Deconstructing Interaction Dynamics in Knowledge Sharing Communities

Ablimit Aji and Eugene Agichtein

Mathematics and Computer Science Department
Emory University
{aaji,eugene}@mathcs.emory.edu

Abstract. Online knowledge sharing sites have recently exploded in popularity, and have began to play an important role in online information seeking. Unfortunately, many factors that influence the effectiveness of the information exchange in these communities are not well understood. This paper is an attempt to fill this gap by exploring the dynamics of information sharing in such sites - that is, identifying the factors that can explain how people respond to information requests. As a case study, we use Yahoo! Answers, one of the leading knowledge sharing portals on the web with millions of active participants. We follow the progress of thousands of questions, from posting until resolution. We examine contextual factors such as the topical area of the questions, as well as intrinsic factors of question wording, subjectivity, sentiment, and other characteristics that could influence how a community responds to an information request. Our findings could be useful for improving existing collaborative question answering systems, and for designing the next generation of knowledge sharing communities.

Keywords: *Social media, Collaborative Question Answering*.

1 Introduction

Asking questions and contributing answers can be an effective way of sharing information and expertise. Online, this way of knowledge sharing manifests itself in the form of Collaborative (or Community) Question Answering (CQA) sites, such as Naver, Baidu Knows, Live QnA, and Yahoo! Answers. In the U.S., Yahoo! Answers attracted more than 100 million users, and has a growing archive of more than 400 million answers to questions (2008 estimates). Already, for many information needs, these sites are becoming valuable alternatives to search engines.

Previous studies of CQA focused on the ultimate outcome of the knowledge sharing activity (e.g., the quality of the finally contributed answers). However, as responses and ratings arrive, the perceived quality of an item may change significantly. Our goal is to understand the factors that influence the *dynamics* of the interactions in knowledge sharing communities - that is, to understand *how* this content is generated, and which aspects of the process can affect the resulting content quality. As far as we know, there has been no prior scientific study of how different question characteristics influence the interaction dynamics in CQA. We attempt to fill this gap by systematically exploring the contextual and intrinsic characteristics of the questions posted in CQA, and the effects of these factors. Specifically, we address the following research questions:

– *How does the question context influence the community response?*
– *What intrinsic question qualities influence the community response?* For this, we examine how the question aspects such as subjectivity, sentiment, conversational orientation and writing quality affect the dynamics of community responses.
– *How do question context, wording, and response dynamics effect the quality and timeliness of the answers obtained?*

We present, to the best of our knowledge, the first large-scale empirical study of temporal dynamics for collaborative question answering to explore the factors that can explain the CQA interactions.

In addition to better understanding the dynamics and behavior of collaborative question answering communities, our results could have practical applications including better real-time CQA content ranking for search engines, and more accurate and timely filtering for high quality content. More generally, our findings can be useful to improve existing collaborative question answering systems, and for designing the next generation of knowledge sharing communities.

2 CQA Overview and Description

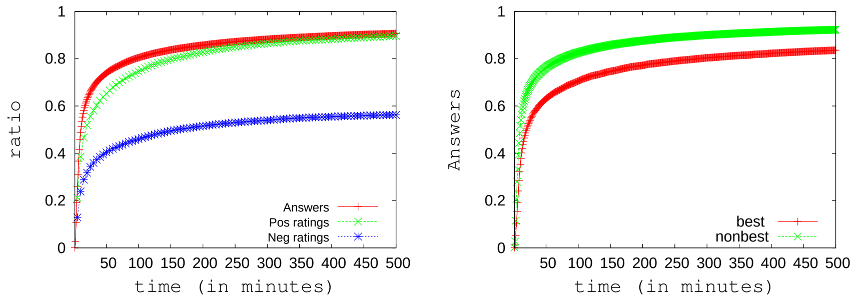
We first briefly review the Community Question Answering setting. A user (asker) selects an appropriate topical category, and posts a question, optionally. Newly posted questions appear in the "Open Question" list for each category, reverse ordered by the time they posted. At this point, other users can answer this question, or can rate the already posted answers. If the asker is satisfied with any one of the submitted answers, he or she can choose it as "best" answer. If the asker has not closed the question during the "Open Question" period, the "best" response chosen by votes from other users.

