

Knowledge Sharing and Yahoo Answers: Everyone Knows Something

Lada A. Adamic¹, Jun Zhang¹, Eytan Bakshy¹, Mark S. Ackerman^{1,2}

¹School of Information, ²Department of EECS

University of Michigan

Ann Arbor, MI

{ladamic,junzh,ebakshy,ackerm}@umich.edu

ABSTRACT

Yahoo Answers (YA) is a large and diverse question-answer forum, acting not only as a medium for sharing technical knowledge, but as a place where one can seek advice, gather opinions, and satisfy one's curiosity about a countless number of things. In this paper, we seek to understand YA's knowledge sharing activity. We analyze the forum categories and cluster them according to content characteristics and patterns of interaction among the users. While interactions in some categories resemble expertise sharing forums, others incorporate discussion, everyday advice, and support. With such a diversity of categories in which one can participate, we find that some users focus narrowly on specific topics, while others participate across categories. This not only allows us to map related categories, but to characterize the entropy of the users' interests. We find that lower entropy correlates with receiving higher answer ratings, but only for categories where factual expertise is primarily sought after. We combine both user attributes and answer characteristics to predict, within a given category, whether a particular answer will be chosen as the best answer by the asker.

Categories and Subject Descriptors

H.5.3 [Information Interfaces and Presentation (e.g. HCI)]: Group and Organizational Interfaces; J.0 [Computer Applications]: General

General Terms

Measurement, Human Factors

Keywords

Online communities, question answering, social network analysis, expertise finding, help seeking, knowledge sharing

1. INTRODUCTION

Every day, there is an enormous amount of knowledge and expertise sharing occurring online. One of the largest knowledge exchange communities is Yahoo! Answers (YA). Currently, YA has approximately 23 million resolved questions. This makes YA by far the largest English-language site devoted to questions and answers. These questions are answered by other users, without payment. Eckhart Walther

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.

WWW 2008, April 21–25, 2008, Beijing, China.

ACM 978-1-60558-085-2/08/04.

of Yahoo Research has claimed that “[YA is] the next generation of search... [it] is a kind of collective brain - a searchable database of everything everyone knows. It's a culture of generosity. The fundamental belief is that everyone knows something” [14]. Indeed, if there is something that someone knows, there is certainly ample opportunity to share it on YA.

Because of the sheer size of the YA community, and its breadth of forums, we wished to conduct a large scale analysis of knowledge sharing within YA. Knowledge sharing has been traditionally difficult to achieve, and yet YA appeared to have solved the problem, providing a society-wide mechanism by which to bootstrap knowledge and perhaps collective intelligence [6].

In short, we found YA to be an astonishingly active social world with a great diversity of knowledge and opinion being exchanged. The knowledge shared in YA is very broad (in several senses) but generally not very deep. In this paper, we examine YA's diversity of questions and answers, the breadth of answering, and the quality of those answers. Accordingly, we analyze the YA categories (or forums), using network and non-network analysis, finding that some resemble a technical expertise sharing forum, while others have a different dynamics (support, advice, or discussion). We then use the concept of entropy to measure knowledge spread based on a user's answer patterns across categories. We find that having lower entropy, or equivalently, higher focus, correlates with the proportion of best answers given in a particular category. However, this is only true for categories where requests for factual answers dominate. Finally, we examine answer quality and find that we can use replier and answer attributes to predict which answers are more likely to be rated as best.

First, however, we discuss the prior literature and describe YA.

2. PRIOR WORK

Sharing knowledge has been a research topic for at least 15 years. At first, it was largely studied within organizational settings (e.g., Davenport and Prusak [4]), but now Internet-scale knowledge sharing is of considerable interest. This knowledge sharing includes repositories (including those socially constructed as with Wikipedia [8]) as well as online forums designed for sharing knowledge and expertise. As mentioned, these forums promise—and often deliver—being able to tap other users' expertise to answer all sorts of questions— from mundane and everyday questions to complex and expert ones.

In general, there is a large body of literature examining on-line interaction spaces, especially Usenet. Four perspectives were important for this study. The first attempts to understand different forums (or newsgroups in Usenet). Whittaker et al. [22] conducted an insightful quantitative data analysis on a large sample of Usenet newsgroups, uncovering the general demographic patterns (i.e. number of users, message length, and thread depth). Interesting findings in their work included the highly unequal levels of participation in newsgroups, cross-posting behaviors across different newsgroups, and a common ground model designed to explore relations between demographics, conversational strategies, and interactivity.

This line of research also used social network analysis to examine forums. For example, Kou and Zhang [25] used network analysis to study the asking-replying network structure in bulletin board systems and found that people's online interactions patterns are highly affected by their personal interest spaces. Fischer et al. [7] and Turner et al. [18] developed visualization techniques to observe various interaction patterns in Usenet groups. These visualization techniques have been very helpful in understanding the big picture of these large online interaction spaces.

While the work above mostly focused on the forum level, there have also been studies focusing on the user level. Wenger [19] discussed the importance of different roles in online communities and how they affect community formation and continuation. Nonnecke & Preece [15] studied lurker behavior in different online forums. Donath [5] explored techniques to mine users' virtual identities and detect deception in online communities. Recently, Welsler et al. [20] argued that one can use users' ego-networks as "structural signature" to identify "discussion persons" and "answer persons" in online forums. This work described role differences in online communities and provided insights on how to analyze user level data. However, the work lacks a strong quantitative basis.

There has also been work focusing on the thread and message level. For example, Sack [16] used visualization to show that there are various conversation patterns in discussion threads. Using message level content analysis, Joyce and Kraut [2, 9] studied whether the formulation of a newcomer's post and related responses influenced whether they continue to participate.

Besides studying the conversation patterns in online communities, researchers have also focused on understanding why people participate in and contribute to online communities. This work has usually been based on small scale data collection and surveys (e.g., Lakhani and von Hippel [11] and Butler et al. [3]), and has informed our study by delineating possible reasons why users engage in different activities in YA.

