An Analysis of the Structure and Dynamics of Large-scale Q/A Communities

Daniel Schall and Florian Skopik

Distributed Systems Group, Vienna University of Technology
Argentinierstrasse 8/184-1, A-1040 Vienna, Austria
{lastname}@infosys.tuwien.ac.at,
WWW home page: http://www.infosys.tuwien.ac.at

Abstract. In recent years, the World Wide Web (WWW) has transformed to a gigantic social network where people interact and collaborate in diverse online communities. By using Web 2.0 tools, people contribute content and knowledge at a rapid pace. Knowledge-intensive social networks such as Q/A communities offer a great source of expertise for crowdsourcing applications. Companies desiring to outsource human tasks to the crowd, however, demand for certain guarantees such as quality that can be expected from returned tasks. We argue that the quality of crowd-sourced tasks greatly depends on incentives and the users' dynamically evolving expertise and interests. Here we propose expertise mining techniques that are applied in online social communities. Our approach recommends users by considering contextual properties of Q/A communities such as participation degree and topic-sensitive expertise. Furthermore, we discuss prediction mechanisms to estimate answering dynamics considering a person's interest and social preferences.

Keywords: Online communities, Expertise mining, Crowdsourcing

1 Introduction

The collaboration landscape has changed dramatically over the last years by enabling users to shape the Web and availability of information. While in the past collaborations were bound to intra-organizational collaborations using company-specific platforms, and also limited to messaging tools such as email, it is nowadays possible to utilize the knowledge of an immense number of people participating in collaborations on the Web. The shift toward the Web 2.0 allows people to write blogs about their activities, share knowledge in forums, write Wiki pages, and utilize social platforms to stay in touch with other people. Task-based platforms for human computation and *crowdsourcing* enable access to the knowledge of thousands of people on demand by creating human tasks that are processed by the crowd. Platforms for human computation including Amazon Mechanical Turk (MTurk)[1] are examples for crowdsourcing applications associated to the business domain. Human tasks include activities such as

---

[1] http://www.mturk.com/

designing, creating, and testing products, voting for best results, or organizing information.

In open and dynamic crowdsourcing environments it becomes essential to manage expertise profiles and reputation of people in an automated manner. Here we argue that the quality of crowd-sourced tasks depends to a large extent on the users' dynamically evolving expertise and interests. Companies desiring to outsource human tasks to the crowd typically demand for certain guarantees such as quality that can be expected from returned tasks. Therefore, we propose expertise mining techniques that are applied in online social communities. Somebody seeking help or advice on a specific problem or businesses issuing task requests using, for example, crowdsourcing platforms need to be able to find the right person who can assist by offering his or her expertise. Work in expert finding, see for example [4], has been addressing the search for persons with the right skill level by using ontologies and by combining diverse semantic information from user skill profiles. Since Web-scale collaborations involving a large amount of people does not only demand for scalable algorithms and ranking solutions, but in many cases it is also desirable to consider the global properties of a human interaction network to determine the importance of users.

In this paper, we present research performed in the context of Q/A communities and crowdsourcing with the following key contributions:

– We discuss the key characteristics of Q/A communities by studying the properties of a real-world dataset obtained from Yahoo! Answers [19]. Investigating interactions and Q/A behavior of people using Yahoo! Answers allows us to analyze community structures and evolution over time. Our analysis has important implications for the design of future crowdsourcing applications that demand for high-quality results of delivered crowd-sourced tasks.
– We highlight expertise mining techniques that are applied in online social communities. Our approach recommends users by considering contextual properties of Q/A communities such as participation degree and topic-sensitive expertise.
– We propose link analysis techniques derived from popular ranking and mining algorithms such as PageRank [14]. The presented approach accounts for *social dynamics* in Q/A communities. Furthermore, we propose prediction mechanisms to estimate answering dynamics considering a person's interest and social preferences.
– Our evaluations and discussions are based on the properties of a real Q/A community. Our experiments confirm that our proposed ranking approach is well suited for expertise mining and recommendations.

This paper is organized as follows: In Sect. 2 we present related work in the area of online communities and social network analysis techniques applied to expertise networks. Sect. 3 is concerned with the definition of interactions in Q/A systems and basic community characteristics. In Sect. 4, we present our expertise mining model that is based on fine-grained contextual answering behavior followed by our recommendation approach in Sect. 5. Finally, we conclude the paper in Sect. 6.

2 Background and Related Work

The availability of experts in emerging Q/A Web communities raises a number of research issues and provides unique opportunities for new business models. Recent crowdsourcing environments enable novel ways to utilize external expert knowledge. When focusing on large-scale networks, where thousands of individuals are interlinked through a social and collaborative network, the broad applicability of efficient and scalable algorithms becomes evident.

Discovery of experts in Q/A communities. Experts contribute their knowledge in various fields by providing answers, e.g., posting comments, in respective sub-communities [1]. However, since Q/A communities are typically open platforms (everyone can join), quality of provided content becomes a key issue [2]. One approach to cope with quality issues is to find authoritative users based on the link-structure in Q/A communities [8, 11]. Well-established ranking models including HITS [12] and PageRank [14] have been applied and tested in online communities [21]. Our approach is to find experts with respect to different topics [15]. Most existing expertise mining approaches do not consider user posting behavior in different topics. An approach for calculating *personalized PageRank* scores was introduced in [9] to enable topic-sensitive search on the Web. In [18], topic-sensitive ranking in Twitter networks was discussed. The connectivity of users in different *contexts* (e.g., topics) may be very different based on their personal interaction preferences such as frequency of exchanged messages [6]. Thus, considering context and discussion topics may help to discover *relevant experts* in sub-communities.

