly growing user community that differs from its competitors by integrating a social network into its basic structure. Various estimates of user growth include numbers such as 150% growth in one month, and nearly 900% growth in one year [23]. Despite its short history (Quora exited beta status in January 2010), Quora seems to have achieved where its competitors have failed, *i.e.* successfully drawing the participation of both a rapidly growing user population and specific domain experts that generate invaluable content in response to questions. For example, founders of Instagram and Yelp answered questions about their companies, Stephen Fry and Ashton Kutcher answered questions about actors, and domain-specific answers come from experts such as Navy Seals sharpshooters and San Quentin inmates.

So how does Quora succeed in directing the attention of its users to the appropriate content, either to questions they are uniquely qualified to answer, or to entertaining or informative answers of interest? This is a difficult question to answer, given Quora's own lack of transparency on its inner workings. While it is public knowledge that Quora differs from its competitors in its use of social networks and real identities, few additional details or quantitative measures are known about its operations. A simple search on Quora about how it works produces numerous unanswered questions about Quora's size, mechanisms, algorithms, and user behavior.

In this paper, we perform a detailed measurement study of Quora, and use our analyses to shed light on how its internal structures contribute to its success. To highlight key results, we use comparisons against Stack Overflow, a popular Q&A site without an integrated social network. We seek to answer several key questions:

- What role do traditional question topics play in focusing user attention? How much do followers of a topic contribute to answering its questions?

WWW 2013, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2035-1/13/05.

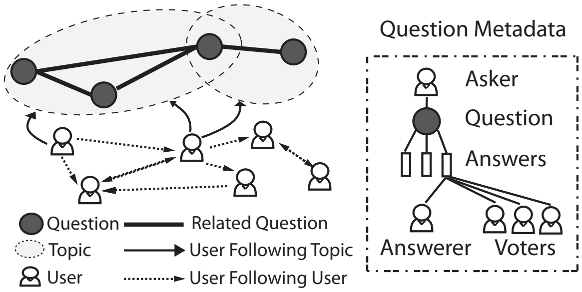


Figure 1: Structure of questions, topics and users in Quora.

- What impact do super users have on specific patterns of user activity? Can they generate and focus user attention on individual questions, thus setting them apart from questions on related topics?
- Given the rapid growth of questions on question-and-answer sites, how does Quora help users find the most interesting and valuable questions and avoid spammy or low-value questions? What role do the "related questions" feature play?

Our analysis reveals interesting details about the operations of Quora. We find that while traditional topic-followers generate traffic, social relationships help bring a significant amount of answers to questions, and generally provide much higher quality answers than strangers. Most surprisingly, we find that the related-question graph plays a significant role in funneling user traffic and attention to key questions on a given topic. It exhibits a power-law structure where the degree of a question correlates strongly with the number of views and answers it receives. Finally, we use graph partitioning to identify clusters of related questions in the graph, and find that the large majority of viewers and answers focus on very few questions within each cluster. This further supports our hypothesis that Quora's social graph and question graph have been extremely effective at focusing user attention and input on a small subset of valuable questions.

2. BACKGROUND

Quora is a question and answer site with a fully integrated social network connecting its users. In this section, we introduce Quora, using Stack Overflow as a basis for comparison. We then give details on the key Quora graph structures that connect different components together. Specifically, we describe three types of graphs in Quora: a social graph connecting users, a user-topic following graph and a related question graph.

2.1 Quora and Stack Overflow

Quora. Quora is a question and answer site where users can ask and answer questions and comment on or vote for existing answers. Unlike other Q&A sites where all users exist in a global search space, Quora allows users to follow each other to form a social network. Social connections in Quora are directional like Twitter. A user A can follow user B without explicit permission, and B's actions (new questions, answers, comments and topics) will appear in A's activity stream. We say A is B's *follower* and B is A's *followee*. In addition, users can follow *topics* they are interested in, and receive updates on questions and answers under this topic.

Each Quora user has a *profile* that displays her bio information, previous questions and answers, followed topics, and social connections (followers and followees). Each user has a "Top Stories" page, which displays updates on recent activities and participated questions of their friends (followees), as well as recent questions under the topic they followed. A small subset of registered users are chosen by Quora to be *reviewers* and *admins*, and have the power to flag or remove low quality answers and questions.

Finally, each Quora question has its own page, which includes a list of its answers and a list of related questions. Users can add new answers, and comment, edit and vote on existing answers.

Website	Data Since	Total Questions	Total Topics	Total Users	Total Answers
Stack Overflow	Jul. 2008	3.45M	22K	1.3M	6.86M
Quora	Oct. 2009	437K	56K	264K	979K

Table 1: Data Summary.

Stack Overflow. Stack Overflow is another successful Q&A site started in 2008. Stack Overflow differs from Quora in two main aspects. First, while Quora covers a broad range of general topics, Stack Overflow focuses specifically on computer programming questions. Second, users in Stack Overflow are fully independent and no social connections exist between users.

2.2 Graph Structures In Quora

The internal structure of question-and-answer sites are often a complex mix of questions, answers, question topics, and users. We summarize the relationships between different entities in Figure 1. Users can follow individual topics and other users for news and events; questions are connected to other "related" questions, and each question can be tagged with multiple topics. Finally, for each question in the system, there is a user who asked that question (the *asker*), users who answered that question (*answerers*), and users who voted on an answer (*voters*).

Quora's internal structure is dominated by three graphs that act as channels that guide user interest and deliver information to users.

1. *User-Topic Graph:* Quora users follow different topics, and receive updates about questions under topics they follow.
2. *Social Graph:* Quora users follow each other to form a Twitter-like social graph. Users receive newsfeed about questions their friends participated in.
3. *Question Graph:* Each question has a list of related questions used by users to browse related questions. The "related" relationship is considered symmetric.

We believe these three graphs are largely responsible for guiding the attention of Quora users. In this paper, we will perform detailed analysis on these graphs to understand how they impact user activities, especially how they help users separate a small subset of interesting questions from the larger number of less interesting questions/answers.

3. DATASET AND PRELIMINARY RESULTS

Before diving into main analytical results of our work, we begin in this section by first describing our data gathering methodology and presenting some preliminary results. Here we describe the properties and limitations of our Quora and Stack Overflow datasets. We also analyze some high level metrics of the Quora data, while using Stack Overflow as a baseline for comparison.

3.1 Data Collection

Our analysis relies on two key datasets. A publicly available dataset periodically released by Stack Overflow, and a dataset crawled

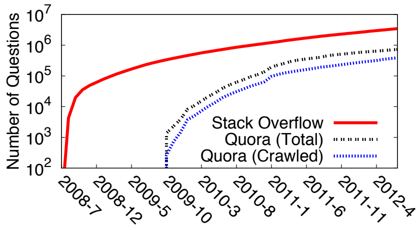


Figure 2: Questions growth.

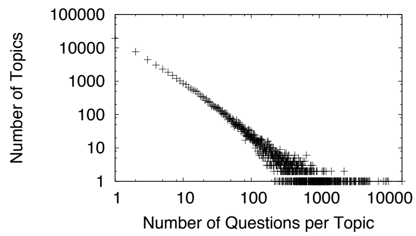


Figure 3: # of Questions per topic.