In our own work, we have been studying one kind of online forum, online expertise sharing communities – those spaces devoted to answering one another's technical questions. We analyzed a technical question answering community (Java Forum) and explored algorithms using network structure to evaluate expertise levels in [23]. Using simulations, we explored possible social settings and dynamics that may affect the interaction patterns and network structures in online communities [24]. The goal of these studies was to design better systems and online spaces to support people in sharing knowledge and expertise in the Internet age.

During the course of our studies, we realized that relatively little is known about extremely large scale knowledge sharing and expertise distribution through online communities. YA presents an excellent place to study this problem because of its breadth of topic and high level of participation. More importantly, YA is a space that was designed for the sole purpose of knowledge sharing, although as we will see, it is used for much more. To our knowledge, there have been only two studies examining YA to date. Su et al. [17] used YA's answer ratings to test the quality of human reviewed data on the Internet. Kim et al. [10] studied the selection criteria for best answers in YA using content analysis and human coding. This has left open both the need for a large scale systematic analysis of YA, and the opportunity to study the depth and breadth of direct knowledge sharing from several perspectives that are only visible in such a large space.

3. YAHOO ANSWERS AND DATA SET

The format of interaction on YA is entirely through questions and answers. A user posts a question, and other users reply directly to that question with their answers. On YA, questions and their answers are posted within categories. YA has 25 top-level and 1002 (continually expanding) lower level categories. The categories range from software to celebrities to riddles to physics to politics. There are some "fact"-based threads, such as the following from the Programming & Design (Programming) category. In this thread, a user asks for information on how to read a file using the C programming language¹.

Q: How to read a binary file in C ?
I want to know what function from which header I must use to read a binary file. I will need to know how big a file is in byte. Then I want to move N byte into a char * variable.

She garners two responses. One is:

use the function fopen() with the last parameter as "rb" (read, binary).

The other, selected as the "best answer" by the asker, is more detailed:

```
#include <stdio.h>
FILE *fp;
fp = fopen("Data.txt", "rb");
fseek(fp, 0, SEEK_END);
filesize = ftell(fp);
rewind(fp);
fread(DataChar, 5000, 1, fp);
fclose(fp);
```

This is a typical level of depth and complexity of the questions and answers for the Programming category. Indeed, many questions and their answers on YA are relatively simple. For example, math and science categories appear to be dominated by high school students seeking easy solutions to their homework. Not all categories are strictly focused around expertise seeking, however. The following question is from the Cancer category and appears to be soliciting both help and support:

"My uncle was recently diagnosed with some rare cancer and does not have medical insurance. He has tried to apply for medical but has been denied. He does not have much money because he had to quit his job because he is getting too weak. Who can help him?"

¹We anonymized any identifying data and reworded the questions slightly for publication

This question received 10 answers, including a pointer to the local cancer society office. On average, a question in the Cancer category receives 5.2 replies, and only 6% go unanswered. What is surprising is just how much of the interaction in YA is in fact pure discussion, in spite of the question-answer format. There are many categories where questions are asking for neither expertise nor support, but rather opinion and conversation. For example, in the celebrity category, one finds the following question:

Who is the better actress, Angelina Jolie or Jennifer Aniston?

This question has appeared at least twice, once garnering 33 answers and once garnering 50. While one might expect a large question-answer forum to show a more diverse range of behavior than a narrowly focused software forum, we were nevertheless surprised to see the full range of topic and user types previously seen in general online newsgroups [7].

It is important to note that these discussions are constrained by the setup of the YA system for technical expertise sharing with a strict question and answer format. Threads must still start with a question. YA users discuss by answering the question, not by addressing one another. Furthermore one cannot answer more than once nor can one answer oneself, making Usenet-type discussions difficult. This clearly changes the thread interactions relative to other online systems.

In order to study the characteristics and dynamics of YA in a systematic manner, we harvested one month of YA activity. The dataset includes 8,452,337 answers to 1,178,983 questions, with 433,402 unique repliers and 495,414 unique askers. Of those users, 211,372 both asked and replied. These numbers are already a hint to the diversity of user behavior in YA. Many users make very few posts. Even those who actively post will sometimes reply without asking much, while others do the opposite. These behaviors will vary by YA category, so we will briefly describe our analysis of those categories first.

4. CHARACTERIZING YA CATEGORIES

4.1 Basic characteristics

Based on an initial examination of YA, we expected that every category would have some mix of requests for factual information, advice seeking, and social conversation or discussion. While it would be difficult to determine the precise mix for each category without reading the individual posts, we can indirectly infer the category type by observing characteristics such as average thread length (the number of replies per post) and average post length (how verbose the answers are). Figure 1 shows a scatter plot of such data, with several categories highlighted.

We observe that factual answers on technical subjects such as Programming, Chemistry, and Physics will tend to attract few replies, but those replies will be relatively lengthy. In fact, all of the math and science subcategories have a relatively low answer-to-question ratio, from 2 answers per question in chemistry to 4 answers per math question. Astronomy has a higher question-answer ratio at 7, due to occasional questions about extraterrestrial travel and life that garner many replies (e.g. “What will you think if NASA comes clean about UFOs?” attracted 21 answers in 3 hours). In fact, the one science subcategory that stands out starkly is Alternative Science with 12 replies on average per question. These questions deal with the paranormal and by their

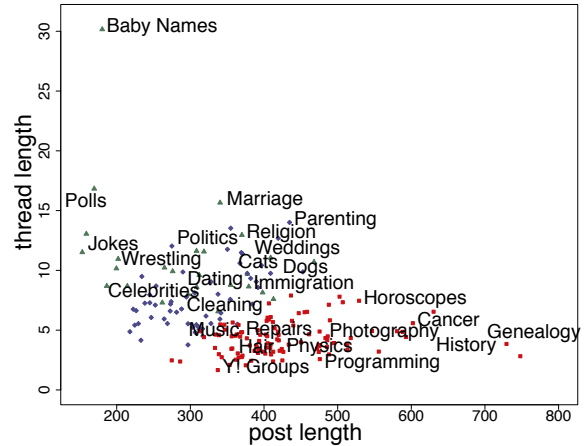


Figure 1: Thread length vs. post length, with some categories labeled.

very nature can lead to long discussions. (A typical question might be: “Can you use a RMS Multimeter for Ghost Hunting?”)

