

3M1-OS-07c-3

Who are the archivists?: Understanding the role of reputation, seniority, and power in social Q&A

Andrew W. Vargo*¹

Shigeo Matsubara*²

*¹ Kyoto University

*² Kyoto University

Determining the archival value of information in Question and Answering (Q&A) forums is often a difficult and costly undertaking. Most current research focuses on how Q&A submissions gain page views or appeal to third party reviewers. The advantage of this approach is that it signals transcendent archival value. What it overlooks is the intrinsic archival value to the community itself since the Q&A community is the initial archivist. In this paper, we look at how these archives are created by the users of a reputation and voting system, specifically the system of Stack Overflow. We find that users with higher reputation points are more likely to vote negatively, regardless of their identity type within the system.

1. Introduction

Aggregated point based reputation systems in online communities are a relatively new construct in the human experience. In the past, reputation could be gained or lost with the ebbing and flowing of inherited and given titles and positions, or earned by accomplishments within a social system. The exact standing of any entity is ambiguous and depends largely on who is considering the question. The most trusted investment bank by institutional traders may have earned low standing with private citizens. A way for online social systems to remove ambiguity from understanding reputation is by having the members of the community give and take reputation points for certain actions.

In the same way, determining archival value is a difficult thing to achieve. Again, in the past, we have relied on qualified people to determine what is saved, made accessible, and even promoted for general cultural knowledge. In the web, archiving can be done a number of ways. For instance, stewards can have paid workers on Mechanical Turk archive text and pictures, Facebook has a like button that indicates popular material, and YouTube largely uses views as a type of currency to determine interesting and worthwhile material.

A problem arises when trying to determine the archival value of information which has exacting technical merit from a Question and Answer session. It is too expensive and impractical to use third party annotators, and while views of Q&A session might be an indication of archival value, it might not give enough information about what part of the session contains the value. Community voting is one way to solve this impasse. This mechanism allows

In Stack Overflow (SO), we can view crossing point between the two concepts of reputation and social archiving. This crowd-sourcing Q&A site allows members to vote for content, with votes directly affecting the reputation of their fellow members. They can either choose to give reputation, or take it away in the form of points. In addition, members can choose to partake in SO activities using a variety of identities, and developing a

reputation via open communication and sharing their real world job identities and connections. This brings us to an interesting consideration: Is there a connection between different types of reputations and how users choose to archive?

2. Stack Overflow

SO is a crowd-sourcing platform regarding software development and programming. The website encourages users to contribute both questions and answers in order to build an accurate corpus. SO uses an incentive and reputation system to encourage users to not only contribute, but also to manage and maintain the quality of information on the platform. This is done through giving privileges to users based on reputation, as well as encouraging certain users to become official and de facto moderators.

The reputation system in SO is quite straightforward. Most points are earned via having material voted up by the rest of the community. There are some caveats to gaining points, however. Users can only gain a maximum of 200 points per day via answer upvotes. This means that a particularly hot or popular answer or question may receive more in gross reputation than in net reputation. Therefore, as Anderson, et al. determined, top users target receiving extra points by having their answer chosen as the best answer. There are some other ways of earning points as shown in Table 1.

<i>Action</i>	<i>Points Earned/Lost</i>
Answer is Voted Up	+10 (Max 200 Points per day)
Question is voted up	+5
Answer is accepted	+15
Question maker accepts answer	+2
Question is voted down	-2
Answer is Voted down	-2
Voter votes answer down	-1
Edit of another post is accepted by peer review	+2 (Maximum of 1000 points gained through editing)

Table 1: How to Earn Points on Stack Overflow

Reputation in SO is not without tangible meaning. One of the stated goals on SO, is that reputation is a signal to the community of: How much one is trusted. the quality of one's communication

skills, and the quality and of relevancy of one's questions and answers. In order to give strength to this signal, SO implements a privilege scheme in which users gain or lose privileges based upon their points. Some privileges are essential for full enfranchisement. For example, users must gain at least 125 reputation points in order to be able to vote a question or answer down. Until one has gained a strong enough signal, one is essentially unable to tangibly express a negative opinion on material in the corpus.