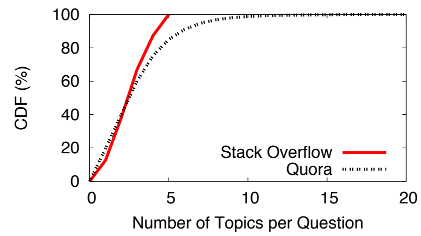


Figure 4: # of Topics per question.

from Quora that contains multiple groups of data on users, questions, topics and votes. We describe details below. The basic statistics of both datasets are shown in Table 1.

Stack Overflow. Stack Overflow periodically releases all of their data to the public. Our site trace was released in August 2012, and covers all activity on Stack Overflow between July 2008 and July 2012.

Quora. We gathered our Quora dataset through web-based crawls between August and early September 2012. We tried to follow crawler-etiquette defined in Quora's `robots.txt`. Limited portions of data were embedded in Ajax calls. We used FireWatir, an open-source Ruby library, to control a FireFox browser object, simulating clicking and scrolling operations to load the full page. We limited these crawls to 10 requests/second to minimize impact on Quora.

Since Quora has no predefined topic structures for its questions (questions can have one or more arbitrary topic "labels"), getting the full set of all questions is difficult. We followed the advice from a Quora data scientist [3] and start our question crawls using 120 randomly selected questions roughly evenly distributed over 19 of the most popular question topics. The crawls follow a BFS pattern through the related questions links for each question. In total, we obtained 437,000+ unique questions. Each question page contains the topics associated to the question, a complete list of answers, and the answerers and voters on each answer. As shown in Table 1, this question-based crawl produced 56,000+ unique topics, 979,000+ answers, and 264,000+ unique users who either asked or answered a question, or voted on an answer.

Our biggest challenge is trying to understand how much of the Quora dataset we were able to gather. The simple answer is we don't know, since there are no official quantitative measures about Quora available. But we found a post by a Quora reviewer [2] that hinted the question ID (or *qid*) in Quora is sequentially assigned. Thus we can infer the total number of questions by inspecting the qid of the newly added questions. To validate this statement, we performed several small experiments where we added small bursts of new (meaningful) questions to Quora. Each burst contains 10 new questions sent seconds apart, and consistently produced 10 sequential qid's. We separated experiments by at least 30 minutes, and observed increments to the qid consistent with the expected number of new questions in the gap between experiments. Finally, we plotted qid values for all questions found by our crawl and correlated them with the estimated date of question creation. The result, discussed below, provides further support that this qid can be used as an estimate of total questions in the system. The largest qid from our crawled questions is 761030, leading us to estimate that Quora had roughly 760K questions at the time of our crawl, and our crawl covered roughly 58% of all questions. Note that not all questions remain on the site, as Quora actively deletes spam and redundant questions [5]. This estimate might provide an upper bound

Topic in Quora	# of Questions	Topic in Stack Overflow	# of Questions
Startups	16.3K	C#	333K
Survey Questions	10.3K	Java	277K
Movies	9.7K	PHP	257K
Medicine / Healthcare	9.3K	Javascript	242K
Food	8.7K	Android	211K
Facebook	7.4K	jquery	207K
Music	5.5K	iPhone	143K
Google	5.4K	C++	139K
Psychology	5.2K	ASP.net	132K
Startup Advice	5.2K	.net	125K

Table 2: Top 10 topics based on number of questions.

of actual number of questions, and our coverage of 58% would be a lower bound.

We also crawled the user profiles for users extracted from the crawled questions. Each user profile contains 6 parts: the list of the user's followers, list of users they follow (followees), their previous answers, their previous questions, their followed topics and boards. Out of the 264K extracted users, we found that roughly 5000 (1.9%) profiles were no longer available, likely deleted either by Quora or the user.

***Qid* Over Time.** Assuming we are correct about the use of qid, we can plot an estimate of the growth of Quora (and Stack Overflow), by plotting qid against time. Since Quora does not show when a question is posted, we estimate the posting time by the timestamp of its earliest answer. For open questions with no answer, we infer the question posting time based on the latest activity timestamp on the question page. Since reading the question does not update this "latest activity" timestamp, this timestamp can estimate posting time for unanswered questions. We estimate the total number of questions in Quora for each month by looking at the largest *qid* of questions posted in that month. For Stack Overflow, we use the timestamp for questions creation in the data trace.

We see in Figure 2 that Stack Overflow is an older site with more questions than Quora. We plot two lines for Quora, a black dashed line for the total number of questions estimated by *qid*, and the blue dashed line is the number of questions we crawled from each month. Both lines increase smoothly without gaps, suggesting that Quora did not reset *qid* in the past and the questions we crawled are not biased to a certain time period. Our estimated number of questions in Quora for June 2012 is 700K, which is consistent with previously reported estimates [24]. As Quora continues to grow, it is clear that helping users easily identify and find the most meaningful and valuable questions and answers is a growing challenge.

3.2 Initial Analysis

Topics. Quora is a general Q&A site with a very broad range of topics. We observed 56K topics in our dataset, which is twice more than that of Stack Overflow, even though Quora is smaller by

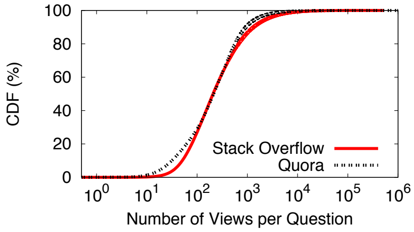


Figure 5: # of User views per question.

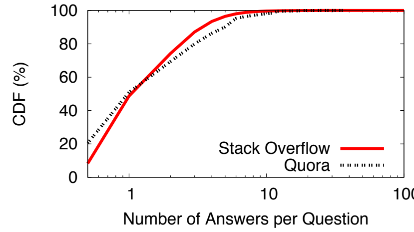


Figure 6: # of Answers per question.

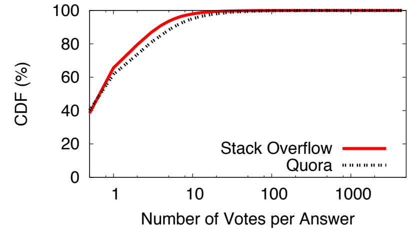


Figure 7: # of Votes per answer.

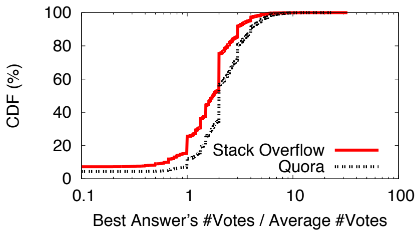


Figure 8: Votes for the best answer vs. the average.

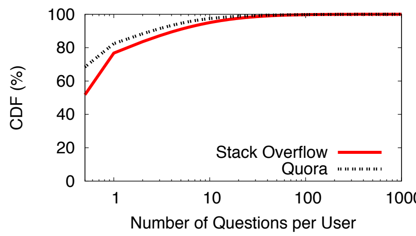


Figure 9: # of Questions per user.

