

Modelling User Collaboration in Social Networks Using *Edits and Comments*

Ifeoma Adaji
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
lta811@mail.usask.ca

Julita Vassileva
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
jiv@cs.usask.ca

ABSTRACT

Research has shown that in Q&A social networks, collaboration between respondents results in quality answers. Since good answers are required to keep any Q&A social network active, it is important to understand the characteristics of these collaborations and the collaborators. In this paper, we investigate how Stack Overflow promotes collaboration by allowing users to edit existing questions and answers in order to improve them. Using over 40,000 answer posts, our study reveals that collaboration in answer posts is not a function of achievement earned in terms of badges, as most edits associated with “best answer” rewards were posted by users who have not earned any answer badge. Our study further shows that posts that earned the “best answer” reward have more comments than those that did not. This study though, work in progress, can aid developers in implementing collaboration strategies in social networks that work.

Keywords

User modelling, collaboration, social networks

1. INTRODUCTION

The use of collaboration has been shown to result in a greater value creation [1]. In Q&A social networks, collaboration between authors results in better answers [2] and quality answers result in an active Q&A network [3]. Hence, the importance of studying collaboration among users who answer questions cannot be overemphasized.

In this study, we examined the characteristics of collaboration and collaborators in answers posted on Stack Overflow (SO), a successful Q&A social network. In SO, users post questions and other users attempt to answer them. The asker of a question selects the best answer provided to his or her question irrespective of how highly an answer is rated by the community. Users can collaborate to improve existing questions and answers by editing them independently of who posted the questions or answers. Only those who have earned the privilege can approve such edits. Users also collaborate by commenting on existing questions and answers. Unlike question and answer posts, comments cannot be edited by other users. However, they can be voted as being useful by other users.

To determine the characteristics of collaboration and collaborators, we analyzed over 40,000 answer posts and their corresponding edits and comments. Our study showed that the collaborators are not the highest achievers in terms of badges earned as most edits associated with “best answer” reward were posted by users who have not

earned any answer badge. Hence, collaboration in answer posts is not a function of achievement earned in terms of badges.

To investigate collaboration using comments, we analyzed the comments of over 20,000 answer posts. The result of our analysis showed that posts that earned the “best answer” reward have more comments than those that did not. We also discovered that the answer posts with the highest number of comments were posted by users that have earned between 1 and 5 answer badges.

This paper is still work in progress. However, the preliminary results discussed in this paper can aid developers and stake holders on strategies to adopt when incorporating collaboration in Q&A social networks.

2. RELATED WORK

2.1 Stack Overflow

Stack Overflow is a Q&A platform where users can ask and answer specific IT related questions. Through active participation and providing high quality answers, the users can gain incentives such as reputation score, badges and privileges. While all users can upvote or downvote other users’ questions and answers, only the user that asked a question that can select the best answer to his/her question. Users have to earn the privilege to edit answers posted by others in a bid to improve them. Users can also comment on answers to improve them or to seek clarification. Similar to questions, comments can also be upvoted or downvoted. Upvotes and downvotes contribute to the reputation score of users.

In Stack Overflow, badges are earned by users who are especially helpful in the community¹. Badges are awarded in several categories including question, answer, participation, tag and moderation badges. We only considered badges in the answer category because we are interested in collaboration in answer posts since quality answers is one of the factors that keeps a Q&A social network active [3].

2.2 Collaboration in social networks

Dalle et al [4] applied game theory to the study of collaboration in social networks. Their study was based on a class of local contribution games where only neighbors of a contributor benefit from the positive effects of a contribution. This form of network is totally different from SO where respondents who post helpful answers benefit from their contribution in the form of upvotes and higher reputation scores.

In [5], McDonald focused on the use of social networks in an organization to recommend people for possible collaboration. Unlike the case of McDonald, in this study, we examined existing

¹ <http://stackoverflow.com/help/badges>

collaborations through *edits* and *comments* in order to identify the characteristics that make them successful.