It takes a until 20,000 reputation points in order to earn all of the privileges. An appropriate consideration is how many users fall into the purview of the reputation scheme. The system has close to a million users, but only a few dozen have more than 100,000 points. The vast majority of users have between 1 and 200 points.

An important note is the privileges that are given as shown in Table 2:

15 Points - Vote up
15 Points - Flag for moderator attention
50 Points - Leave comments [†]
100 Points - Edit community wiki posts
125 Points - Vote down (costs 1 rep on answers)
200 Points - Reduced advertising
250 Points - Vote to close, reopen, or migrate your questions
500 Points - Retag questions
1000 Points - Show total up and down vote counts
1500 Points - Create new tags
2000 Points - Edit other people's posts, vote to approve or reject suggested edits
3000 Points - Vote to close, reopen, or migrate any questions
5000 Points - Vote to approve or reject suggested tag wiki edits
10000 Points - Vote to delete closed questions, access to moderation tools
15000 Points - Protect questions to prevent answers by new users
20000 Points - Vote to delete negatively voted answers and

Table 2. Points Needed for Privileges

Once a user has 2,000 points, they have the ability to edit other user's posts, in order to improve them. Once they have 20,000 points, they can vote to delete negatively voted answers and questions. The reputation scheme directly gives members control over the information in the system.

2.1 Voting Mechanism of Stack Overflow

The power of affirmative voting is greater than that of negative voting. For each affirmative vote, the question asker can earn +5 points, while a downvote is only worth -2 points. The more extreme case is within regards to voting on answers. The effect here is +10 points for affirmative voting, and -2 for downvotes. The design intent seems to be clear: SO wants to encourage the production of information, by making affirmative voting more powerful than negative voting. Affirmative voting holds such an advantage over negative voting, that an answer could have a gross negative vote count, but still receive a net gain in points.

Voters are the first line of archiving. SO uses an algorithm that incorporates the amount of votes a particular question is receiving in order to promote certain questions over others. In this way, there is a direct effect between voting and the reception of page views. We also chose to look at voters, because it is something that we can confirm as a real act of archiving. While many pages might receive views, it is hard to qualify what these views are. They could be random browsing, lost researchers, or even non-human entities. The act of voting, however, is a tangible attempt to either accept or reject content. Indeed, voters overwhelmingly choose to vote Upwards. In our sample, the ratio of Upvotes to Downvotes was close to 100 to 1.

An issue that can be raised that of why users participate in negative voting at all. A user, when faced with reading a question has three options: No Action, Upvote, or Downvote. Given the above, we would assume that voters with more reputation would care less about the risk associated with voting down.

3. Previous work in Q&A and Determining Quality

Previous research on Q&A has largely focused on empirical reviews to understand which questions and answers are have the highest archival value, which questions are answered, and which questions have transcendent value to third-party evaluators.

Harper, et al. looked at the relationship between the quality of answers and the relationship with price. In this work, the authors conducted a field experiment in order to judge the quality of the answers on Google Answers and Yahoo! Answers. Using undergraduate students to rate the content of answers, the results were that the more money a question was worth, the better, and that Google Answers had better content overall compared to Y! Yahoo! Answers.

Chen conducted a similar field experiment, with the exception of using graduate students and discounting unanswered questions. With these two changes, the authors found that there was no real relationship with price and quality. Instead, there was a clear relationship between the quality of the answer and the reputation of the answerer. The better the track record of the answerer, the better the answer. Jeon, et al., supported Chen's findings when analyzing both Harper and Chen's data sets. Higher prices mean that a question is more likely to receive an answer, but for questions that are answered price does not guarantee better answers. The reputation of the answerer seems to be a better signal for answer quality. Hsieh, et al., confirmed Chen and Jeon's observation in a field study looking at Mahalo's paid question asking service. Interestingly enough, however, they do argue that while the questions with a higher reward may not receive better answers, the questions themselves may be of higher archival value. The results are not conclusive.