Routing of requests within Q/A communities. The best candidate user for answering a given question may be the person with the highest reputation in a specific sub-community *and* with the highest number of trusted links. Aardvark [10], for example, is a social search engine that utilizes the social network structure of its users to route questions to be answered by the community. In such a system, it is essential to maintain an up-to-date view on users interests and trust relations. A wide range of computational trust models to control interactions have been proposed (e.g., [3]). Our approach is related to social trust [7, 16, 22] relying on previous behavior. In such Q/A communities, one of the key issues also is to provide recommendations to users whose questions they should answer based on their evolving interests.

Integration of crowds in enterprise systems. An emerging new business model is to integrate the capabilities of crowds in enterprise systems through crowdsourcing techniques [5]. The first commercially available platform was MTurk offering the crowd's capability via Web service-based interfaces. One of the key challenges in crowdsourcing is to estimate the quality that can be expected from a particular crowd member. This could be done based on the user's expertise. If the quality of returned task results is expected to be low (due to missing expertise), different actions can be taken such as issuing multiple instances of the same task and voting for best results. However, since open Web-based platforms are subject to frequent changes (members are joining and leaving), the user's expertise must be calculated automatically.

3 The Yahoo! Answers Community

The goal of this work is to study the characteristics of Q/A communities. To date, Yahoo! Answers (YA) [19] is one of the biggest and most successful Q/A community on the Web. YA was established in 2005 and attracts a large number of users to ask questions regarding different topics. The goal of our analysis of the YA community is twofold:

1. In this work, we analyze the basic properties of the popular YA community with the aim of providing insights in the structure and Q/A behavior of users. The results could be a valuable input when designing, for example, new rewarding or reputation schemes in online communities.
2. By understanding the basic properties of the community, we design a topic-sensitive expertise mining algorithm. Also, understanding the evolution of user interest plays a significant role when recommending questions or users in Q/A communities.

3.1 Interactions and Actor Relations

In the following (Fig. 1) we give an overview of the basic interactions as commonly found in Q/A communities such as YA.
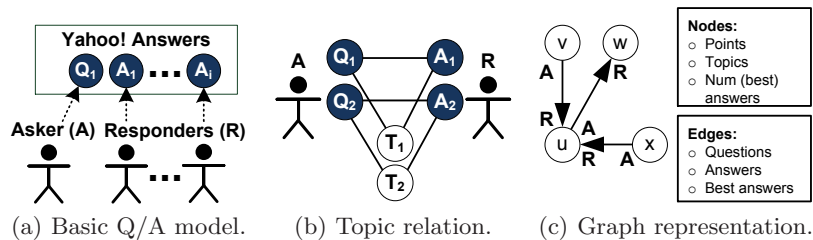

(a) Basic Q/A model.   (b) Topic relation.   (c) Graph representation.

Fig. 1. Representation of Q/A communities.

Figure 1(a) shows the very basic actions taken by *Askers* (**A**) and *Responders* (**R**). Askers post questions that may be replied by answers of one or more responder(s). Typically, one answer is selected as the best answer, thereby setting the status of a question from open to closed. However, if the asker does not opt for a best answer, other community members may vote for a best answer and close a given question. Users get points for answering questions and more points if answers are selected as best answer (see [20] for details about the YA scoring system). The YA system has a predefined set of categories (hereafter refereed to as *topics*) in which questions are posted. Categories typically depict popular *question topics*. Thus, as shown in Fig. 1(b), each question and answer between **A** and **R** is associated with a topic identifier. Of course, users may post and answer questions regarding different topics.

A central theme of our work is the automated analysis of expertise based on the users' answering behavior. The basic idea of our approach is to apply sound

techniques from the information retrieval and social network analysis domain. An intuitive approach is to model the Q/A community as a directed graph $G(N, E)$ composed of the set of node N and the set of edges E. Nodes represent users and edges are established based on interactions that are derived from the underlying question/answering behavior. Given the YA system, each node has a set of properties such as the total collected points, set of topics the user is engaged in, number of answers as well as best answers, and so forth. Fig. 1(c) shows the essential graph model and the meaning of directed edges. In our approach, an edge points from, e.g., v to u, in standard graph notation written as $(v, u) \in E$, if u responded to v's question. Each edge between nodes holds further information such as the number of questions being asked and answered between a pair of nodes u and v.

3.2 Q/A Structure and Community Evolution

The presented analysis of the answers community is based on user interactions (questions and answers) that were posted from 2005 to 2008. We start our discussion of the YA community by analyzing basic statistics. In the following we show the amount of questions and answers posted over time.

Table 1. Q/A volume over time.

	2005	*2006*	*2007*	*2008*
Number of questions $|Q|$	15886	76084	74501	50088
Number of answers $|A|$	40797	828282	790918	304138
Ratio A/Q	≈ 2.6	≈ 11	≈ 11	≈ 6
$|Q|_{year}/|Q|_{total}$	7%	35%	34%	23%
$|A|_{year}/|A|_{total}$	2%	42%	40%	15%

In Table 1, the number of questions and answers in consecutive years from 2005 to 2008 is shown. YA was established in 2005 so the number of questions $|Q|$ and answers $|A|$ is the lowest in this year. An answer per question ratio A/Q of approximately 11% can be observed (i.e., there is a high chance that a question receives multiple answers). Crawling of the dataset was performed until half of 2008. Thus, 2006 and 2007 represented the only years of full Q/A volume. In 2008, the A/Q ratio is lower due to a cutoff in the YA dataset.