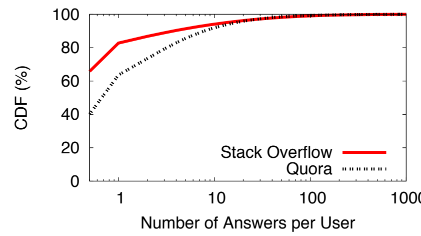


Figure 10: # of Answers per user.

question count. Table 2 lists the top 10 topics with most number of questions in each site. In Quora, the top 10 includes topics in various areas including technology, food, entertainment, health, etc. "Startups" is the most popular one which takes 3.7% of the questions. While all topics in Stack Overflow are different, they are all related to programming. The most popular topic is "C#," which represents roughly 10% of all questions.

Figure 3 plots the distribution of number of questions per topic in Quora in a log-log grid. It shows that for the large majority of topics, each topic contains only a handful of questions, while a few popular topics are responsible for most of all questions. The distribution of number of questions per topic mirrors a power-law distribution. Performing a power-law fitting produces alpha value 2.28 with error 0.03.

Questions and Answers. In both systems, one question can have multiple topics. Figure 4 shows the number of topics per question. Stack Overflow requires a minimum of 1 topic and a maximum of 5 topics per question, and the results are evenly distributed between 1 and 5. Although Quora does not have such requirements, a majority (85%) of questions have no more than 5 topics. Very few ($<$1%) of questions end up with more than 10 topics, which might be an attempt to draw more attention to the question.

Next, we plot the distribution of views and answers per question in Figure 5 and Figure 6. We are surprised to find that the curves from Stack Overflow and Quora are nearly identical. Although 20% of questions in Quora remain unanswered (10% for Stack Overflow), almost all questions got at least 1 user view. In addition, 99% of questions end up with less than 10 answers, and 20% of all Quora questions managed to collect $\geq$4 answers. We use this as a minimum threshold for our later analyses on social factors on system performance.

In terms of votes, both Quora and Stack Overflow allow users to upvote and downvote answers. Quora makes visible the list of upvoters, but hides downvoters. Downvotes are processed and only contribute to determining the order answers appear in. Thus in our analysis of Quora, we only refer to upvotes and disregard downvotes. In contrast, Stack Overflow anonymizes all voters and only displays the accumulated number of votes, which can be negative if an answer is poorly received. We plot the distribution of votes per answer in Figure 7. There is still a fairly big portion of answers (about 40% for both sites) that received no votes from users.

Next, we look at how votes impact the order that answers are displayed. Quora uses a proprietary algorithm [1] to rank the answers, where best answers show on the top of the page. In Stack Overflow, the question asker can *accept* one of the answers as the best answer. First, we examine how well votes work to identify the "best answer." We select questions with at least 2 answers, 180K or 40% of all questions in Quora and 1.76M or 51% in Stack Overflow. Figure 8 plots the ratio of the best answer' votes over the average votes per answer under this question. We call this as "best answer vote ratio." Overall, vote count was very effective at identifying the best answers, and the differences between the two sites might be explained by the more concrete (right or wrong) nature of Stack Overflow's questions compared to general questions on Quora. Surprisingly, some of the best answers have less votes than the average answer. 5% of Quora questions ranked answers with fewer upvotes on top, likely due to other features used by Quora's ranking algorithm such as answerer reputation or downvotes. On Stack Overflow, 7% of the answers chosen by the asker had lower votes than average.

Finally, we note that both sites use crowdsourcing to moderate user-generated content. Stack Overflow has administrators who actively flag unqualified questions and close them [4]. Roughly 3% of all questions in Stack Overflow have been closed, and Figure 11 shows the reasons why they were closed. The top two reasons were "not-real," *i.e.* ambiguous, vague, incomplete, overly broad or rhetorical, and redundant questions. In contrast, Quora relies on a total of 43 admins and 140 reviewers chosen from the user population to flag low quality answers and redundant questions [6, 7]. The number of flagged or removed answers and questions is unknown. While it is unclear whether these reviewers are responsible for keeping Quora largely free of fake or scripted accounts (Sybils) [37, 39], recent work has shown that human reviewers can be extremely effective at detecting fake or forged content [36].

Users Activity. Finally, we compare levels of user activity in Quora and Stack Overflow. Figure 9 and Figure 10 show the

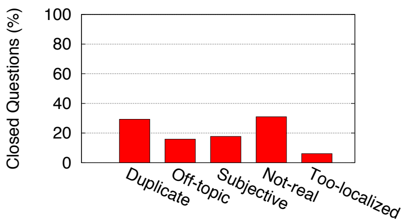


Figure 11: Reasons for deleting questions in Stack Overflow.

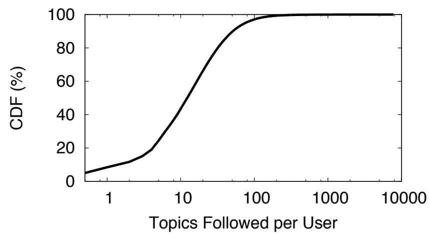


Figure 12: Topics followed per user.

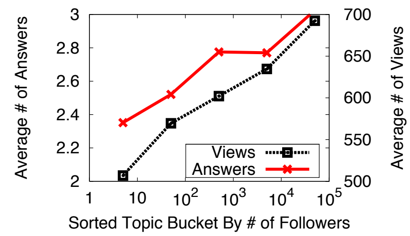


Figure 13: Average views and answers for questions under sorted topic buckets.

total number of questions and answers posted by each user. 60% of Stack Overflow users did not post any questions (or answers), while less than 1% of active users post more than 1000 questions (or answers). We observe similar trends in Quora. 40% of the users in our dataset did not post any answers, and 70% of the users have not asked any questions, indicating that a small portion of users have contributed most of the content.

Summary. Despite their different topics of interest, Quora and Stack Overflow share many similarities in distribution of content and activity. A key observation is that given the broad and growing number of topics in Quora, identifying the most interesting and useful content, *i.e.* separating the wheat from the chaff, is a very difficult problem. Without built-in mechanisms to lead users to useful content, the service will overwhelm users with the sheer volume of its content, much like the Internet itself. This is the focus of the rest of our paper, where we will study different Quora mechanisms to understand which, if any, can keep the site useful by consistently guiding users to valuable information.

4. THE USER-TOPIC GRAPH

Quora allows users to track specific fields by following the corresponding topics, such as "Startups," "Facebook," and "Technology." This also directly connects users to questions (and associated answers). A question, once created or updated under a topic, will be pushed to the newsfeeds of users who follow the topic. In this section, we model the interaction between Quora users and topics using a *user-topic graph*, and examine the impact of such interactions on question answering and viewing activities. Specifically, we seek to understand whether there is a direct correlation between followers of a topic and views and answers to questions, *i.e.* do highly-followed topics draw a large number of views and answers to their questions?

4.1 High-level Statistics

We first examine the number of topics followed by each user[1]. Figure 12 shows the cumulative distribution of the number of topics followed per user. We make three key observations. First, the large majority (95%) of users have followed at least 1 topic. This is because Quora recommends topics during the sign-up process. Second, Quora users each tend to follow a moderate number of topics, *e.g.* more than 50% of users followed at least 10 topics, but 97% of users followed no more than 100 topics. Finally, a very small portion of users (27 or 0.01%) followed more than 1000 topics. We manually checked these users and found that they were legitimate accounts, and come from various backgrounds such as CEOs, co-founders, bloggers, students, and were all very active Quora users.