Anderson, et al. look at the properties that can predict whether a question can be sufficiently answered in a group knowledge sharing and creation environment. Looking specifically at SO, Anderson finds that there is a reputation pyramid that is largely supported by users answering as many questions as they can, as fast as they can. Anderson uses page views to identify archival quality. The reasoning behind this method is that page views is a transcendent property. The community and the world at large

determine the the value of the information. It is, however, narrow in scope in that it ignores voting altogether.

Zhuolun, et al. find that gamification in the form of badges in SO encourage particular types of behavior. Badges are specific medals that users can earn by completing certain tasks. For instance, a user can earn a “citizen” badge by voting a total of 300 times. In interest to this paper, badges can be used to encourage the amount of voting and the types of voting a member does. The expected results due to the power of voting might be affected by the draw of earning badges.

4. Methodology

SO, like many open Q&A sites, is made up of a diverse collection of users and stakeholders which benefit from and effect the system in different ways. For instance, passive viewers of Q&A sessions may derive significant personal or professional benefit from the exchanges, all while not directly contributing to the system with questions, answers, comments, or votes. We do consider them users, as they have a direct effect on how the site functions. As we previously indicated, they do influence the system by weighing on the algorithm for determining hot questions. In the same way, there are literally tens of thousands of users who have contributed material a few times, but have mostly remained inactive for various reasons. Their collective effect on the system may be significant, and they may derive much value from the site, but they are also close to passive viewers in their individual effect on the system.

The users that are more interesting for this study are those who have invested in the SO reputation system, either through earning points or casting many votes. These users give us a solid foundation for understanding strategy and behavior in a competitive social archiving situation.

In order to better understand the behavior of active voters, a random sample 714 ($p=.05$) was taken of the 24,000 voters who have voted at least 200 times. This threshold was determined as to try to identify active voters. While all votes have an effect, and not voting might be an identity, we wanted to look at specific patterns that arise from those who are frequent voters.

Data was collected from these users along points 11: Reputation, Months as a Member, Profile Picture, Website, Total answers, Highest vote for answer, Total questions, highest vote for total questions, up votes, down votes, votes on answers, and votes on questions.

4.1 Reputation Class

SO has a number things that divides its user base, none such as strong as the amount of points a member has. SO itself divides its community into three groups:

- **Users** - Those with 0 - 999 Points
- **Established Users** - 1000 - 19999 Points
- **Trusted Users** - 20,000+ Points

We used this classification system to perform the analysis on voting behavior. In the sample of those with more than 200 votes, we have 175 Users, 412 Established Users, and 130 Trusted Users.

4. 2 Identify Classification

SO users can choose the amount of information they reveal about themselves, with no special privileges given or taken away due to disclosure. There are three classifications that we determined fit most (but not all) users:

- **Full ID:** Users have a Profile Picture and a link to a Personal Website. All websites were checked at time of data collection.
- **Website Only:** Users have a link to a Personal Website. However, that is the only like to their out of system existence.
- **No ID:** These users have no link to the outside world, or any identification in the system. For all intents, these users are essentially anonymous.

There is also the existence of users who have a profile picture, but no website or direct link to personal work. However, these users are quite rare, and were excluded from the samples as they overlap classification.

An interesting problem is what to do with Profile Names. While names can often tell a lot about a user, there are problems with assumption. While many users use their real names, and others user obvious avatars, others use names that seem real, but in fact, are not. After a brief sample test, we determined that we could not reasonably guarantee the veracity of classification.