On the other extreme are categories with many short replies. The Jokes and Riddles category contains many jokes whose implicit question is “Is this funny?” Most of the replies are short, “hahaha. that’s funny” or “I’ve heard that one before”. Also in this corner of the figure is the category Baby Names, where threads center around brainstorming and suggestions of names, and many users chime in (24 people per question on average).

We can recognize discussion categories, those attracting many replies of moderate length: sports categories like Wrestling, as well as other categories such as Philosophy, Religion, and Politics. Also among those categories attracting many replies of moderate length are topics where many individuals have some experience and advice is sought. These include Marriage & Divorce Marriage and several parenting-related categories for newborns, toddlers, grade schoolers, and adolescents. The Cats and Dogs categories generate fairly long threads of moderate reply lengths as well.

Another distinguishing characteristic for categories is the asker/replier overlap: whether the people who pose questions are also the ones who reply. In a forum where users share technical expertise, but the majority of askers are novices, one might expect that the population of askers and repliers is rather distinct [18]. Those who have expertise will primarily answer, while those who do not have it will be posing the majority of the questions. In a forum centered on advice and support, users may seek and offer both, becoming both askers and repliers. In a discussion forum, both asking and replying are ways of continuing the conversation. It is therefore unsurprising that the technical categories have a lower overlap in users who are both askers and repliers, while the discussion forums have the highest overlap. We will revisit this question in Section 5.1.

4.2 Cluster analysis of categories

We calculated several aggregate measurements for each category. The activity in each category ranged from 216,061 questions in Singles & Dating, to 129,013 questions about

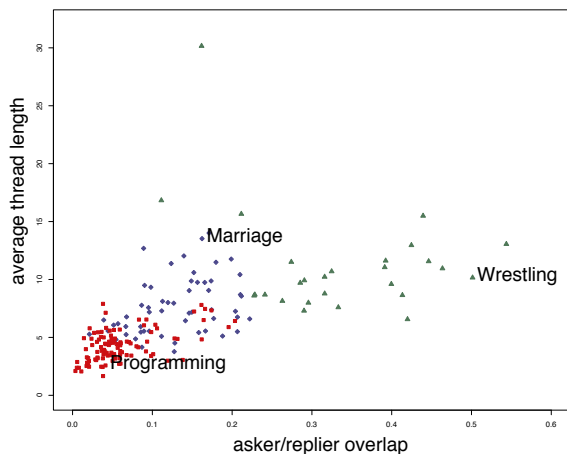


Figure 2: Clustering of categories by thread length and overlap between askers and repliers

Religion and Spirituality, 48,624 Mathematics questions, and only 5 questions on Dining Out in Switzerland. We classified the most active categories (> 1000 posted questions) using k-means clustering on three primary metrics: thread length, content length, and asker/replier overlap. The thread length for a category is given by the average number of responses for each answered question. Content length is given by the average number of characters in all responses within a category. The asker/replier overlap is the cosine similarity between the asking and replying frequency for each user. The analysis considers 189 categories, which together constitute over 91% of all questions posed on YA.

We find that clustering the categories into three groups yields a result we find the most intuitively meaningful. Figure 2 shows how these three clusters are distributed according to thread length and asker/replier overlap.

The first cluster consists of discussion forums (green triangles in Figure 2), having a high proportion of users who both pose and answer questions. In these categories, users discuss likely winners in various sports categories, squabble over partisan issues in Politics, or debate the true nature of a god in the Religion & Spirituality category. These kinds of stimulating questions tend to attract long thread lengths.

The second cluster (blue diamonds) consists of categories in which people both seek and provide advice and common-sense expertise on questions where there may be several legitimate answers or no single factual answer. Perhaps because there is rarely a definitive answer, and at the same time many feel qualified to give advice, the threads tend to be long. This cluster includes the categories Fashion, Baby Names, Fast Food, Cats, and Dogs.

In the third cluster (red squares), we observe categories where many questions have factual answers, e.g. identifying a spider based on markings. People tend to either ask or reply, and thread lengths tend to be shorter. These categories include Biology, Repairs, and Programming.

In next section, we examine the question-answer dynamics further by analyzing how network structure differs in representative categories for each of these clusters. This more carefully considers how expertise and knowledge is arranged and structured in YA.

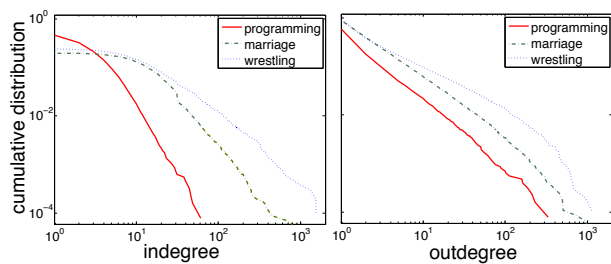


Figure 3: Distributions of indegree (number of users one has received answers from) and outdegree (number of users one has answered)

4.3 Network structure analysis

By connecting users who ask questions to users who answer them, we can create an asker-replier graph; we call these QA networks. The analysis of QA networks sheds light on important aspects of interaction that are not easily captured by non-network measures. In this section, we examine three categories whose social dynamics are typical of the three clusters: Wrestling, Marriage & Divorce (Marriage), and Programming & Design (Programming).

4.3.1 Degree distributions

Figure 3 shows the number of people one has answered (outdegree) and the number of people one has received replies from (indegree) for categories corresponding to each of the three clusters. From these figures we can see that the users differ in their activity level in all three categories. Some answer many questions, others merely stop by to ask or answer a question or two. On the other extreme there were users who asked or answered dozens of questions. Second, we can also see that there are differences among these three categories. Although all three categories display heavy tailed distributions, Marriage and Wrestling have much broader indegree distributions, with a few people receiving thousands responses in the one month sample considered in this study. In contrast, the most active users posing questions in Programming initiated threads garnering only a few dozen replies.

In general, forums in Yahoo answers tend to have broad outdegree distributions. In the Programming category this reflects a few highly active individuals who consistently help others with their tasks and problems, but do not necessarily ask for help themselves. In the Marriage category, these could be users who regularly offer advice, or are there for the fun of discussion, as is the case in the Wrestling category. Note that this separation of roles is evident even when one considers whether a user posted a single question or answer. For instance, in Programming, about 57% of the users who asked questions did not answer any during this time period, and similarly 51% who answered questions did not ask. As seen in Figure 2, of the three categories, wrestling has the most significant overlap in asking/replying activity, followed by Marriage and Programming.