2.1 Data Collection and Statistics

To obtain the data for our study, we repeatedly crawled the Yahoo! Answers using the provided API, to capture the arrival of user contributions nearly real time. For this, we tracked a total of approximately 10,000 questions, sampled from 20 categories, over each of the questions' lifetimes. Specifically, for each category, after a new question appears on the "Open Questions" list, we begin tracking it (up to 200 questions per run per category) every five minutes until the question is closed or moved to the "Undecided Question" status. As a result, we obtained approximately 22 million question-answer-feedback snapshots in total, capturing feedback and contribution dynamics for a representative sample of CQA. We refer to this dataset as *AnswersTemporal*. Ultimately (after filtering and cleaning), the *AnswersTemporal* dataset consists of the interaction data for 9,747 questions and for the 47,780 answers posted for these questions.

2.2 Temporal Dynamics of CQA

In popular CQA sites, most questions stay on the front page of the respective category for only a few minutes. These questions tend to receive many answers initially, but the rate of arrivals of answers and user ratings decreases over time (Fig 1(a)). In contrast, user feedback ratings continue to arrive hours or even days after the question and answers have been posted. We conjecture that as the questions and answers are indexed by the web search engines, additional users view the questions and the answers, but tend to rate the content instead of contributing new answers. Interestingly, answers eventually chosen as “Best” by the asker appear to arrive somewhat later than non-best answers (Figure 1 (b)). That is, 56% of the eventual “Best” answers arrive within first 30 minutes, compared to 70% for Non-Best answers (a statistically significant difference over our large sample of questions). Having discussed the overall patterns of answer arrival, we now delve into more detailed analysis of the *content* of the questions.



(a) (b)

Fig. 1: Answer and rating arrival time, averaged across all categories, for the first 500 minutes of question lifetime (a), and arrival rate for “Best” vs. “Non-Best” answers for the first 500 minutes after posting (b).

3 Factors behind CQA Dynamics

In this section we describe the factors that could influence interaction dynamics, namely the question context, the intrinsic characteristics of the questions, and answer quality.

Question Context: Yahoo! Answers has 26 top level categories and more than 1000 lower level subcategories. These categories differs from each other in topical focus and community demographics. Based on previous work [1], we expect question category (and more generally, a forum chosen) to have significant effects on content creation speed and quality.

Intrinsic Question Characteristics: We now define the intrinsic question dimensions, drawn from the literature to be important aspects of textual communication.

– **Subjectivity**: CQA sites increasingly attract users interested in obtaining opinions of other users about social norms, preferences, and popularity. As has been suggested in previous work (e.g., [2] and references therein), question subjectivity can significantly affect the quality, quantity, and the tone of the answers posted by the community. Our hypothesis, is that question subjectivity will also influence interaction dynamics such as the rate of answer and ratings arrival, and other factors such as agreement of the asker rating (“stars”) with community ratings (“thumbs”).
– **Conversational Orientation**: Another important dimension, distinct from subjectivity, is whether a question is a fact seeking information or is simply a start of a conversation. Note that conversational questions may in fact be objective (e.g., “What

are the problems with [Windows] Vista?"), thus this factor is different from subjectivity [3]. Our hypothesis is that this dimension will also have a measurable effect on interactions and behavior of the community.
– **Sentiment**: An important feature that distinguishes CQA content from other web content is that both questions and answers can carry a sentimental orientation (that is, have a negative or positive connotation). For example, a question ("How stupid must you be to wear shorts and flipflops when it's cold?") makes a clear negative statement. We expect such questions to attract answers at different rates and of different quality than questions that are positive or neutral.
– **Quality**: Previous research showed strong correlation between question quality and answer quality [4]. We hypothesize that question quality would have an effect on the arrival of best answer (and popular answers in general) as questions that are well stated and interesting should attract good answers faster.

Answer Quality: Ultimately, the goal of CQA is to obtain information (answers) for the asker's information needs. Thus, quality and timeliness of answers obtained are perhaps the most important "output" variables of a CQA system. We also examine the effects of the contextual and intrinsic question factors above on the answer quality, as rated by both the asker and the community.

4 Experimental Setup

To analyze interaction dynamics, we first report on the large-scale manual labeling of question characteristics, and then describe the metrics used to analyze the responses.