Table 1, points out ($P=0.0004$, $F= 6.184$) that Full ID users and Website Only Users are more likely to have higher reputation that their cohorts with NO ID. As Table 2 shows:

	Subjects	Reputation Mean	Std Dev
Full ID	195	10,647	2,100
Website Only	229	8,374	1,140
No ID	252	4,298	9,130

Table 2 ID Type and Reputation Points
 $F=6.184$ $P=0.0004$

5. The Results of Negative Voting

While positive voting has no differentiation along reputation lines, we can find strong correlation between monthly negative votes and reputation. As Table 3 shows:

	Subjects	Mean	Std Dev
Users	175	0.74509	2.8859
Established	412	1.8557	5.2003
Trusted	130	6.108	13.7574

Table 3. Reputation Points and Negative Votes Per Month
 $F=23.047$ $P=0.0001$

We might expect that Full ID and Website Only Users would also have a higher level of monthly negative votes. As Table 3 shows:

	Subjects	Reputation Mean	Std Dev
Full ID	195	2.8489	11.7
Website Only	229	2.2357	4.79
No ID	252	2.3015	5.49

Table 4. ID Type and Negative Votes Per Month

F=20.52 P=0.7594

However, there is no significant difference between the ID types. Indeed, the differentiation between users seems to rely solely on the amount of reputation they have accrued. No ID users may be more likely to have lower reputation scores compared to their counterparts, but those with higher reputation pull their weigh accordingly.

6. Conclusion and Future Work

The results are somewhat expected, and yet confusing at the same time. Users who have more powerful tools at their disposal, like the ability to edit and close Q&A session, are still much more likely to vote negatively. Part of this is likely due to the fact that they are more likely to have a lesser fear of voting negatively. Another factor is probably the tie that users with a lot of reputation points have with the point system. Negative votes, while weak in point damage, do have a tangible effect within the system that the community works. Perhaps those with the most points feel the strongest about this effect.

The fact that users, regardless of ID type, vote negatively reflects positively on the archival system of SO. While many elements of Q&A sessions receive no votes either way, users are engaged to Downvote and attempt to remove material that they determine to be harmful. In addition, by SO's own determination, of whom they consider "Trusted Users," it is interesting to see that as users approach or pass this status, they take part in this form of archiving.

In future work, we plan to look at the effects of earning badges and voting behavior. As we understand from Zhuolun, et al. that badges can have tangible effects on behavior within the system, it is interesting to consider badges that are relevant to voting. For example, SO has badges that directly relate to voting. For instance, the badge "Civic Duty" is awarded to users who have contributed 300 or more votes.

Acknowledgements

This research was partially supported by a Grant-in-Aid for Scientific Research (S) (24220002, 2012-2016) from Japan Society for the Promotion of Science (JSPS).

References

Adamic, L.A., Zhang, J., Bakshy, E., and Ackerman, M.S. Knowledge Sharing and Yahoo Answers: Everyone Knows Something. WWW (2008)

Anderson, A., Huttenlocher, D., Kleinberg, J., and Leskovec, j. Discovering Value from Community Activity on Focused Question Answering Sites: A Case Study of Stack Overflow. KDD (2012)

Chen, Y., Ho, T., Kim, Y. Knowledge Market Design: A Field Experiment at Google Answers. *Journal of Public Economic Theory*, 12 (4) (2010)

Harper, F.M, Raban, D., Rafaeli, S, and Konstan, J.A. Predictors of Answer Quality in Online Q&A Sites. CHI (2008)

Hsieh, G., Kraut, R.E., and Hudson, S.E. Why Pay?: Exploring How Financial Incentives are Used for Question & Answer. CHI (2010)

Jeon, G.YJ., Kim, Y., and Chen, Yan. Re-examining Price as a Predictor of Answer Quality in an Online Q&A Site. CHI (2010)

Treude, C., Barzilay, O., Storey, M. How Do Programmers Ask and Answer Questions on the Web? ICSE (2011)

Zhuolun, L. Huang, Ke-Wei, Cavusoglu, H. Can We Gamify Voluntary Contributions to Online Q&A Communities? Quantifying the Impact of Badges on User Engagement. ICIS (2012)