4.3.2 Analysis of ego networks

Welser et al. [20] suggested that one can distinguish an “answer person” from a “discussion person” in online forums by looking at users’ ego networks. Each ego network consists

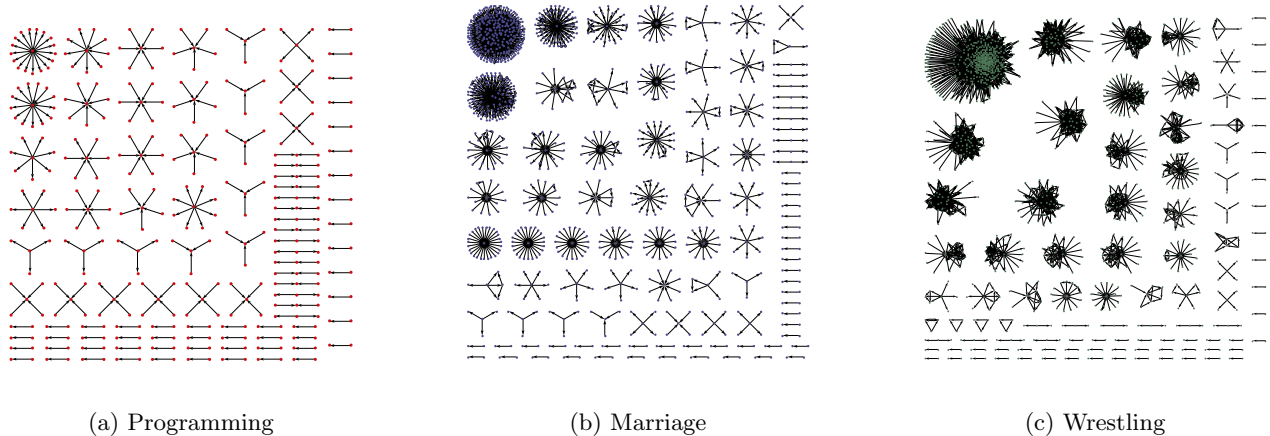


Figure 4: Sampled ego networks of three selected categories

Table 1: Summary statistics for selected QA networks

Category	Nodes	Edges	Avg. deg.	Mutual edges	SCC
Wrestling	9,959	56,859	7.02	1,898	13.5%
Program.	12,538	18,311	1.48	0	0.01%
Marriage	45,090	164,887	3.37	179	4.73%

of the user, the ties to other users the person interacts with directly, and interactions between those users. Thus, one can examine what types of users appear in different categories. Figure 4 shows the ego networks of 100 randomly sampled users from the three categories.

From this figure, we can see that the neighbors of some of the highly active users in Wrestling are themselves highly connected, which indicates that they are more likely to be “discussion persons”. On the contrary, in the Programming category, the most active users are “answer people” because most of their neighbors, the people they are helping, are not connected [20].

4.3.3 Strongly connected components

Given that some people reply almost exclusively, and others ask almost exclusively, it is unclear whether these categories contain giant strongly connected components (SCCs). Strongly connected components represent those sets of users, such that one user can be reached from any other, following directed edges from asker to replier. A large SCC indicates the presence of a community where many users interact, directly or indirectly. Table 1 gives the sizes of the SCCs and other general statistics of the networks of the three selected categories.

From this table, we can see, consistent with the degree distributions shown above, that the Wrestling category is more connected. More importantly, it has a strongly connected component and a relatively large number of mutual edges (two users who have answered each other’s questions), which indicates that there may be a core social group forming in this category. There is almost no strongly connected component in Programming (even a random network of this size and density should have a modestly sized SCC), and

reciprocal edges are entirely absent. We believe that this is due to the separation of roles of “helpers” and “askers” in the Programming category. The Marriage category lies in-between. The proportion of mutual edges is small, but not zero, and the giant component is small, but not absent. We delve into this further in the next section.

4.3.4 Motif analysis

Motif analysis allows one to discover small local patterns of interaction that are indicative of particular social dynamics. Here we focus on all possible directed interactions between three connected users within a forum. Figure 5 displays the motif profiles of the three selected categories, showing, for example, how often interactions are reciprocal (the asker becomes the replier for another question) and how often the triads are complete (three users who have all replied to one another). These profiles are constructed by counting the actual frequency of each triad in the QA network for that category, and then comparing that frequency against the expected frequency for randomized versions of the same network [13, 12, 21].

From Figure 5, we can see that all three categories have a significantly expressed feed forward loop (see triad 38 in the figure) compared to random networks. In this motif, a user is helped by two others, but one of the helpers has helped the other helper. The motif, most pronounced in the Programming category, indicates a common characteristic in help-seeking online communities where people with high levels of expertise are willing to help people of all levels, while people of lower expertise help those with even less expertise than their own [23].

As well, we can see that both the Wrestling and Marriage categories have a high number of fully reciprocal triads, indicating symmetric interaction. Another triad that is significant in these two categories involves two users who have replied to one another (who may be regulars in the forum) and have also replied to a third user, perhaps someone who is just briefly joining the discussion to ask a question. Interestingly, the triad of two users who have replied to one another, and have also both received replies from a third user, is not significant for Programming; it would imply that the regulars who have had a chance to reply to one another’s

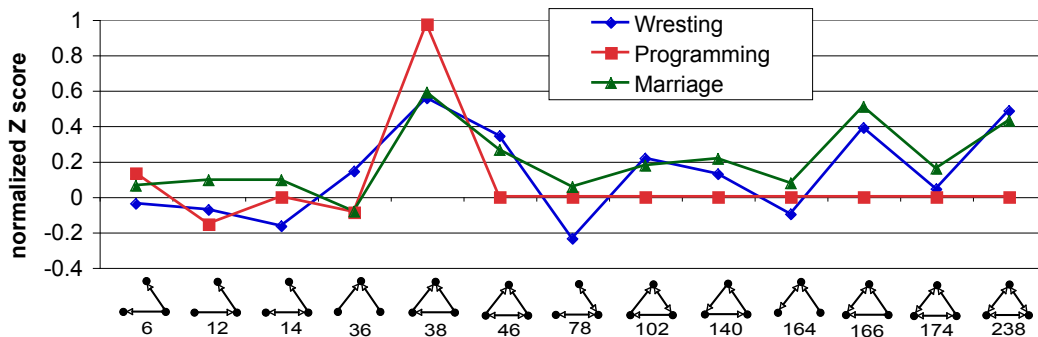


Figure 5: Motif profiles of selected categories

programming questions are drawing answers from less active users. It is a significant motif, however, for both Wrestling and Marriage. Perhaps there even questions posed by regulars are of an inviting nature.