The Q/A behavior in the YA community is best approximated as a power-law distribution, which is commonly used to describe the scaling laws of many naturally evolving social networks such as email communications or the access of Web resources [17]. We show the distributions of both *questions per user* and *answers per user* (see Fig. 2). The results show the distributions for the entire interval 2005-2008. The distributions in respective sub-figures can be modeled as $N(k) \sim k^{-x}$ with exponent x. First, we analyzed the distribution of questions per user (see Fig. 2(a) with $x = -1.79$) and questions per user who also *respond to answers* (see Fig. 2(b) with $x = -1.77$). As in many Q/A communities, some users only ask questions (being askers only) and some users ask and answer questions. Second, we analyzed the community structure with respect to answers per users (see Fig. 2(c) with $x = -1.19$) and best answers per user (see Fig. 2(d) with $x = -1.39$).
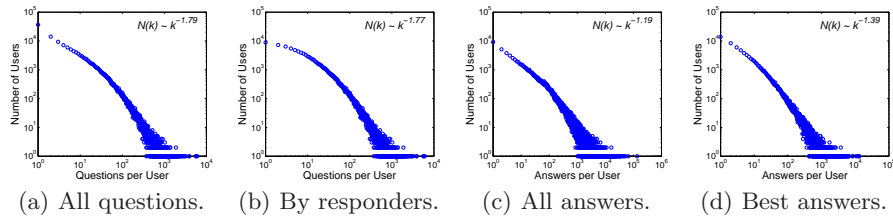
(a) All questions. (b) By responders. (c) All answers. (d) Best answers.

Fig. 2. Question and answer distributions.

In Fig. 3 we show the answering behavior per question. In Fig. 3(a), one can see that a large number of questions is answered by at least one response. Some questions are answered by a large number of responders. This is typically the case when users ask for opinions rather than factual answers. For example, in some cases the YA platform is used like a discussion forum to chat with other users about current news or celebrities [8]. This behavior is reflected by the tail of the distribution in Fig. 3(a).
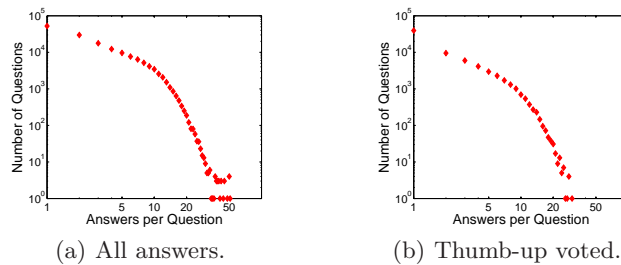

(a) All answers. (b) Thumb-up voted.

Fig. 3. Answers per question.

By selecting only those answers that were voted using the *thumb-up* feature of YA (a 'like' button), the maximum number of answers per question moves from about 50 to about 30 (see the tail of the distribution in Fig. 3(b)). Given that a large number of questions is answered by more than one answer, recommending users who are able to provide high-quality answers becomes important in Q/A communities. For example, a ranked list of answers could be presented to the asker based on the responders community reputation and experience.

The next step in this work is to analyze the role of askers and responders. The interesting aspect in YA is the role of *two-sided markets* [13]. In YA, users get 100 points by signing-up to the platform [20]. For each answer being provided, users get additional points (more points if the answer is selected as best answer). However, users lose points if they ask questions, thereby encouraging members to provide answers. Based on the rewarding scheme in YA, users tend to have either role – being asker or responder – instead of having both roles. This behavior is shown by Fig. 4 where we plot the total number of points and the count of users considering (i) askers only (see diamond shaped symbols) and (ii) responders (star-shaped symbols). One can see that users who only ask question have typically considerably less points than those who also answer questions (responders). Given those two sets of users, the highest scoring asker collected a total points value of 16.431 whereas the maximum value for responders is 425.225.
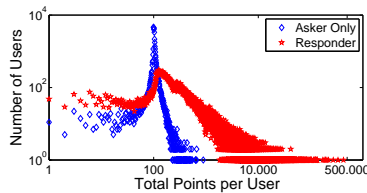
Fig. 4. Total points asker and responder.

This observation means that many users who respond to questions also tend to use the YA platform over a longer period of time since more points are accumulated over time. On the contrary, many users just try using the platform by signing up to YA and asking a single question (see the spike at 100 points in Fig. 4). The details regarding user participation over time are shown in Table 2 by analyzing the overlap of community members (based on their Q/A behavior) in consecutive years and within the total duration from 2005 to 2008.

Table 2. User participation over time.