4.1 Manual Labeling

To obtain human judgements, we utilized the Amazon Mechanical Turk (MTurk) service. Briefly, MTurk provides the infrastructure for "workers" to select Human Intelligence Tasks (HITs) that can be, for example, questions for which we wish to obtain human judgments regarding sentiment. Our typical HIT would include 10 questions, each with the text and the context (category) provided to the rater. Five workers rated each question for the various dimensions described above; the ratings were filtered by using majority opinion (that is, picking the label chosen by at least three out of the five Turk workers). From the 2000 initially labeled questions, 1570 remained after filtering, which we consider reliable human ratings. We refer to this dataset as *QuestionsLabeled*, and report the annotator agreement in Table 1. This dataset can be downloaded, by request, from `http://ir.mathcs.emory.edu/shared/sbp2010/`.

	Orientation	*Subjectivity*	*Question Quality*	*Sentiment*	*Answer Quality*
Agreement	0.7765	0.7396	0.701	0.647	0.682

Table 1: Annotator agreement for question and answer labels

4.2 Asker Satisfaction vs. Popularity

To complement the understanding the dimensions of the questions, it is important to examine the perceived quality of the contributed answers. Intuitively, the most important factor for an answer should be whether the asker considered it to be the best answer for the question. However, we observed that this selection does not always agree with the community ratings. To quantify this discrepancy, we used the traditional measure from information retrieval, Mean Reciprocal Rank (MRR), computed by ranking the answers in order of decreasing "Thumbs up" ratings, and identifying the rank of the actual "best" answer, as selected by the asker. More precisely:

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{N}\frac{1}{rank_i}$$

where $rank_i$ is the rank of the best answer among the answers submitted for question i.

5 Results: Effects of Question Factors on CQA Dynamics

This section reports the effects of both contextual and intrinsic question characteristics on content creation, namely answer and rating arrival and answer quality.

5.1 Contextual Factors: Question Category

Question category influences many aspects of CQA dynamics, such as time to first answer, time to closing a question, and answer quality [4]. In particular, as Figure 2 shows, there is a large difference in answer arrival rates across categories. For example, in the top level category "Health," which is relatively popular, nearly 65% of the answers arrived in the first 10 minutes, whereas only 20.4% arrive within the first 10 minutes for the category "Travel".

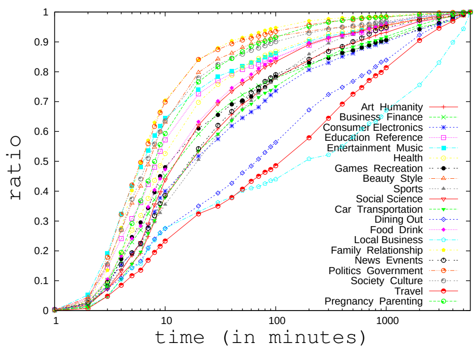


Fig. 2: Answer arrival patterns for categories

Category Name	MRR	Answer
Local Business	0.46	2.96
Yahoo Products	0.46	4.28
Travel	0.43	4.86
Science and Mathematics	0.41	4.27
Computer and Internet	0.40	4.11
Consumer Electronics	0.39	3.84
Entertainment and Music	0.33	9.74
Family and Relationship	0.30	8.32
Beauty and Style	0.30	6.67
Pregnancy and Parenting	0.29	9.06
Average	0.37	6.4

Table 2: Average MRR values (asker vs. community agreement) for selected categories

Interestingly, the agreement between the asker and the community (modeled with the MRR metric defined above) also varies significantly across question categories, as shown in Table 2. In general MRR is higher for categories with shorter average thread length (which would make high MRR more likely), but is not always the case (eg., Beauty & Style vs. Entertainment & Music). This indicates that in some categories community consensus is more difficult to achieve than in other categories, which would also indicate higher rate of subjective or conversational questions.

5.2 Effects of Intrinsic Question Factors

Question Subjectivity: We expect subjectivity of a question to be less dominant than category but still related to other important factors such as average answer arrival and