[1]The user-topic interaction is one-way where users can follow multiple topics, but the relation is asymmetric, *i.e.* topics do not follow users.

Topic	# of Followers	Topic	# of Followers
Startups	47,084	Google	18,867
Facebook	25,569	Science	17,669
Twitter	23,034	TechCrunch	13,313
Technology	21,852	Music	13,084
Entrepreneurship	20,661	Venture-Capital	12,863

Table 3: Top 10 topics in Quora based on number of followers.

Next, we rank the topics by the number of followers. Since each Quora user lists the topics she follows in her profile, we estimate the number of followers by examining user profiles in our crawled dataset. Out of 56K topics crawled, 35K topics have at least 1 follower in our dataset. Using these estimates, we list in Table 3 the top 10 topics with the most followers. Clearly, users were highly biased towards certain topics. For example, "Startups" was followed by nearly 18% of users, and "Venture-Capital" by 5% of users. More interestingly, when compared to Table 2 ranking topics by number of questions, only 4 topics ("Startups", "Facebook", "Google", and "Music") are in the top-10 of both rankings. This shows that a high level of interest in a topic, *i.e.* more followers, does not necessarily produce more questions.

4.2 Impact on Question-related Activities

We now examine whether user interest towards certain topics translates into higher level of activities on questions related to those topics. We examine the correlation between the number of views or answers per question, and the number of followers of each topic. Since the number of topics is large (35K), we bucketize topics based on the number of followers in a log scale. For example, topics with number of followers in the range $[1, 10]$ are in one bucket, and topics with number of followers within $[10, 100]$ are in a second bucket. We have a total of 5 buckets. In each bucket, we compute the number of views (answers) per question, averaged over the topics and their questions.

Figure 13 shows the correlation results for both question views and answers. We observe a strong correlation: questions under topics with more followers tend to have a higher number of average page views and answers. This is intuitive: when a user follows a topic, all questions under the topic and their updates show up on the user's newsfeed, thus encouraging page views and answers.

We verify this intuition by examining for each question the percentage of answers that came from followers of the question's topic(s). Unfortunately we could not do the same for question page views, because Quora only reveals the identity of users who answer questions, but not those who browse each question. We focus on questions with some minimum number of user interactions ($\geq$4 answers), which filters out all but 87K (20%) questions from our

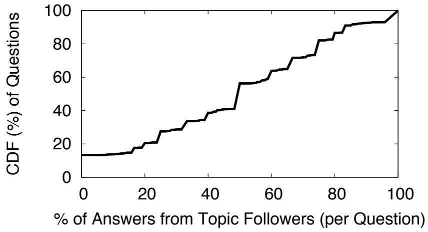


Figure 14: % of Answers added by the followers of the question's topics.

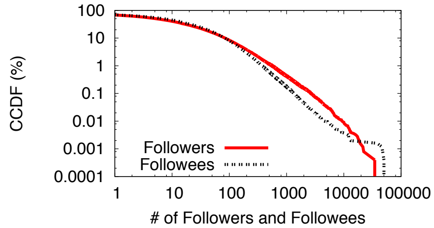


Figure 15: Degree distribution in social graph.

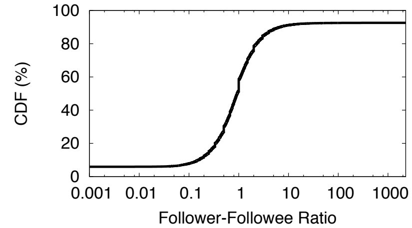
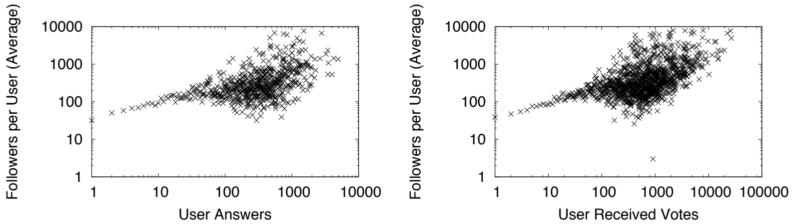


Figure 16: Follower-followee ratio.



(a) Answers

(b) Received Votes

Figure 17: Correlation between user answers (received votes) and followers per user.

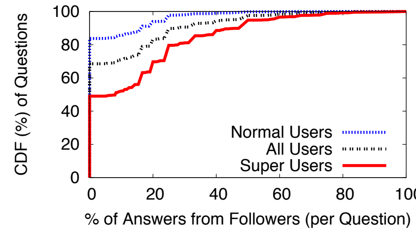


Figure 18: % of Answers written by asker's followers.

dataset. Figure 14 plots the cumulative distribution of the portion of answers contributed by topic followers. It is very close to a uniform distribution with mean of 50%, except for roughly 13% of questions, for which none of the answers were produced by followers of the question's topic(s). At a high level, this suggests that topics are effective ways of guiding users towards questions that are valuable and appealing to them.

Summary. The user-topic interaction has considerable impact on question answering activities in Quora. Not surprisingly, questions under well-followed topics generally draw more answers and views. Following the right topics can introduce users to valuable questions and answers, but is not the only way to access questions.

5. THE SOCIAL GRAPH

In addition to following topics of interest, Quora users also follow each other to form a Twitter-like directed social graph. Questions that a user interact with are dissiminated to their followers in the form of events in their newsfeed. Therefore, social relationships clearly affect Q&A activities, and serve as a mechanism to lead users to valuable information.

In this section, we analyze the Quora social graph to understand the interplay between user social ties and Q&A activities. Specifically, we seek to answer three key questions. First, what triggers Quora users to form social ties? Second, does the presence of popular users correlate with high quality questions or answers? That is, do questions raised by "super-users" with many followers receive more and/or better answers from her followers? Finally, do strong social ties contribute to higher ratings on answers to questions? In other words, do questions answered by super-users get more votes because of the sheer number of their followers?

5.1 Social Ties

We begin by examining the follower and followee statistics of Quora users. Figure 15 plots the complementary cumulative distribution function (CCDF) for both the incoming degree (follower) and outgoing degree (followee). As expected, the degree distribution follows the power-law distribution [10]. Specifically, 23% of users have no followers and 23% do not follow anyone. The vast majority of users (99.6%) have less than 1000 followers, while 23 users have more than 10,000 followers. The exponential fitting parameter α for the incoming degree distribution is 2.49 (with fitting error 0.01). This is very close to that of Twitter (α=2.28), but higher than that of Facebook and Orkut (α=1.5) [38, 26].

Figure 16 plots the distribution of the follower-followee ratio (FFRatio), the ratio of a user's incoming and outgoing degrees. In our data set, 44,091 (17%) of all users have neither followers nor followees (and are thus removed in this particular analysis). For the rest, 6% of users have no followers, and 7% do not follow anyone, representing the two extremes in the FFRatio distribution. Overall, more than half (58%) of all users have more followees than followers. A very small portion (0.1%) have 100 times more followers than followees. Not surprisingly, these are mostly celebrities, *e.g.* editors, actors and CEOs.