	2005/2006	*2006/2007*	*2007/2008*	*2006/2008*
$sim(y_a, y_b)_A$	9%	4%	5%	1%
$sim(y_a, y_b)_R$	47%	31%	39%	12%
$sim(y_a, y_b)_{A \cap R}$	20%	5%	7%	1%

We use the following equations to obtain the entries in Table 2:

$$sim(y_a, y_b) = \frac{|U_{y_a} \cap U_{y_b}|}{|U_{y_a}|} \tag{1}$$

The first equation (Eq. 1) calculates the similarity of sets of users U_{y_a} and U_{y_b} between two years y_a and y_b (these need not to be consecutive years). Thus, Eq. 1 calculates the overlap of how many users participating in year y_a also participated in year y_b in YA. In other words, Eq. 1 denotes the containment of y_a in y_b. The results are shown in Table 2 as $sim(y_a, y_b)_A$ (for askers only) and $sim(y_a, y_b)_R$ (responders only). One can observer that the number of responders continuing using the platform is much higher than the number of askers. This fact is even more drastic by comparing the overlap of users between the years 2006 and 2008. In particular, only 1% of askers who signed up in 2006 also continued to ask questions in 2008, whereas 12% of responders kept using the platform also in 2008.

Second, we calculate the similarity $sim(y_a, y_b)_{A \cap R}$ of users who are askers *and* responders as follows:

$$sim(y_a, y_b)_{A \cap R} = \frac{|U_{y_a}^A \cap U_{y_a}^R \cap U_{y_b}^A \cap U_{y_b}^R|}{|U_{y_a}^A \cap U_{y_b}^R|} \tag{2}$$

The results are again shown in Table 2 (last row). The percentage of users who have both roles and continue using the platform in consecutive years is higher than for askers only but still much lower as compared to the set of responders.

This is especially true for the comparison 2006/2008 where we observe equal values of $sim(y_a, y_b)_A$ and $sim(y_a, y_b)_{A \cap R}$. Thus, users who ask and answer questions do not typically use the platform over a longer period of time.

3.3 Community Topics

The YA community is structured in *topics* helping users to quickly access a set of relevant questions and answers. Before designing a topic-sensitive expertise mining approach, it is useful to understand the preference of users to post in one or more topics (i.e., whether users post in different topics at all). Figure 5 visualizes the number of topics per user.
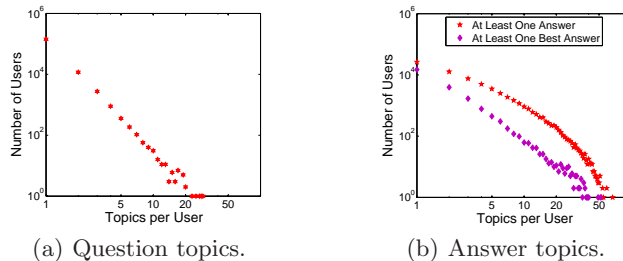

(a) Question topics. (b) Answer topics.

Fig. 5. Number of topics per user.

In Fig. 5(a), the number of topics by askers are shown. Since many users only ask one question (cf. previous section), the majority of askers is active in one topic only. The responders' behavior is shown in Fig. 5(b). We count the number of topics per user if the user provided at least one answer. Based on the results, we expect that a large fraction of responders typically post in 20-50 topics. However, by looking at those users who provided at least one best answer, about 20 topics by responder can be expected.

In the next step, we picked questions and answers in one particular year (2007) and analyzed the following properties: (i) which are the most discussed topics, (ii) how much overlap is between the sub-communities in terms of users posting in multiple topics, and (iii) the evolution and change of user interest. The results are illustrated in Fig. 6. First, we rank each topic based on the number of answers, which is a good indicator for community interest (see Top 20 topics shown in Table 6(c)). For each topic in Table 6(c), we calculate how similar the user base is compared with another top-ranked topic using the formula $sim(T_a, T_b) = \frac{|U_{T_a} \cap U_{T_b}|}{|U_{T_a}|}$. In other words, similarity is expressed by how many users post in both topics T_a and T_b. This measure is useful when comparing the impact and results of topic-sensitive expertise mining. If topics (sub-communities) have virtually now overlap in their user base, it does not bring any benefit to rank experts in joint topics. The similarities of (ranked) topics are shown in Fig. 6(a). Here, the elements are interpreted as follows: given a top-ranked topic, say 'Cancer' (Rank 1), the similarity to other topics is shown from the left to the right were the index corresponds to the rank in Table 6(c). The color shade of each element in the topic comparison matrix corresponds to the numerical similarity value (see the colorbar on the right).

(a) Topic similarities.

(b) Interest change.
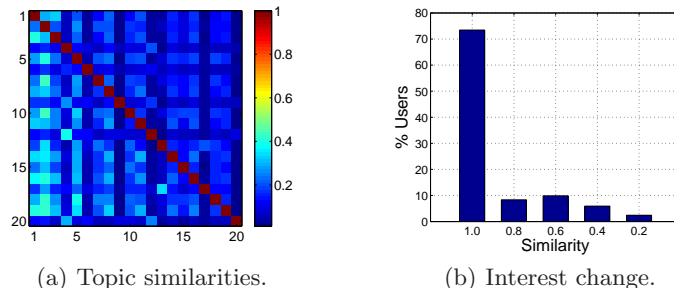
Rank	Topic	Rank	Topic
1	Cancer	11	Women's Health
2	Philosophy	12	Baseball
3	Heart Diseases	13	Men's Premiere Leagues (UK)
4	Football (American)	14	Men's Health
5	Words & Wordplay	15	Mathematics
6	Wrestling	16	Mental Health
7	Books & Authors	17	2006 FIFA World Cup(TM)
8	Diet & Fitness	18	Trivia
9	NASCAR	19	General Health
10	Astronomy & Space	20	Hockey

(c) Top 20 topics.