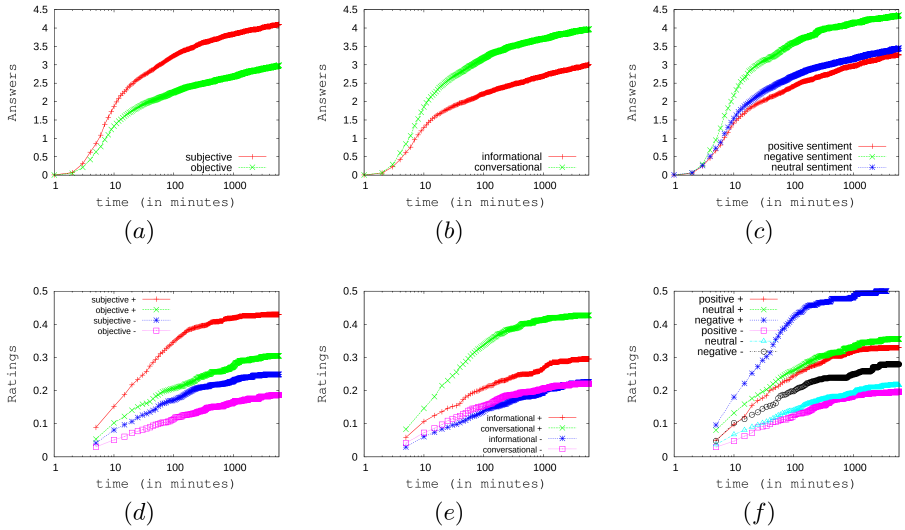


Fig. 3: Effects of Question Characteristics on Answer and Vote Dynamics: answer arrival (a) and answer ratings (d) grouped by subjectivity; answer arrival (b) and answer ratings (e) grouped by informational orientation; and answer arrival (c) and answer ratings (f) grouped by question sentiment.

MRR. In particular, more answers arrive for subjective questions than for objective questions (Figure 3 (a)), and user ratings exhibit a similar skew towards subjective questions (Figure 3 (d)). Interestingly, answer ratings for subjective question arrive faster than ratings to user answers to objective questions. This is more noticeable for positive user ratings.

Conversational Orientation: Answer arrival and rating arrival for questions with different conversational vs. informational orientations are reported in Figure 3 (b) and (e), respectively. We observe that first answers for conversational questions arrive faster than for informational questions. This is not surprising since conversational questions are more common and users are willing to answer conversational questions compared to more difficult factual or informational ones. However, user ratings for informational questions suggest a different dynamic. Positive user ratings for conversational questions arrived faster than for informational questions, while *negative* user ratings for both conversational and informational questions arrive at the same rate.

Question Sentiment: Recall that we hypothesized that question's sentiment would affect the answer arrival rate as well as the ratings. Figure 3 reports answer arrival dynamics (c) and rating arrival dynamics (f) for these different question types. Answers to negative questions arrive significantly faster than answers to positive and neutral questions; interestingly, positive *ratings* arrive much faster to negative questions, whereas positive and negative ratings arrive roughly at the same rate for positive and neutral questions. Based on manual examination of the examples, we conjecture that this effect is caused by the selection bias of the raters participating in negative question threads, who tend to support answers that strongly agree (or strongly disagree) with the asker.

Answer quality: We now consider the effects of all the question factors on answer quality. As reported in previous work [4], question quality indeed moderately correlates with answer quality (Pearson $p = 0.23$), and answer length correlates more strongly with answer quality (Pearson $p = 0.45$). However, we were surprised to find that there is weak or no correlation between question subjectivity, orientation, or sentiment and answer quality. That is, while conversational questions tend to elicit more participation from others, the overall quality of the contributed content does not exhibit significant differences (we omit the detailed results for lack of space). Interestingly, CQA participants respond differently to high quality vs. low quality answers (Figure 4). Surprisingly, there is no difference between the number of positive "thumbs up" ratings for high vs. low quality answers, but there *is* a significant difference in the number of the negative "thumbs down" ratings: in other words, it is the negative ratings (or lack thereof) that can more reliably separate high quality answers from low-quality answers.

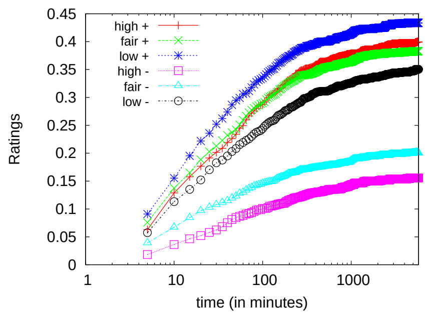


Fig. 4: Positive (+) and Negative (-) rating arrival for high, fair, and low-quality answers.