In their study of social networks, Dalle et al [2] studied *edits* as a means of collaboration. They concluded that collaboration between authors results in better answers. Though our research is an extension of this study, unlike Dalle et al, our focus is only on the answer posts.

Pancieria et al [6] studied collaboration in an open source project, Cyclopath. By conducting a user survey, they studied the reasons behind collaboration from the perspective of the editors and the consumers. Though we also studied the characteristics of collaborators, unlike Panciera et al, we did that by analyzing the existing data of users and not by user study. The context of their study is also different from ours; we studied a Q&A social network where incentives are earned by providing high quality questions and answers, while Panciera studied Cyclopath, a bike route finder.

In their study of interactions in Stack Overflow, Wang et al [7] developed a topic modelling strategy that can assign a new question to existing categories of questions. Their study of collaboration did not include the use of edits and comments.

To the best of our knowledge, no research has been carried out on understanding the characteristics of collaborators in answer posts in Stack Overflow using edits and comments.

3. RESEARCH METHOD & RESULTS

To study the characteristics of collaboration between respondents in SO, we collected detailed information of answers posted in response to questions that scored above the average score. We calculated the average score of questions to be 203 points², hence we considered answers posted to questions that earned at least 203 points. We identified 4,521 questions that met this criteria. These questions received 49,738 answers which were posted by 32,538 users. The summary of the dataset used is described in table 1.

Table 1. Summary of dataset used in this study

Number of questions	4,521
Number of answers	49,247
Number of users who posted answers	32,538
Total number of edits done on these answers	14,343
Number of users who carried out these edits	12,232

3.1 Collaboration using *edits*

To understand the use of *edits* in Stack Overflow as a method of collaboration, we identified the edits that were carried out on the 49,738 answers in the dataset. There were 14,343 edits of which 7,909 were approved. Figure 1 shows the spread of the data based on the number of edits carried out. Most posts were edited less than five times. These edits were approved because they improved the initial answer post³.

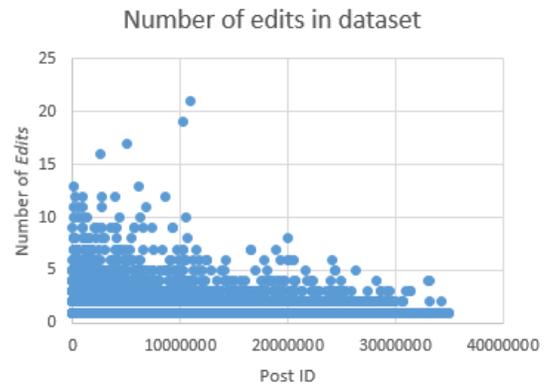


Figure 1. Number of edits in the dataset

In order to understand the characteristics of users who collaborated using these edits, we aimed to answer the following research question; are commenters high achievers in the network? In other words, is collaboration in answer posts a function of achievement in terms of badges earned? To answer these questions, we categorized users into 4 based on the number of answer badges the users have earned. In Stack Overflow, badges are earned by users who are especially helpful⁴. We only included users who have earned answer badges because this study is based on collaboration in answer posts. Table 2 describes the number of badges in each category and their description. The use of categories is important in order to have a narrower view of who collaborators are in the network.

Table 2. Categories of badges

Category	Description
EQ1	Users who have earned only 1 answer badge
GT1	Users who have earned between 2 and 5 answer badges inclusive
GT5	Users who have earned between 6 and 10 answer badges inclusive
GT10	Users who have earned more than 10 answer badges
NON	Users who have not earned any answer badges

Figure 2 shows the breakdown of edits carried out by users in the various categories. The users that edited most of the answer posts were users who have not earned any answer badge in the past. On the other hand, the users who have earned the most answer badges carried out the least number of edits. Hence collaboration using edits was done by non-achievers (users who have not earned any answer badge) in the network.

² In calculating average, we only included question posts that had scored at least one point.