4.4 Expertise depth

Is expertise being shared? We have already alluded to the relative simplicity of many questions. It often seems as though users are sharing the answers to one another’s homework questions. To determine the depth of the questions asked in YA, we rated 100 randomly selected questions from the Programming category. We rated these questions into 5 levels of expertise (as discussed in [23]). In this rating scheme, level 3 expertise is that of a student with a year’s experience in a programming topic, for example, someone who could pull details from an API specification. A level 4 expert, on the other hand, would be a professional programmer, someone with experience in implementation or deployment issues and their effects on design (such as compiled Java applications and their speed). We found only one question (1%) in the Programming category that required expertise above level 3. In short, the questions are very shallow. This is not a definitive test, of course, but it indicates that YA is very broad but not very deep. We explore that breadth in the next section.

5. EXPERTISE AND KNOWLEDGE ACROSS CATEGORIES

Given the wide variety of behavior and interests in the different forums, we saw an opportunity to describe how knowledge and expertise are spread across different domains. In this section, we describe the breadth of YA from two perspectives. The first considers the extent to which users who are actively answer questions in one category are also likely to do so in another. The second measures users’ entropy, namely the breadth of topics their answers fall in.

5.1 Relationships between categories

By tracking answer patterns, it is easy to discern related categories, shown in Figure 6(a), where people who answer questions in one category are likely to answer questions in related categories. Computer-centric categories, including Computers & Internet, Consumer Electronics, Yahoo! Products, and Games & Recreation (dominated by questions about video and online games), are all clustered together. Similarly, Politics and Government is linked to News and

Events, while the Home and Garden category is linked to Food and Drink, which is in turn linked to Dining Out, which is in turn linked to the topic of Local Businesses. The above cross-category correlations suggest a focus of interest on the part of the users.

Reply patterns only reveal the topics that a user feels comfortable discussing. The overlap of asking and replying patterns, on the other hand, indicates whether people who reply regarding one topic are likely to ask questions in the same topic or another. In Figure 6(b), we can observe that users are likely to post both questions and replies in the same forum, if that forum deals with topics that are prone to discussions: Sports, Politics, and Society & Culture (including Religion). Topics dominated by straightforward factual questions, such as those found in the Education & Reference and Science & Math subcategories, have a smaller percentage of users who both seek and offer help. Most users almost exclusively either ask for help (as mentioned, many apparently looking for easy answers to their homework questions), or provide help without posing questions of their own.

Other interesting patterns emerge when one looks at question/ answer patterns across categories. As a silly hypothetical example, consider users who answer many car repair questions, but may need a bit of advice about beauty and style. As amusing as it would be to find this connection, we find that those posting answers about cars and transportation tend not to ask for help in other categories, as much as people answering in other categories asked for help with cars. In fact, sports and politics were the only other large categories from which the helpers were less likely to be the ones asking questions about beauty and style.

No matter the category that users post answers in, they almost uniformly also ask about Yahoo products, including YA itself. Health was a category that many users asked questions in, no matter where else they answered. But it was also a category that many offered help in. The latter was also true of Family & Relationships, but asking questions about relationships typically did not correlate with answering in other categories. There was again an asymmetry between technical and support categories: people who answered in Relationships, Health, or Parenting tended to ask in the Computers & Internet category, while the opposite was not true: those answering in Computers & Internet did not have a high proportion of questions in Health, Relationships, or Parenting.

The above connections between categories are apparent because at least some users are not replying in all categories

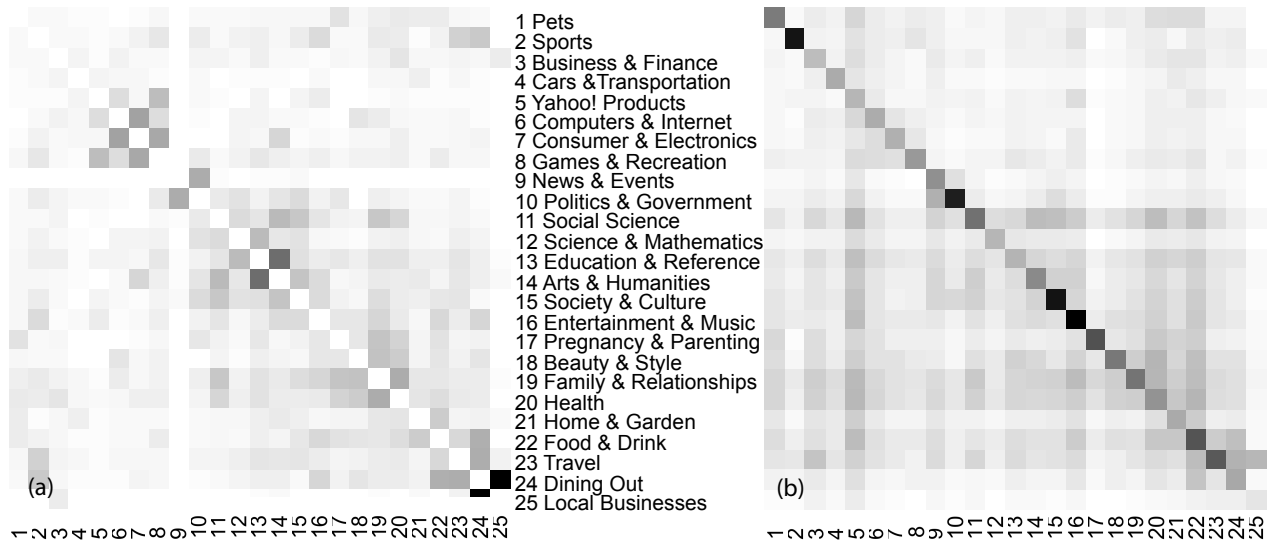


Figure 6: Similarities between categories: a) overlap in users who replied in both categories, ordered using hierarchical clustering b) overlap in users who answered in one category (rows) and asked in another (columns). A cosine similarity was used in both, but the shades correspond to different scales.

at random; they have a certain degree of focus. So while YA gives the opportunity to individuals to seek and share knowledge on a myriad of different topics, any individual user is likely to only do so for a limited range. In the next section, we will turn to studying YA on the individual level in order to pinpoint just how broad users' participation is.