Fig. 6. Top answer topics in YA (2007) and interest evolution.

For all users who were active in 2007, we compared their interest evolution by analyzing user topics (in which topics users were active) in 2007 and 2008. We created a chart (Fig. 6(b)) that shows the percentage of users within the similarity intervals $[1, 0.8), [0.8, 0.6), [0.6, 0.4), [0.4, 0.2)$, and $[0.2, 0]$ according to the similarity value calculated using the definition of set overlap as discussed in the previous paragraph. Fig. 6(b) shows that 73% of users mostly post in the same topics they had posted in the past year. The remaining 27% have changing interests (to some degree) whereby a very low fraction of users have a high interest change (see most right category 0.2).

4 Topic-sensitive Expertise Mining in Q/A Communities

Here we propose link-based expertise mining techniques. Open Web-based communities are governed by changes such as people joining and leaving the community. It is therefore important to automate aspects such as the management of interest and expertise profiles due to scale and temporary nature of communities. We propose the application of well-established ranking methods used by search engines to estimate users' experiences based on the link structure of Q/A communities. The goal of this work is to establish a *relative order* of users based on their participation degree. The top-ranked users (i.e., responders) are considered to be authoritative experts who are recommended for answering questions regarding a specific topic.

Specifically, we propose the PageRank method [14], which can be personalized for different topics as discussed in [9]. Consider a graph structure as previously introduced (cf. Fig. 1(c)), where edges point from askers $\mathbf{A}$ to responders $\mathbf{R}$. An edge is thus established whenever $\mathbf{R}$ responds to a question from $\mathbf{A}$, regardless whether the answer was selected as best answer. The PageRank $PR(u)$ of a node u is defined as follows:

$$PR(u) = (1 - \alpha)p(u) + \alpha \sum_{(v,u) \in E} \frac{PR(v)}{\texttt{outdegree}(v)} \tag{3}$$

The equation consists of two parts. The first part $p(u)$ is called *personalization vector* that is used to assign preferences to certain nodes (we shall discuss shortly how this is done). Without any preferences or bias towards a particular node, a personalization vector $p = [\frac{1}{|N|}]_{|N| \times 1}$ is assumed. The second part of Eq. 3 can be regarded as u's relative standing (reputation) within the entire network. The factor α is typically set to 0.85 because a random surfer on the Web is assumed to follow six links $(1 - \alpha) = 1/6 = 0.15$ and 'teleports' with probability 0.15 to a randomly selected Web page.

Our approach is to assign preferences to users based on their answering behavior in a particular topic of interest. For example, if experts need to be discovered with regards to a topic T (see footnote[2]), then preferences are assigned to those users who have answered questions with respect to this topic T. Formally, the topic-sensitive personalization vector p is assigned as follows:

$$personalization_{topic} : (u, T) \mapsto \begin{cases} 0, & \text{if } u \text{ has not been active in } T \\ answers(u, T), & \text{otherwise} \end{cases}$$

The function $answers(u, T)$ returns the number of *all* answers of u in topic T. Answers may include best answers or even answers that were not useful at all (e.g., even thumb-down voted answers). Our hypothesis is that both answering behavior of users in a certain topic (i.e., personalization) and also a user's global (community wide) reputation are important.

Results. We performed three experiments in different topics and combinations thereof. All experiments are visualized in Fig. 7 (top-25 ranked users that are ordered based on their ranking scores). Ranking results have been obtained by using the personalized PageRank model as discussed before using the parameter $\alpha = 0.75$. If α was set to a higher value, ranking scores were biased too much towards the global reputation of a user. On the other hand, by setting α too low, also 'spammers' would be ranked high with, however, low community reputation. Each node in Fig. 7 is labeled by its position in the ranking result. The position of a node corresponds to an entry (# column) in Table 3. Only edges that connect nodes within the top-25 rankings results are shown. Also, to understand the quality of our results, we rank nodes according to (i) best answer count BAC (not topic sensitive) and topic-based best answer count BAC^T (the rank R is shown in parenthesis in Table 3).

---

[2] Notice, T could be a single topic or a joint topic with $T = \{T_a, T_b\}$.

This information is also encoded in the nodes's color and shape. The border of nodes is octagon-shaped and nodes have black fill color if nodes are ranked within the top-25 list by BAC and within the top-25 list by BAC^T. Nodes with fill color gray are only ranked in a high position by our method, but not by BAC or BAC^T. Nodes with dark borders and gray core are ranked high by our approach and also by BAC. Nodes with gray borders and a dark core are ranked high by our approach and also by BAC^T.
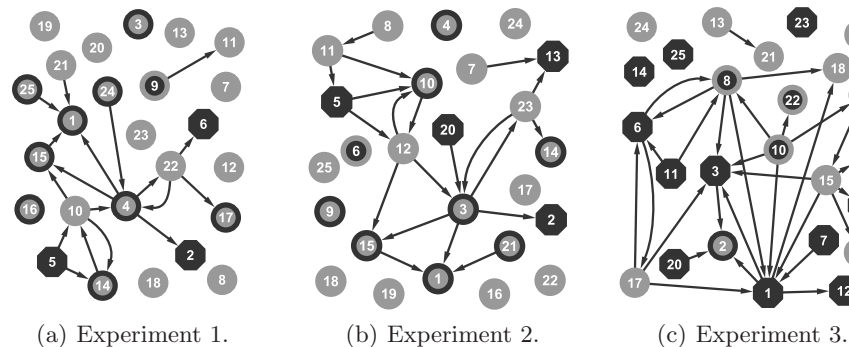

(a) Experiment 1.   (b) Experiment 2.   (c) Experiment 3.