Triggers of Social Ties. To understand how Quora's social network functions, a basic question of interest is how users choose their followees. According to a recent survey of Quora users [31], they tend to follow users who they consider interesting and knowledgeable. Thus our hypothesis is that, outside of the small portion of celebrities who get followers just by their mere presence, the majority of Quora users attract followers by contributing a large number of high-quality answers.

To validate our hypothesis, we examine the correlation between a user's follower count and the quantity and quality of her answers to questions. We approximate the quality of an answer by the number of votes received. We put users with the same number of answers (votes) into a group and compute the average number of followers per user for each group. Figure 17(a) plots the correlation results, which confirm our hypothesis. The correlation is particularly strong for users with less than 100 followers, which account for 91% of the users in our dataset.

5.2 Impact on Question Answering

Quora is unique because it integrates an effective social network (shown above) into a tradition Q&A site. Thus it is important to

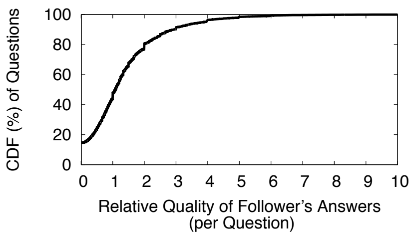


Figure 19: Relative answer quality ratio.

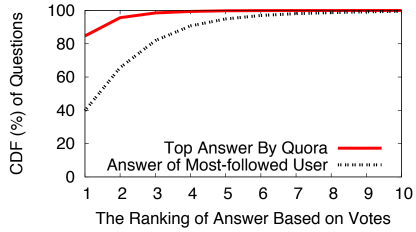


Figure 20: Ranking answers by author popularity vs top answer selected by Quora.

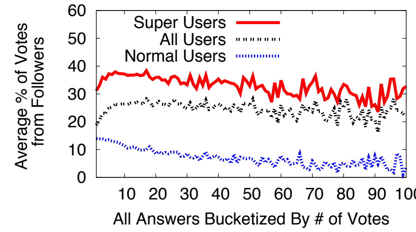


Figure 21: Votes from answerer's followers.

understand how social ties affect Q&A activities. Specifically, we explore whether super users (or users with many followers) draw more and better answers from their followers. To answer this question, we first examine for each question the number of answers, and the portion of answers coming from the asker's followers. We then measure the quality of answers based on votes and explore whether followers provide better answers. We define "Super User" as top 5% of all users by followers. In our dataset, we have 12K super users, each with more 160 followers.

For questions in our dataset, the asker is not shown on the question page. Instead, we match the originator of the question (the "asker") to each question based on user profiles. Each user's profile page contains a list of user's previously asked questions. Using this list, we managed to find the askers for 285K (65%) questions in our question dataset. Since our analysis targets user social activities in the question thread, we do not consider open questions and questions that have not gained enough answers. We only consider questions with known askers and at least 4 answers, which still leaves a large number of questions (59K) for our analysis.

Number of Answers. In Figure 6 we have plotted the distribution of the number of answers received per question across all the questions. We repeat this analysis for both questions raised by super users and non-super users (regardless of the number of answers received), and found that they follow the same distribution (figure omitted due to the space limitations). This shows that users do not get more answers for questions just by having more followers.

Answers by Followers. Next, we examine for each question the portion of answers contributed by the asker's followers. Figure 18 plots the cumulative distribution across all the questions (marked as "All"), across the questions raised by super users ("Super User"), and across the questions raised by non-super users ("Normal").

We make two key observations. First, a big portion of the questions (68% for "All") did not receive answers from the asker's followers. Even half of the questions raised by super users received no answers from their followers. This is likely because users who follow someone tend to seek her (helpful) answers to questions, rather than looking for questions to answer. This also implies that if we build a Q&A site *solely as a social network* that expects answers only from friends (followers), most questions will remain unanswered. Second, compared to normal users, super users do draw more answers from their followers, indicating a moderate level of social influence on question answers.

We also compare the effectiveness of drawing answers using social ties to that of drawing answers from following topics (discussed in Section 4), by comparing the results in Figure 18 and Figure 14. We see that in general, questions received more answers from users who follow the associated topic(s). But neither channel appears to be the primary way of attracting answers, and both channels appear to complement each other in this process.

Answer Quality. We now examine whether answers contributed by the asker's followers have better quality. Again we use the number of votes received to serve as an approximate measure of the quality of an answer. For each question thread, we first compute the average votes per answer for all the answers V_{all} and for all the answers contributed by the asker's followers $V_{follower}$. We define $R = \frac{V_{follower}}{V_{all}}$ as the relative quality of the followers' answers. Thus $R > 1$ indicates that the followers' answers are of higher quality in general.

Figure 19 shows the cumulative distribution of R, where for more than 50% of the questions, answers from the followers were of higher quality, and for 20% of the questions, answers from the followers got more than 2 times the votes than average. This result is consistent with a recent survey study [27] on Q&A behaviors in Facebook, which suggests that close friends have stronger motivation to contribute good answers.

5.3 Impact on Voting

Quora applies a voting system that leverages crowdsourced efforts to promote good answers. By positioning good answers at the top of the questions page, Quora allows users to focus on valuable content. However, the social interaction among Quora users could impact voting in various ways. The key concern is users who have many followers can get their followers to vote for their answers, thus gaining an "unfair advantage" over other users. In the following, we study this issue in detail by exploring two key questions. First, do user votes have a large impact on the ranking of answers in Quora? Second, do super users get more votes, and do these votes mainly come from their followers?

Votes and Ranking. Quora has indicated that the number of votes is the key metric to determine quality of answers [1]. In fact, our results in Figure 8 show that more than 96% of the best answers (ranked 1st by Quora) received more votes than average. Thus our goal is to explicitly examine how much the number of votes matters in Quora's ranking algorithm, and whether social connections give user advantage to gain more votes.

For each question thread, we start by ranking the answers by the number of votes received. Answers with the most votes are ranked first. We then take the best answer (ranked 1st) chosen by Quora's built-in algorithm and study their vote-based ranking. Figure 20 plots the cumulative distribution of these best answers' vote-based ranking. We see that for 85% of the questions, Quora's best answers also ranked the highest in votes, and for 96% of the questions, the best answers from Quora are among the top-2 most votes. This result confirms that the number of votes is the dominating feature for selecting best answers. The same result also implies that potential bias in the voting process could lead to unfair ranking of answers, which we study next.

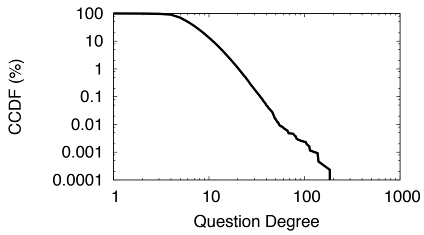


Figure 22: Node degrees in the related question graph.

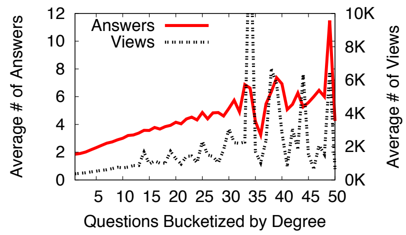


Figure 23: Question degree versus average views and answers per question.