Type	MRR	Answer
Informational	0.43	3.90
Conversational	0.39	4.80
Subjective	0.38	4.91
Objective	0.44	3.85
High Quality	0.40	4.37
Low Quality	0.42	4.52
Positive	0.41	4.10
Neutral	0.41	4.40
Negative	0.38	5.00

Table 3: Agreement(MRR) of asker's choice of best answer with the community for different question types.

Asker vs. Community Agreement: Finally, we consider the Asker and community agreement for different types of questions (Table 3). The most noticeable differences confirm our hypotheses: askers of informational questions tend to agree with the most popular answers more often than conversational (MRR of 0.43 vs. 0.39), and askers of objective questions tend to agree with community more than askers of subjective questions (MRR 0.44 vs. MRR 0.38). Interestingly, question quality and question sentiment do not appear to have significant influence on asker's and community agreement.

6 Related Work

One of the main goals of CQA is to enable the exchange of high-quality, relevant information between community participants. Finding such quality information, where in QA communities quality varies significantly, provides a unique challenge, which recently has been addressed in references [4], [5]. This previous work treated CQA content as static and no attempt was made to classify content while it is still being updated/rated.

References [6, 7] introduced more fine-grained models of individual user actions generating content in blogs and other social media. Our study is also related to recent work by Harper et al. [3] which consider some of the similar features for automatic classification of questions in CQA into informational or conversational, but in a static, off-line setting. In contrast to these efforts, our work attempts to recover the underlying factors for content generation where we attempt to exploit factors such as question subjectivity. Other related work focuses on the temporal evolution of the social media or web graph structures, such as [8] that analyzes the temporal evolution of the wikipedia graph; [9] predicts controversy of Slashdot posts based on social network and discussion structure. Closest to our work, Adamic et al. [10] examined the category-centric variations in link structure in the Yahoo! Answers community. Additionally, Leskovec et al. [11] developed models for microscopic evolution, including one for Yahoo! Answers. In contrast, our work focuses on *understanding* the content generation dynamics, that appear to be more influenced by question characteristics than structural properties.

7 Conclusions

We presented the first, to our knowledge, large-scale study of the underlying factors influencing the dynamics of participation in knowledge sharing communities. In addition to confirming previous findings on a new dataset (e.g., the effects of each community or forum category), other more subtle characteristics such as the question sentiment, subjectivity, and informational orientation all influence the arrival of answers, ratings, and agreement between the "popular" answers and the ones chosen by the asker. These findings could be useful both for improving existing CQA systems and for designing the next generation of collaborative information sharing environments.

References

1. Liu, Y., Bian, J., Agichtein, E.: Predicting information seeker satisfaction in community question answering. In: SIGIR. (2008)
2. Li, B., Liu, Y., Agichtein, E.: CoCQA: Co-Training over questions and answers with an application to predicting question subjectivity orientation. In: Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNL). (2008)
3. Harper, F.M., Moy, D., Konstan, J.A.: Facts or friends?: distinguishing informational and conversational questions in social qa sites. In: CHI '09: Proceedings of the 27th international conference on Human factors in computing systems. (2009) 759–768
4. Agichtein, E., Castillo, C., Donato, D., Gionis, A., Mishne, G.: Finding high quality content in social media. In: Proc. of WSDM. (2008)
5. Jeon, J., Croft, W., Lee, J., Park, S.: A framework to predict the quality of answers with non-textual features. In: Proc. of SIGIR. (2006)
6. Hogg, T., Szabo, G.: Diversity of user activity and content quality in online communities. In: Proc. of ICWSM. (2009)
7. Goetz, M., Leskovec, J., Mcglohon, M., Faloutsos, C.: Modeling blog dynamics. In: Proc. of ICWSM. (2009)
8. Buriol, L.S., Castillo, C., Donato, D., Leonardi, S., Millozzi, S.: Temporal analysis of the wikigraph. In: Proceedings of the International Conference on Web Intelligence (WI). (2006)
9. Gómez, V., Kaltenbrunner, A., López, V.: Statistical analysis of the social network and discussion threads in slashdot. In: Proc. of WWW. (2008)
10. Adamic, L.A., Zhang, J., Bakshy, E., Ackerman, M.S.: Knowledge sharing and yahoo answers: everyone knows something. In: WWW '08: Proceeding of the 17th international conference on World Wide Web. (2008) 665–674
11. Leskovec, J., Backstrom, L., Kumar, R., Tomkins, A.: Microscopic evolution of social networks. In: Proceeding of KDD. (2008)