Fig. 7. Topic-sensitive expertise mining (users in top-25 results).

Further information shown in Table 3 includes: AC answer count, number of topics T the user posts answers in, and IL^T depicting the number of inbound neighbors (askers) given a specific topic. For each experiment, we create 5 rows depicting the properties of the top-10 ranked nodes (reduced from top-25 to top-10 due to space limits).

Table 3. Experiment node properties.

#	AC	$BAC(R)$	T	IL^T	AC^T	$BAC^T(R)$	#	AC	$BAC(R)$	T	IL^T	AC^T	$BAC^T(R)$
1	2072	1219 (1)	3	n/a	n/a	n/a	6	595	275 (6)	12	5	5	5(2)
2	396	257 (9)	22	29	29	26(1)	7	172	8 (386)	42	8	8	1(2)
3	505	118 (14)	18	18	18	0(589)	8	302	11 (266)	43	6	6	0(686)
4	1328	851 (2)	6	1	1	0(918)	9	58	14 (190)	12	7	7	2(8)
5	445	89 (17)	22	10	10	3(5)	10	126	63 (27)	2	n/a	n/a	n/a
1	2072	1219 (1)	3	n/a	n/a	n/a	6	35	5 (637)	2	30	30	5(6)
2	396	257 (9)	22	60	60	50(1)	7	370	33 (57)	14	19	19	1(496)
3	1328	851 (2)	6	1	1	0(4268)	8	97	3 (1359)	23	26	26	0(3895)
4	505	118 (14)	18	38	39	1(396)	9	845	316 (4)	10	1	1	0(2821)
5	445	89 (17)	22	29	29	4(9)	10	365	258 (8)	2	n/a	n/a	n/a
1	1328	851 (2)	6	1167	1300	842(1)	6	365	258 (8)	2	338	354	252(2)
2	2072	1219 (1)	3	n/a	n/a	n/a	7	491	128 (13)	7	456	482	124(6)
3	618	258 (7)	3	441	570	240(4)	8	126	62 (27)	2	121	125	62(11)
4	595	275 (6)	12	362	544	247(3)	9	393	15 (175)	9	360	380	15(39)
5	845	316 (4)	10	445	465	163(5)	10	286	59 (28)	2	270	285	59(12)

The **first experiment** (see Fig. 7(a) and first segment of rows in Table 3) shows the ranking results within the topic *General - Health*, which we selected because health related issues are of general interest in YA (a top ranked topic). The first ranked user at position 1 would have been ranked at position 1 based on BAC as well. However, the user never responded to a question in the specific topic (thus, topic related info is not shown in the table – n/a). In this case, the user's community-wide reputation was predominant over the personalization. By using a parameter $\alpha = 0.65$ (giving more preference towards topic-based personalization), the order of users ranked at position 1 and 2 would have been changed. However, we noticed that many other globally low ranked (by BAC) users would be ranked too high. The results of experiment 1 show that 8 of 10 have been active in the given topic, 4 of 10 would have been ranked top-10 by BAC^T in the topic, and 4 of 10 would have been ranked top-10 by BAC.

The **second experiment** (see Fig. 7(b) and second segment of rows in Table 3) shows the ranking results within the topics *General - Health* and *Diet & Fitness*, which we selected because both topics have a good overlap of users. The visualization shows that adding another overlapping topic results in more edges being added among the top-25 ranked nodes. The same node as in experiment 1 is ranked at position 1. However, the user previously ranked at position 4 is now ranked at position 3. Notice, the node with label 3 is now connected with a set of other nodes and was therefore ranked at a higher position (i.e., the community embedding of the user improved by adding another topic). In experiment 2, 8 of 10 have been active in the given topics, 3 of 10 would have been ranked top-10 by BAC^T given both topics, and 5 of 10 would have been ranked top-10 by BAC.

The **third experiment** (see Fig. 7(c) and third segment of rows in Table 3) shows the ranking results within the topics *General - Health* and *Heart Diseases*, to test the impact of combining the previously selected topic with the highest ranked topic. Using these topics, the top-ranked user is now also top-ranked by BAC^T and also high ranked by BAC. The combination with a high answering amount topic results in even more connections between the nodes. In experiment 3, 9 of 10 have been active in the given topics, 6 of 10 would have been ranked top-10 by BAC^T given both topics, and 6 of 10 would have been ranked top-10 by BAC.

To conclude the discussion on topic-sensitive expertise mining, personalization based on topic activity combined with community-wide reputation delivers very good results. Thus, the automatic discovery of experienced (authoritative) users with respect to topics is possible by using the ***link-structure of Q/A communities***.

5 Personalized Recommendations

Due to the high number of questions that are typically asked in today's Q/A community, it becomes essential to provide recommendations (e.g., which questions should be answered). Our previous expertise ranking approach was based on the assumption that authoritative users also provide good answers, which we confirmed by comparing the rank of a user with the user's best answer count.

Our recommendation approach works in a manner similar. Instead of analyzing questions, we recommend users that ask 'good' questions. This approach can be used to order, for example, askers in a responder's buddy list.