5.2 User entropy

We sought a measure that would capture the degree of concentration in a person's reply patterns to particular topics. Entropy is just such a measure – the more concentrated a person's answers, the lower the entropy, and the higher the focus. We also wanted our entropy measure to capture the hierarchical organization of the categories, such that a user who answers in a variety of subcategories of the same top level category would have a lower entropy than someone who answered in the same number of subcategories, but with each falling into different top level categories.

$$H_L = - \sum_i p_{L,i} \log(p_{L,i})$$

Figure 7 illustrates a hypothetical user's distribution of questions. To obtain the total entropy for a user, we first calculate the entropy H_L for each level separately.

$$H_T = \sum_L H_L$$

The users' apparent breadth depends in part on the extent of their activity in terms of the number of posted answers. This activity level varies considerably among users, as we observed in section 4.3.1. In order to discern whether users are truly focused on just a few topics, or simply had not been active enough to reveal their full range of interests, we selected just the 41,266 users who had posted at least 40 replies in the month of our crawl. Among those users, we can observe a range of entropies. For one user, who describes herself as a dog trainer who shows shelties at dog shows, we find that all her answers are in the Dog subcategory.

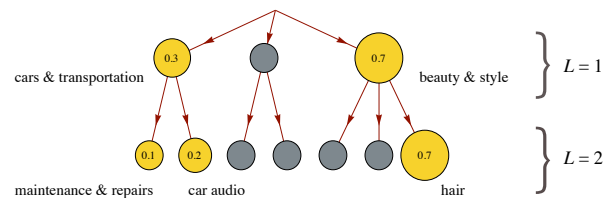


Figure 7: illustration of the hierarchical entropy calculation: $H_1 = 0.3 * \log(0.3) + 0.7 * \log(0.7) = 0.61$, $H_2 = -0.2 * \log(0.2) - 0.1 * \log(0.1) - 0.7 * \log(0.7) = 0.81$, and $H_T = H_1 + H_2 = 1.42$

Therefore her entropy is 0. On the other end of the spectrum is a user whose 40 questions are scattered among 17 of the 25 top-level categories and 26 subcategories. He posted no more than 4 answers in any one category and his combined 2-level entropy is 5.75.

Figure 8(a) shows the entropy distribution of all users who posted 40 or more questions. The distribution is surprisingly flat. It is not the case that only a few users are very diverse. Rather, some users have a very low entropy, but higher entropies are relatively common, until one encounters a limit in terms of the number of possible categories that are specified by the YA hierarchy.

We also examined the proportion of best answers by users. (Again, best answers are those answers rated as such by the asker or voted as such by YA users.) This distribution, shown in Figure 8(b), is skewed, with a mode around 6-8% best answers. Some users obtain much higher percentages of best answers. In the next section we will correlate the two metrics applied to users in order to determine whether being focused corresponds to greater success in having one's answers rated as best.

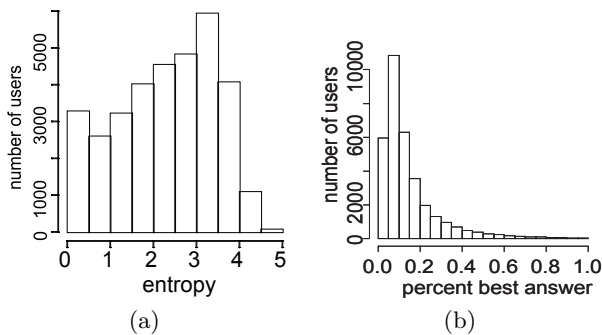


Figure 8: The distribution of (a) entropy and (b) proportion of best answers for users who had answered at least 40 questions.

5.3 Correlating focus to best answers

Intuitively, one might expect that users who are focused to a limited range of topics tend to have their answers selected as best more frequently. For example, a dog trainer/breeder who answers questions about dogs may be expected to have a higher proportion of best answers because all of her answers are focused on her specialty. Interestingly, we found no correlation between the total entropy of a user across all categories and their overall percentage of best answers ($\rho = -0.02, p < 10^{-3}$). Users do not provide better answers (at least according to their best answer count) when they specialize. The value of the correlation has the correct sign (more scattered users have a lower proportion of best answers), but is only significant because of the large number of users.

While it may well be the case that posting answers in several discussion forums does not correlate with whether others like those answers, we still expected to see a correlation in some cases. From our earlier examination of the different categories, we know that only some topics reflect requesting and sharing factual information. This brings to question what the criteria for best answer selection are in other forums. In support forums, the best answer may be the one with the most empathy or most caring advice. In a discussion forum, the best answer may be the one that agrees with the askers’ opinions, while for entertainment categories, the wittiest reply may win. A previous study that sampled users comments upon selecting a best answer to their question found that content value (such as accuracy and detail) was used in selecting the best answer in just 17% of the cases, compared to 33% for socio-emotional value, including agreement, affect, and emotional support[21].