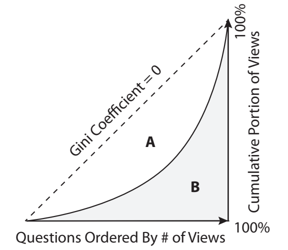


Figure 24: Gini coefficient, $G = A/(A+B)$.

Votes on Super Users. We repeat the above analysis on answers offered by super users (most followed users). Results in Figure 20 show that for 40% of questions, super users' answers received the highest votes, and for 60% of cases, their answers are among the top-2 most votes. This implies that regardless of the quality of their answers, super users can often get more votes over other users.

To better understand the bias, we examine whether a large portion of votes come from the answerer's followers. For this we gather answers to all the questions and group them by the number of votes received. For each group of answers, we compute the average percentage of votes from the answerer's followers. We also repeat the same process on answers offered by super users and those from non-super users. Figure 21 shows the average percentage of followers' votes across different answer groups. We cut the line at the points where the number of votes reaches 100, which covers 99.9% of all answers (see Figure 7). These results show that answers contributed by super users do receive a large portion of votes (30-40%) from their followers, which is significantly larger than normal users (<10%). This shows that users with more followers tend to get more votes from their followers, which could introduce potential unfairness in answer ranking. For example, an answer contributed by super users gets a much higher rank even though the true quality of the answer is not high.

Summary. In Quora, users who contributed more and good answers tend to have more followers. These well-connected users also gain advantage by having more friends (followers) to answer their questions and upvote for their answers.

6. THE RELATED QUESTIONS GRAPH

One of Quora's core features is the ability to locate questions "related" to a given question. This effectively creates a *related question graph*, where nodes represent questions, and links represent a measure of similarity as determined by Quora. The related question graph provides an easy way for users to browse through Quora's repository of questions with similarity as a distance metric.

In this section, we extract the question graph from our dataset, and seek to determine if the structure of the graph plays a role in helping users to find top questions. Intuitively, a similarity-based question graph would produce large clusters of questions around popular topics, with less popular questions relegated to sparse regions of the graph. Thus users following related question links could encounter popular questions with a higher probability.

6.1 Impact of Degree in the Question Graph

We build the question graph by crawling and extracting related questions links. By default, Quora lists a fixed number (5) of related questions on each question's main page. These are deemed by Quora to be the most related to the question on the current page. Since the "related" relationship is intuitively a bidirectional property, the question graph is a unweighted, bidirectional graph.

Our final question graph has a total of 437K nodes and 1.6M edges. We plot the distribution of question degree in Figure 22. Although each question only has at most 5 outgoing related questions, most questions have incoming connections from other questions, and thus have a total "related" degree greater than 5. However, there are 9% questions with degree less than 5. This is because some of their related questions were not crawled (questions deleted by Quora) and thus are not included as nodes. 99% of the questions have degree less than 50. The distribution shows a distinctive power-law shape, and when we fit the question degree CCDF to the power-law, we get an α value of 3.5 with fitting error 0.048.

Next, we examine the connectivity of the question graph. The question graph is dominated by a single large connected component that covers 98% (430K) of all questions. On closer inspection, we see that the remaining 2% of the questions are either newer questions whose related questions have not yet been computed, or they belong to esoteric topics with very few questions and low user interest.

Stability. One concern we had about the question graph is whether it is stable, *i.e.* does it change on a frequent basis as new questions are added to the system. We test the long-term stability of the related question graph by comparing the related question graph across two snapshots. The first snapshot was taken in our primary measurement period of August 2012. We also took another snapshot in October 2012 (two months after the first snapshot). When we compared the related question set for each question in the system, we found that 60% of all question had no changes in the time between our snapshots, and 30% of the questions have only one new entry (out of five) in its related question list. Thus we can assume that the related question list is relatively stable over moderate time periods, and our snapshots are a reasonable approximation for earlier versions of the question graph.

Question Degree vs. Attention. On each question page, users can browse a series of questions via the related question edges. This leads to the hypothesis that a question with higher question degree can receive more attention, *i.e.* more user views, and potentially more answers as a result.

We validate this hypothesis as follows. We first group all questions based on question degree in the related question graph. Then we compute the average number of answers and views for questions in each group. We plot the results in Figure 23. The dashed line represents the average number of user views across all questions with a given node degree, and the solid red line represents the average number of answers received by all questions with a given degree. There are clear trends in both cases. For questions with higher degree (they are listed as being related to more questions), they are accessible to users via a larger number of incoming

ID	Question Title
459576	What percentage of questions on Quora have no answers?
370857	Can I search Quora only for questions that have been answered?
45022	How many questions have been answered on Quora?
20195	What percentage of Quora questions receive at least one answer?
17363	What percentage of questions on Quora are answerable?
13323	How many questions are on Quora, answered or not?
...	...
Top Topics	Quora, Quora-Usage-Data-and-Analysis, Quora-product

Table 4: A cluster of 43 questions, produced by graph partitioning. The top three tags covers 90% of the questions in the cluster.

links. Hence these high degree questions receive both more page views as well as more answers. The takeaway here is that questions with high degree in the question-relation graph correlates strongly to questions that receive more attention and answers from users.

6.2 Locating Similar Questions

In the question graph, questions on similar topics are clustered together, while irrelevant questions are likely to be "related" to popular questions. Thus they are likely positioned in sparser regions of the graph. In this subsection, we first leverage the graph structure to identify groups of similar questions. We then ask two key questions: do similar questions receive equal attention from users? If not, what are the potential mechanisms that drive users to certain questions while ignoring other similar questions?

Graph Clustering. We first locate similar questions using the question graph. More specifically, we want to generate question clusters where questions within the cluster are more tightly connected than those outside the cluster. This is a simple definition easily characterized by modularity.

We formalize this problem as a graph partition problem, and use the popular graph partitioning tool METIS to perform a multilevel k-way partitioning [18] on our question graph. In this case, we predefine K as the number of clusters we want to generate. We run the graph partitioning algorithm, with K equal to 100, 1000, 10,000 and 100,000. When K is too big, we end up with many small clusters after cutting many edges. On the other hand, when K is too small, we get a small number of big clusters which take in many questions under related topics, but are not truly similar. Since there is no good way to get the ground-truth assessment on how "similar" the questions are, we randomly sample 10 clusters from each run with different K values, and manually inspect questions within each cluster. We find that the best match between semantic clusters and automatically detected clusters occurs when $K =$10,000.

So we partition the graph into 10,000 clusters of similar sizes. Table 4 shows an example of one generated cluster. This cluster contains 43 questions, and all questions are related to "Quora." We also extract the topics of the questions in the cluster and rank the topics based on how many questions they are associated with. The top 3 topics of the cluster are listed in the table. We see that the three topics are different but all related. In fact, the top three topics cover 90% of the questions in this cluster, which indicates a good cluster focused around a single subject.

Cluster Analysis. Based on the generated clusters, we can now answer the high level question: do similar questions receive equal attention? We answer this question by assessing the distribution of user views and answers between questions in the same cluster. We choose to use *gini coefficient*, a uniformity metric commonly used to evaluate the equality of distributions in economics [11].