However, before actually designing a recommendation algorithm, the *peer answering behavior* of users must be analyzed. If responders rarely (or never) answer questions from the same asker, it is obviously not useful to provide recommendations because users would not prefer to interact repeatedly. The results are shown by Fig. 8.
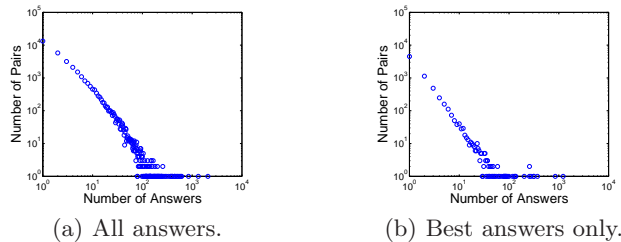

(a) All answers. (b) Best answers only.

Fig. 8. Peer answering behavior.

Here we take all edges in G and count the number of answers associated with $(v, u) \in E$. In Fig. 8, we show the number of asker/responder pairs and how often questions are typically answered by the same responder. On average, 6 answers are provided between the pair v, u (calculated as the weighted average). The average of best answers only is much lower (i.e., ≈ 0.4).

Notice, the objective is to recommend good askers to responders. To provide recommendations, we use essentially the same equation (cf. Eq. 3) as previously. What we do, however, is to change the direction of edges. For example, $(v, u) \in E$ becomes $(u, v) \in E'$ where the link inverted graph is defined as $G'(N, E')$. This means that nodes (users) are ranked by their *question authority* instead of their *answer authority*. In particular, those users who ask good questions are recommended. However, not all responders should receive the same recommendations but rather based on their previous interactions. Thus, recommendations should be given from each responder's individual point of view. Thereby, recommendations are calculated using Eq. 3 based on the inverted graph $G'(N, E')$ with $n \in N$ and the following personalization of the preference vector:

$$recommendation : (u) \mapsto \begin{cases} 1, & \text{if } n = u \\ 0, & \text{otherwise} \end{cases}$$

This means that for each node $n \in N$ personalized PageRank scores are created by setting the preference to n with $p(n) \leftarrow 1$. To evaluate our approach we use the standard precision and recall metrics.

Recall: A measure of the ability of the system to present *all* relevant items.

$$recall = \frac{|\{rel\_nodes\} \cap \{rec\_nodes\}|}{|\{rel\_nodes\}|}$$

The set $\{rel\_nodes\}$ holds those nodes that are actually relevant and the set $\{rec\_nodes\}$ those nodes that were recommended.

Precision: A measure of the ability of the system to present *only* relevant items.

$$precision = \frac{|\{rel\_nodes\} \cap \{rec\_nodes\}|}{|\{rec\_nodes\}|}$$

Results. We performed recommendations for the top-200 ranked users based on their answer count AC (see Table 4). We calculated recall and precision for recommendations for a reduced list of 25, 50, 100, 150, and the full list of 200 users. Recommendations are performed as follows: based on the year 2007 we calculate recommendations as discussed before. So for example, a node u would be recommended a set of users s/he should also interact in the following year 2008 (i.e., $\{rec\_nodes\}$ based on previous interactions). Notice, the set $\{rec\_nodes\}$ would contain *all* users within YA since recommendations are available for all nodes in the graph G'. We reduced the number of users in $\{rec\_nodes\}$ by populating the set with users based on the ranked (recommended) list with a cutoff at IL in 2007 (i.e., the number of inbound (askers) neighbors in 2007). We compare recommendation results with the actual neighbors (asker) in year 2008 to check whether recommendations were accurate. The parameter α was set to 0.2. In PageRank nomenclature, the restart probability of a random walk on the graph G' is set to $(1 - \alpha) = 0.8$; thus with high probability a random walk would be restarted at node u. The simplest approach to provide recommendations, however, is to recommend the exact same asker neighborhood from 2007 in 2008 again. These results are also calculated as *noRank* and compared against our approach *rank*.

Table 4. Recall and precision in ranking results.

	@25	*@50*	*@100*	*@150*	*@200*
Avg. IL (2007)	413	263	178	141	122
Avg. IL (2008)	426	284	199	163	143
Recall (noRank)	0.948	0.890	0.836	0.816	0.803
Precision (noRank)	1.000	1.000	1.000	1.000	1.000
Avg. RQ (noRank)	436	296	212	172	152
Avg. PQ (noRank)	426	284	199	163	143
Recall (rank)	0.944	0.883	0.827	0.804	0.789
Precision (rank)	0.995	0.992	0.988	0.984	0.980
Avg. RQ (rank)	437	297	214	174	153
Avg. PQ (rank)	427	285	200	164	145

The first two rows in Table 4 show the average IL neighborhood comparing the sets of users in 2007 and 2008. One can see a rising number IL which means that responders actually increased answering questions (more asker neighbors). This shows that recommendations are useful because responders increasingly answer questions of 'new' askers which they have not previously interacted with. In the next segment of Table 4, the average recall and precision of *noRank* recommendations are shown. These recommendations achieve surprisingly high

accuracy in terms of precision and recall. This means that responders tend to interact with the same askers also in 2008 compared to 2007.

The next two rows show the average recall quality RQ and the average precision quality PQ. Instead of calculating recall and precision as a percentage value, we calculated a weighted average that takes the number of IL neighbors of askers into account. For example, recall and precision of responders who are very active in terms of answering questions are given more weights since they contribute more answers to the community. The following rows show the results in terms of recall and precision of our recommendation approach (*rank*). Our approach results are approximately as accurate as the *noRank* results. However, by comparing the average RQ and PQ, our approach outperforms *noRank* recommendations.