Another idiosyncrasy of selecting just one best answer, instead of rating individual ones, is that there may be several good answers, but only one is selected. In a preliminary analysis, we randomly sampled 100 questions each from categories of Programming, Cancer and Celebrity and coded them according to how well they answered the question. We found that replies selected as best answers were indeed mostly best answers for the question. For those best answers not rated as the best answer by us, we found that they could still be second or third best answers. This beneficial glut of good answers means that even if a user always provides good answers, we may not be able to discern this, because their good answer will not always be selected as best, depending

Table 2: Entropy within a category and % best

level 1 category(ies)	Pearson $\rho(\text{entropy}, \text{score})$
computers & internet	-0.22***
science and math	
family & relationships	-0.13***
sports	-0.01

Table 3: Correlation between focus and % best

moderate	low	none
$\rho > 0.1$ ***	$0.05 < \rho < 0.1$ **	$\rho < 0.05$
physics	programming	marriage&divorce
chemistry	gardening	wrestling
math	dogs	alternative medicine
biology	hobbies&crafts	religion & spirituality
Y! products	cooking&recipes	baby names

on how many other replies were posted. This means that users focusing on categories with a high answer-to-question ratio will on average have a lower best answer percentage, and any correlation between user attributes and this percentage will be weakened by the noise introduced through answers being pitted against one another for first place.

Despite these caveats, we still expected lower entropy to be correlated with performance for categories where many questions were of a technical or factual nature. To verify this claim we computed separate second level entropies, shown in Table 2, for several first level categories. Indeed, for the technical categories of Computers & Internet and Science & Math, we find a significant correlation between the users’ entropy within those top level categories and their scores. The correlation is weaker, but still present for the advice-laden category of Family & Relationships. It is absent in the discussion category of Sports.

Finally, we used a very simple measure, the proportion of a user’s answers in the category, and correlated it with a user’s proportion of best answers in that category across all of YA. We found that for technical categories, focus tends to correlate with better scores. For categories that still require some domain knowledge to answer questions, there was a weaker, but significant correlation. Lastly, in discussion categories, there was no relationship between focus and score within that category. A listing of typical categories for each level of correlation is shown in Table 3. Note the predominance of a single cluster corresponding to low asker-replier overlap and short thread length for the categories where correlation between focus and score is highest.

6. PREDICTING BEST ANSWERS

So far, we have observed distinct question-answer dynamics in different forums. We have also observed a range of interests among users – some focusing quite narrowly on a particular topic, while many participate in several forums at once. Furthermore, focusing on a particular category (having low entropy) only correlated with obtaining “best” ratings for one’s answers in categories where questions centered on factual or technical content. Here, we test our ability to predict whether an answer will be selected as the best answer, as a function of several variables, some of which will correspond closely with our previous observations. A com-

Table 4: Predicting the best answer

	Programming	Marriage	Wrestling
reply length	+ ***	+ ***	+ ***
thread length	- ***	- ***	- ***
user # best ans.	+ ***	+ ***	+ ***
user # replies	- ***	- ***	- ***
prediction accuracy	0.729	0.693	0.692

+ (positive coefficient), - (negative coefficient)
 *($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$)

plementary, and concurrent, study of question and answer quality was performed by Agichtein et al. [1].

We constructed randomly selected balanced sets of answers that were and were not chosen as best answers. We excluded those instances where the answer was the only answer, which would make it very likely to be selected as best. We then ran a logistic regression on a number of variables.

We omitted entropy and focus measures because the majority of users had posted too few replies to produce meaningful entropy values. We ran a logistic regression to predict whether an answer would be selected as best, and performed a ten-fold cross-validation to obtain a prediction accuracy, with a baseline of 0.5 for random guesses.

Table 4 summarizes the prediction results for the three categories from the category clusters. For all categories, the length of the reply and the number of other answers the asker had to choose from were the two most significant features. We can achieve about 62% prediction accuracy across all three categories based on answer length alone – showing a preference by the asker for receiving lengthier replies. Figure 9 shows the difference in length distribution for best answers and non-best answers in the Programming category. Also important is the track record of the user, in terms of the number of other answers posted within the category, and how many were selected as best. This feature proved more predictive for the Programming category than for either the Marriage or Wrestling categories. Interestingly, the number of best answers users provide outside of the category is not significant, once their track record within the category is taken into account. The simple number of replies (a user’s activity level) improves the odds of an answer being selected a best answer only slightly; and once the number of best answers by the user is taken into account, the coefficient actually becomes negative to reflect a higher number of non-best answers given by the user.

Our results for Yahoo Answers stand in stark contrast to our previous analysis of Sun’s Java Forum, where the number of previous replies strongly correlated with the expertise level as judged by independent human raters. Here, we see that when the raters are the askers themselves, there is a preference for longer answers, but not always by more active repliers. It would of course be interesting to pit the askers’ choice of best answer against best answers selected by experts in the subject. It would also be interesting to examine whether frequency of replies correlates with expertise level, and even whether there is as much of a differentiation in expertise level on a general community such as Yahoo Answers, as opposed to a specialized community such as the Java Forum. We leave these and other questions for future work.

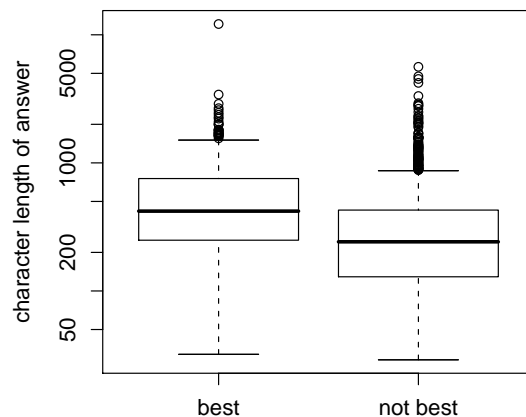


Figure 9: Difference in length between answers selected as best by the asker or other users, and those that were not selected.

7. CONCLUSIONS

Yahoo Answers is a large and diverse question answer community, acting not only as a medium for knowledge sharing, but as a place to seek advice, gather opinions, and satisfy one’s curiosity about things which may not have a single best answer. One may dispute the validity of the knowledge in Alternative Science and even the degree of knowledge in Celebrities. However, the YA participants believe this is knowledge, and they are certainly exchanging it.