We explain how we compute gini coefficient for each question cluster using Figure 24. As an example, the x-axis has the questions sorted by increasing number of views, and the y-axis represents the cumulative portion of the views. So the curve represents y% of contribution (of views) by the bottom x% of questions. By definition, the curve is always at or below the dashed line which represents perfect equality of the distribution. Gini coefficient is defined to quantify how close the curve is to the dashed line: $G = \frac{A}{A+B}$, where A and B represent the corresponding areas above and below the curve. As each axis is normalized to 100%, the gini coefficient G is always within the range of $[0, 1]$, where G=0 means perfect equality or uniformity (the dashed line in our example) and G=1 means an extremely skewed distribution.

We compute the gini coefficient for the distribution of number of views (and answers) of questions in each cluster. As shown in Figure 25, the solid curve shows the gini coefficient of number of views is highly skewed towards 1. More than 90% of clusters have gini coefficient >0.4. This shows that the numbers of views are extremely uneven among similar questions within each cluster. The same trend applies to answers, as the vast majority of clusters have extremely skewed answer distributions. This means that user attention is tightly focused on a small portion of (valuable) questions within each cluster of similar questions.

Our results suggest that the structure of the related question graph (*e.g.* question degree) is at least partially responsible for focusing user attention and answers on a small subset in each cluster of related questions. Next, we ask whether super users play a role in directing traffic towards specific questions in each cluster of related questions.

Super User Effect. We evaluate whether the skew in the distribution is caused by super user effect. Intuitively, when a user adds new answers or upvotes existing answers on a question, that question will be pushed to all her followers. Thus super users with more followers can disseminate the question to a larger audience. We use the same definition of super users as in previous analysis by taking the top 5% of most followed users. We measure the super user effect by comparing the number of views (answers) of questions involving super users to other questions with no super user involvement. Among all 10000 clusters, only 1 cluster has no super user in any of its questions, and is not considered in the analysis.

Figure 26 shows the scatter plot of average views (and answers) of super user involved questions and normal user questions in each cluster. The X-axes are presented in ascending order of the views (answers) of super user questions, thus the super user question points form a near-continuous line. We first compare the average number of user views in Figure 26(a). In the vast majority of the clusters, the super user questions have more views than that of questions with no super user involvement. There is only a small number of clusters (4%) where normal user questions receive more user views then super user questions.

Figure 26(b) compares the two type of questions with respect to average number of answers per question. The result shows that super user involved questions have significantly more answers than normal user questions. Compared to user views, it shows a even stronger impact of super users on drawing answers. In different clusters, super user questions have an average number of answers ranging from 2 to 10, while questions without super user involvement almost always stays below 2 answers across clusters. Both the number of user views and answers can reflect how much attention each question receives. The result shows choices made by a small number of super users on questions usually affect the focus of attention for the whole community.

Summary. We build the related question graph, and find that it is a relatively stable structure even as new questions are constantly

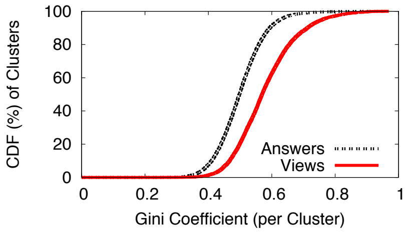
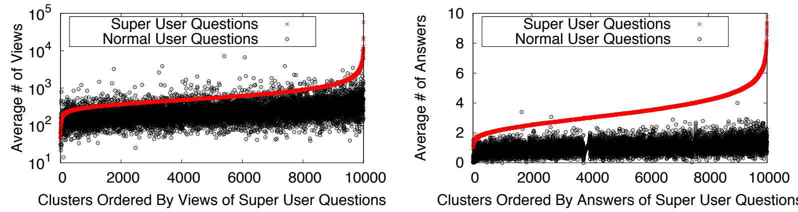


Figure 25: Gini coefficient of view (answer) distribution in each cluster.



(a) Views

(b) Answers

Figure 26: Average # of views (answers) of super user questions vs. normal user questions in each cluster.

added to the system. We find that high degree questions generally receive more answers and views compared to others. More specifically, the spread of user views and answers within clusters of related questions is extremely skewed towards a small subset of questions. This bias is likely created by the structure of the question graph, and enhanced by super users, as the questions they interact with receive additional views and answers from their followers.

7. RELATED WORK

Community based Q&A. Researchers have studied community based Q&A (CQA) sites such as Yahoo Answers [13, 8, 12, 25, 33, 34], MSN QnA [15, 32], Stack Overflow [9, 22], Math Overflow [35] from different perspectives. One perspective focuses on managing questions and topics in CQA sites. Some studies look at question archiving and tagging [32]. Others focus on classifying factual questions with conversional questions [12, 25], or reusing the knowledge collected from old questions to answer new similar questions [34]. Finally, others evaluate the quality of user generated content, including answer quality [33, 35, 8, 16] and question quality [9, 20].

A second group of work studies user communities in CQA sites. These projects aim to develop algorithms to identify users with high expertise. One direction is to rank users based on expertise measures generated from user history data (*e.g.* questions, answers, votes) [8, 28, 21]. Another direction is modeling user interaction to design network-based ranking algorithms to identify experts [17, 19, 40]. Finally, other works study user community from perspectives such as answering speed [22] and user incentives in CQA sites [15].

Our work differs from prior art, since we are the first to analyze a social network based Q&A site using large-scale data measurement and analysis. Instead of treating all users as one big community, we explore the impact of a built-in social network as well other graph structures on the Q&A activities. A recent report [31] looks at Quora's reputation system in depth with a small dataset of 5K questions.

Q&A in Social Networks. Studies have also looked into the question and answering behaviors in existing online social networks. Users can ask their friends questions by posting tweets in Twitter [30] or updating status in Facebook [29, 27, 14]. These studies answer high-level questions like what types of questions are suitable to ask in social networks, and whether strong ties (close friends) provide better answers than weak ties.

8. CONCLUSION

Community question and answer sites provide a unique and invaluable service to its users. Yet as these services grow, they face a common challenge of keeping their content relevant, and making it easy for users to "find the signal in the noise," *i.e.* find questions and content that are interesting and valuable, while avoiding an increasing volume of less relevant content.

In this paper, we use a data-driven study to analyze the impact of Quora's internal mechanisms that address this challenge. We find that all three of its internal graphs, a user-topic follow graph, a user-to-user social graph, and a related question graph, serve complementary roles in improving effective content discovery on Quora. While it is difficult to prove causal relationships, our data analysis shows strong correlative relationships between Quora's internal structures and user behavior. Our data suggests that the user-topic follow graph generates user interest in browsing and answering general questions, while the related question graph helps concentrate user attention on the most relevant topics. Finally, the user-to-user social network attracts views, and leverages social ties to encourage votes and additional high quality answers. As Quora and its repository of data continues to grow in size and mature, our results suggest that these unique features will help Quora users continue find valuable and relevant content.

9. ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their valuable feedback. This work is supported in part by NSF grant CNS-1224100, DARPA GRAPHS program, and gifts from Cisco and Google. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

10. REFERENCES

[1] What is quora algorithm/formula for determining the ordering or ranking of answers on a question? Quora, Feb. 2011. `http://www.quora.com/Quora-product/What-is-Quoras-algorithm-formula-for-determining-the-ordering-ranking-of-answers-on-a-question`.