Final remarks on personalization and performance issues. Considering non-personalized recommendations using the standard PageRank model with G' as input, we obtained an *average recall value of 0.011* and *average precision value of 0.014*. This approach is clearly not usable for recommendations because personalization greatly increases the accuracy of results (i.e., recall and precision of the recommended neighborhood of askers).

Personalized recommendations and topic-sensitive expertise rankings using Eq. 3 are much faster computed due to faster convergence of the ranking algorithm towards stable ranking results (i.e., the iteratively computed ranking scores do not change the order of nodes in the ranked result list). Computing non-personalized recommendations or expertise rankings takes a magnitude longer. For example in graphs with $|N| = 59000$ computing personalized results takes a couple of minutes whereas computing non-personalized results takes several hours. This leads us to the conclusion that personalization is achievable also in larger graphs. Other approaches such as the PageRank linearity theorem as proposed in [9] could be applied to compute personalized PageRank results even faster.

6 Conclusion

In this paper, we provided a comprehensive analysis of the structure and dynamics of Q/A communities with the focus on Yahoo! Answers – one of the most popular Q/A platforms on the Web. Our analysis was based on fundamental properties such as Q/A answering behavior of users over time. Also, we studied context-based properties of Yahoo! Answers by analyzing topic-sensitive answering behavior. Our analysis provides important insights for the design of topic-based expertise mining algorithms. Furthermore, we discussed experiments focusing on recommending users to responders whose questions may be relevant to answer. Our research has important implications for future crowdsourcing environments in terms of designing incentive mechanisms for user participation as well as algorithms to assist question routing to relevant experts in open Web-based communities. In our future work, we will study different rewarding models in *two-sided markets* and crowdsourcing with the aim of encouraging crowd users to solve more complex tasks instead of just answering simple questions.

References

1. Adamic, L.A., Zhang, J., Bakshy, E., Ackerman, M.S.: Knowledge sharing and yahoo answers: everyone knows something. pp. 665–674. WWW '08, ACM (2008)
2. Agichtein, E., Castillo, C., Donato, D., Gionis, A., Mishne, G.: Finding high-quality content in social media. In: WSDM '08. pp. 183–194. ACM (2008)
3. Artz, D., Gil, Y.: A survey of trust in computer science and the semantic web. J. Web Sem. 5(2), 58–71 (2007)
4. Becerra-Fernandez, I.: Searching for experts on the Web: A review of contemporary expertise locator systems. ACM Trans. Inter. Tech. 6(4), 333–355 (2006)
5. Brabham, D.: Crowdsourcing as a model for problem solving: An introduction and cases. Convergence 14(1), 75 (2008)
6. De Choudhury, M., Mason, W.A., Hofman, J.M., Watts, D.J.: Inferring relevant social networks from interpersonal communication. pp. 301–310. WWW '10, ACM (2010)
7. Golbeck, J.: Trust and nuanced profile similarity in online social networks. ACM Transactions on the Web 3(4), 1–33 (2009)
8. Gyongyi, Z., Koutrika, G., Pedersen, J., Garcia-Molina, H.: Questioning yahoo! answers. Technical Report 2007-35, Stanford InfoLab (2007)
9. Haveliwala, T.H.: Topic-sensitive pagerank. In: WWW '02. pp. 517–526 (2002)
10. Horowitz, D., Kamvar, S.D.: The anatomy of a large-scale social search engine. pp. 431–440. WWW '10, ACM (2010)
11. Jurczyk, P., Agichtein, E.: Discovering authorities in question answer communities by using link analysis. In: CIKM '07. pp. 919–922. ACM (2007)
12. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. J. ACM 46(5), 604–632 (1999)
13. Kumar, R., Lifshits, Y., Tomkins, A.: Evolution of two-sided markets. In: WSDM '10. pp. 311–320. ACM (2010)
14. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web. Tech. rep. (1998)
15. Schall, D., Dustdar, S.: Dynamic context-sensitive pagerank for expertise mining. pp. 160–175. SocInfo'10, Springer-Verlag (2010)
16. Skopik, F., Schall, D., Dustdar, S.: Modeling and mining of dynamic trust in complex service-oriented systems. Information Systems 35, 735–757 (2010)
17. Vázquez, A., ao Gama Oliveira, J., Dezsö, Z., Goh, K.I., Kondor, I., Barabási, A.L.: Modeling bursts and heavy tails in human dynamics. Physical Review E 73(3), 036127+ (Mar 2006)
18. Weng, J., Lim, E.P., Jiang, J., He, Q.: Twitterrank: finding topic-sensitive influential twitterers. pp. 261–270. WSDM '10, ACM (2010)
19. Yahoo: Yahoo! answers - home. (last access Apr 2011), `http://answers.yahoo.com`
20. Yahoo: Yahoo! answers scoring system. (last access Apr 2011), `http://answers.yahoo.com/info/scoring_system`
21. Zhang, J., Ackerman, M.S., Adamic, L.: Expertise networks in online communities: structure and algorithms. In: WWW '07. pp. 221–230. ACM (2007)
22. Ziegler, C.N., Golbeck, J.: Investigating interactions of trust and interest similarity. Decision Support Systems 43(2), 460–475 (2007)