We took advantage of the range of user behavior in YA to inquire into several aspects of question-answer dynamics. First, we contrasted content properties and social network interactions across different YA categories (or topics). We found that we could cluster the categories according to thread length and overlap between the set of users who asked and those who replied. Discussion topics or topics that did not focus on factual answers tended to have longer threads, broader distributions of activity levels, and their users tended to participate by both posing and replying to questions. On the other hand, YA categories favoring factual questions (what are usually called question-answer forums) had shorter thread lengths on average and users typically did not occupy both a helper and asker role in the same forum. We found differing interaction motifs in the question-answer networks corresponding to these distinct dynamics. Consistent with prior work on online forums, we found that the ego-networks easily revealed YA categories where discussion threads, even in this constrained question-answer format, tended to dominate.

Second, we identified related categories, by asking whether a user who answers questions in one category is also likely to answer in another. We found many expected relationships between categories about related topics, but also some interesting asymmetries when linking asking questions in one category with answering questions in another. Many users answered questions about familiar topics such as Family & Relationships, no matter where they tended to ask their questions. On the other hand, users who answered in specialized, technical categories, such as Car Maintenance &

Repair or Computers & Internet, asked fewer questions in other categories, where the users they were helping predominantly supplied answers.

This led us to examine the range of knowledge that users share across the many categories of YA. We found that while many users are quite broad, answering questions in many different categories, this was of a mild detriment for specialized, technical categories. In those categories, users who focused the most (had a lower entropy and a higher proportion of answers just in that category) tended to have their answers selected as best more often.

Finally, we attempted to predict best answers based on attributes of the question and the replier. Our results showed that just the very basic metric of reply length, along with the number of competing answers, and the track record of the user, was most predictive of whether the answer would be selected. The number of other best answers by a user, a potential indicator of expertise, was predictive of an answer being selected as best, but most significantly so for the technically focused Programming category.

In future work we would like to further examine the level of expertise being shared on YA. By democratizing knowledge sharing, YA has accomplished a large feat – everyone knows something, and through our analysis, we know that many know even several things and can share them on YA. But it remains unclear whether depth was sacrificed for breadth. We would like to know whether different incentive mechanisms could encourage YA participation by top level experts – who may currently still prefer more specialized, boutique forums – while at the same time allowing the rest of us to get our everyday, simple questions answered.

8. ACKNOWLEDGEMENTS

We would like to thank Mark Newman for formulating the entropy measure used in this paper. We also acknowledge Intel Research, the Army Research Institute, and the National Science Foundation (0325347) for their financial support of this work.

9. REFERENCES

- [1] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne. Finding High-Quality Content in Social Media. *WDSM'08*, 2008.
- [2] J. Arguello, B. S. Butler, L. Joyce, R. Kraut, K. S. Ling, and X. Wang. Talk to me: foundations for successful individual-group interactions in online communities. In *CHI'06*, pages 959–968, 2006.
- [3] B. Butler. Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures. *Information Systems Research*, 12(4):346–362, 2001.
- [4] T. Davenport and L. Prusak. *Working Knowledge: How Organizations Manage What They Know*. Harvard Business School Press, 1998.
- [5] J. S. Donath. Identity and deception in the virtual community. *Communities in Cyberspace*, pages 29–59, 1999.
- [6] D. Engelbart and J. Ruilifson. Bootstrapping our collective intelligence. *ACM Computing Surveys (CSUR)*, 31, 1999.
- [7] D. Fisher, M. Smith, and H. Welser. You Are Who You Talk To: Detecting Roles in Usenet Newsgroups. In *HICSS'06*, 2006.
- [8] T. Holloway, M. Bozicevic, and K. Börner. Analyzing and visualizing the semantic coverage of wikipedia and its authors: Research articles. *Complexity*, 12(3):30–40, 2007.
- [9] E. Joyce and R. Kraut. Predicting Continued Participation in Newsgroups. *Journal of Computer-Mediated Communication*, 11(3):723–747, 2006.
- [10] S. Kim, J. S. Oh, and S. Oh. Best-Answer Selection Criteria in a Social Q&A site from the User-Oriented Relevance Perspective. *presented at ASIST*, 2007.
- [11] K. Lakhani and E. von Hippel. How open source software works: “free” user-to-user assistance. *Research Policy*, 32(6):923–943, 2003.
- [12] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon. Superfamilies of evolved and designed networks. *Science*, 303:1538–1542, 2004.
- [13] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network Motifs: Simple Building Blocks of Complex Networks. *Science*, 298(5594):824–827, 2002.
- [14] Y. Noguchi. Web searches go low-tech: You ask, a person answers. *Washington Post*, page A01, 2006.
- [15] J. Preece, B. Nonnecke, and D. Andrews. The top five reasons for lurking: improving community experiences for everyone. *Computers in Human Behavior*, 20(2):201–223, 2004.
- [16] W. Sack. Conversation map: a content-based Usenet newsgroup browser. In *IUI'00*, pages 233–240, 2000.
- [17] Q. Su, D. Pavlov, J. Chow, and W. Baker. Internet-scale collection of human-reviewed data. In *WWW'07*, pages 231–240, 2007.
- [18] T. Turner, M. Smith, D. Fisher, and H. Welser. Picturing Usenet: Mapping Computer-Mediated Collective Action. *Journal of Computer-Mediated Communication*, 10(4), 2005.
- [19] E. Wegner. *Communities of Practice: Learning, Meaning, and Identity*, 1998.
- [20] H. T. Welser, E. Gleave, D. Fisher, and M. Smith. Visualizing the signatures of social roles in online discussion groups. *Journal of Social Structure*, 8(2), 2007.
- [21] S. Wernicke and F. Rasche. FANMOD: a tool for fast network motif detection. *Bioinformatics*, 22(9):1152–1153, 2006.
- [22] S. Whittaker, L. Terveen, W. Hill, and L. Cherny. The dynamics of mass interaction. *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 257–264, 1998.
- [23] J. Zhang, M. Ackerman, and L. A. Adamic. Expertise networks in online communities: structure and algorithms. In *WWW'07*, pages 221–230, 2007.
- [24] J. Zhang, M. S. Ackerman, and L. A. Adamic. Communitynetsimulator: Using simulations to study online community networks. In *C & T'07*, 2007.
- [25] K. Zhongbao and Z. Changshui. Reply networks on a bulletin board system. *Phys. Rev. E*, 67(3):036117, Mar 2003.