[2] How many questions are on quora, answered or not? Quora, Mar. 2012. `http://www.quora.com/Quora-Usage-Data-and-Analysis/How-many-questions-are-on-Quora-answered-or-not`.

[3] On what topics does quora have the best questions and answers? Quora, Apr. 2012. `http://www.quora.com/Lists-of-Top-Quora-Content/On-what-topics-does-Quora-have-the-best-questions-and-answers`.

[4] What is a closed question? Stack Overflow, Oct. 2012. `http://meta.stackoverflow.com/questions/10582/what-is-a-closed-question`.

[5] What is quora's policy on question deletion? Quora, Jun. 2012. `http://www.quora.com/Quora-Policies-and-Guidelines/What-is-Quoras-policy-on-question-deletion`.

[6] Who are all of the reviewers on quora? Quora, Nov. 2012. `http://www.quora.com/Quora-Reviewers/Who-are-all-of-the-reviewers-on-Quora`.

[7] Who are the current site admins of quora? Quora, Aug. 2012. `http://www.quora.com/Quora-Admins/Who-are-the-current-site-admins-of-Quora`.

[8] ADAMIC, L. A., ZHANG, J., BAKSHY, E., AND ACKERMAN, M. S. Knowledge sharing and yahoo answers: everyone knows something. In *Proc. of WWW* (2008).

[9] ANDERSON, A., HUTTENLOCHER, D., KLEINBERG, J., AND LESKOVEC, J. Discovering value from community activity on focused question answering sites: a case study of stack overflow. In *Proc. of SIGKDD* (2012).

[10] BARABASI, A.-L., AND ALBERT, R. Emergence of scaling in random networks. *Science 286* (1999).

[11] DAGUM, C. The Generation and Distribution of Income, the Lorenz Curve and the Gini Ratio. *Economie Applique 33* (1980), 327–367.

[12] HARPER, F. M., MOY, D., AND KONSTAN, J. A. Facts or friends?: distinguishing informational and conversational questions in social Q&A sites. In *Proc. of CHI* (2009).

[13] HARPER, F. M., RABAN, D., RAFAELI, S., AND KONSTAN, J. A. Predictors of answer quality in online Q&A sites. In *Proc. of CHI* (2008).

[14] HOROWITZ, D., AND KAMVAR, S. D. The anatomy of a large-scale social search engine. In *Proc. of WWW* (2010).

[15] HSIEH, G., AND COUNTS, S. mimir: A market-based real-time question and answer service. In *Proc. of CHI* (2009).

[16] JEON, J., CROFT, W. B., LEE, J. H., AND PARK, S. A framework to predict the quality of answers with non-textual features. In *Proc. of SIGIR* (2006).

[17] JURCZYK, P., AND AGICHTEIN, E. Discovering authorities in question answer communities by using link analysis. In *Proc. of CIKM* (2007).

[18] KARYPIS, G., KUMAR, V., AND KUMAR, V. Multilevel k-way partitioning scheme for irregular graphs. *Journal of Parallel and Distributed Computing 48* (1998), 96–129.

[19] LERMAN, K., AND GALSTYAN, A. Analysis of social voting patterns on digg. In *Proc. of WOSN* (2008).

[20] LI, B., JIN, T., LYU, M. R., KING, I., AND MAK, B. Analyzing and predicting question quality in community question answering services. In *Proc. of CQA Workshop (WWW)* (2012).

[21] LI, B., AND KING, I. Routing questions to appropriate answerers in community question answering services. In *Proc. of CIKM* (2010).

[22] MAMYKINA, L., MANOIM, B., MITTAL, M., HRIPCSAK, G., AND HARTMANN, B. Design lessons from the fastest Q&A site in the west. In *Proc. of CHI* (2011).

[23] MCGEE, M. Quora traffic has grown 882% in the last year, but it's still just a fraction of what yahoo answers gets. MarketingLand, Dec. 2011.

[24] MCGEE, M. Yahoo answers hits 300 million questions, but Q&A activity is declining. Search Engine Land, Jul. 2012.

[25] MENDES RODRIGUES, E., AND MILIC-FRAYLING, N. Socializing or knowledge sharing?: characterizing social intent in community question answering. In *Proc. of CIKM* (2009).

[26] MISLOVE, A., MARCON, M., GUMMADI, K. P., DRUSCHEL, P., AND BHATTACHARJE, B. Measurement and analysis of online social networks. In *Proc. of IMC* (2007).

[27] MORRIS, M. R., TEEVAN, J., AND PANOVICH, K. What do people ask their social networks, and why?: a survey study of status message Q&A behavior. In *Proc. of CHI* (2010).

[28] PAL, A., CHANG, S., AND KONSTAN, J. A. Evolution of experts in question answering communities. In *Proc. of ICWSM* (2012).

[29] PANOVICH, K., MILLER, R., AND KARGER, D. Tie strength in question & answer on social network sites. In *Proc. of CSCW* (2012).

[30] PAUL, S. A., HONG, L., AND CHI, E. H. Is twitter a good place for asking questions? a characterization study. In *Proc. of ICWSM* (2011).

[31] PAUL, S. A., HONG, L., AND CHI, E. H. Who is authoritative? understanding reputation mechanisms in quora. In *Proc. of Collective Intelligence* (Cambridge, MA, April 2012).

[32] RODRIGUES, E. M., MILIC-FRAYLING, N., AND FORTUNA, B. Social tagging behaviour in community-driven question answering. In *Proc. of WI-IAT* (2008).

[33] SHAH, C., AND POMERANTZ, J. Evaluating and predicting answer quality in community QA. In *Proc. of SIGIR* (2010).

[34] SHTOK, A., DROR, G., MAAREK, Y., AND SZPEKTOR, I. Learning from the past: answering new questions with past answers. In *Proc. of WWW* (2012).

[35] TAUSCZIK, Y. R., AND PENNEBAKER, J. W. Predicting the perceived quality of online mathematics contributions from users' reputations. In *Proc. of CHI* (2011).

[36] WANG, G., MOHANLAL, M., WILSON, C., WANG, X., METZGER, M., ZHENG, H., AND ZHAO, B. Y. Social turing tests: Crowdsourcing sybil detection. In *Proc. of NDSS* (San Diego, CA, February 2013).

[37] WANG, G., WILSON, C., ZHAO, X., ZHU, Y., MOHANLAL, M., ZHENG, H., AND ZHAO, B. Y. Serf and turf: crowdturfing for fun and profit. In *Proc. of WWW* (Lyon, France, April 2012).

[38] WILSON, C., BOE, B., SALA, A., PUTTASWAMY, K. P. N., AND ZHAO, B. Y. User interactions in social networks and their implications. In *Proc. of EuroSys* (Nuremburg, Germany, April 2009).

[39] YANG, Z., WILSON, C., WANG, X., GAO, T., ZHAO, B. Y., AND DAI, Y. Uncovering social network sybils in the wild. In *Proc. of IMC* (Berlin, Germany, November 2011).

[40] ZHANG, J., ACKERMAN, M. S., AND ADAMIC, L. Expertise networks in online communities: structure and algorithms. In *Proc. of WWW* (2007).