³ <http://stackoverflow.com/help/privileges/edit>

⁴ <http://stackoverflow.com/help/badges>

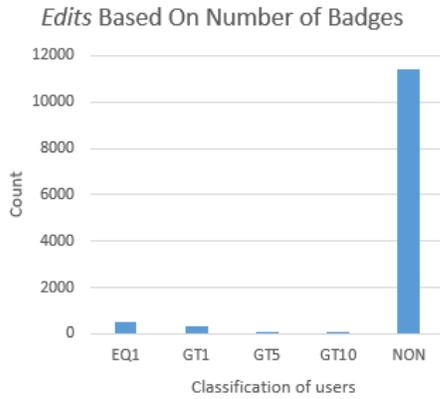


Figure 2. Number of edits in each user category

Since edits have to be approved before the suggested changes are committed, we further analyzed the approved edits. Of the 14,343 answer edits, only 7909 were approved by the community. Of these, 7,353 were posted by users who have not earned any answer badge. Hence, the users who collaborated the most were not high achievers in terms of badges earned. Only 556 edits were posted by users who have earned at least one answer badge. The breakdown of these 556 users based on the category of badges is shown in figure 3.

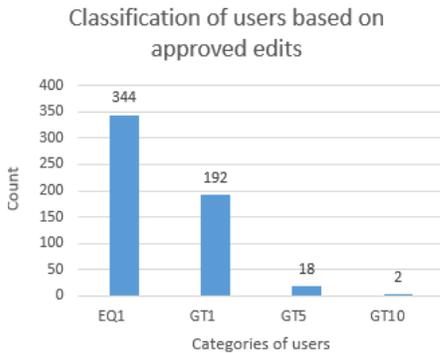


Figure 3. Classification based on approved edits

Of the approved edits posted by answer badge earners, over 60% of them were posted by users who have at least one answer badge. On the other hand, users with the most answer badges only have a few edits approved.

3.2 Collaboration using comments

The use of comments is another form of collaboration in Stack Overflow. Users can improve on existing questions and answers by posting comments related to such questions and answers. These comments could be upvoted based on how useful they are to other users.

To explore the use of comments for collaboration, we aimed to answer the following research question; does collaboration using comments lead to better answers? In other words, do posts that earn the *best answer* reward receive more comments?

To do this, we analyzed the comments of the answers identified in table 1. We considered the answers in response to questions that scored higher than the average score of 203. For this study, we only considered comments that had a vote of more than 1, hence we excluded questions that had 1 or no comment associated with it. We also excluded questions that did not have a corresponding “best answer” post. The final data set was reduced to 54,212 comments

made on 20,776 answer posts in response to 4,443 questions. These were posted by 26,728 users. On the average, an answer post had 5 comments. Figure 4 describes the distribution of comments among the answer posts.

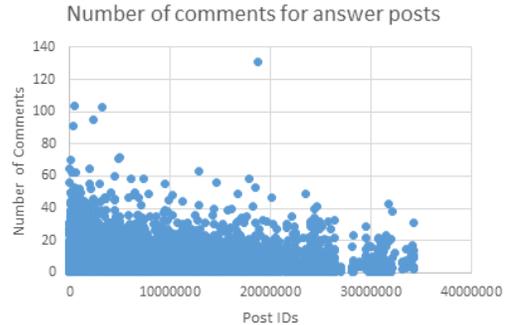


Figure 4. Number of comments for answer posts

For each question post, we identified the corresponding answer post with the highest number of comments, *MaxComment*. We also determined the answer posts with the “best answer” reward, *BestAnswer*, for each of the questions. We then compared *MaxComment* to *BestAnswer* to determine if the posts with the maximum number of comments also earned the best answer reward. 3,374 posts had *MaxComment* and also earned *BestAnswer* reward. We concluded that 75% of posts with the highest number of comments also earned the best answer reward, hence the use of comments enhances collaboration as it leads to improved answers.

To determine the level of achievement of the users who collaborate by commenting, we studied the 3,374 posts that had the highest number of comments, *MaxComment*. Figure 5 shows the participation of users based on the category of badges in table 1.

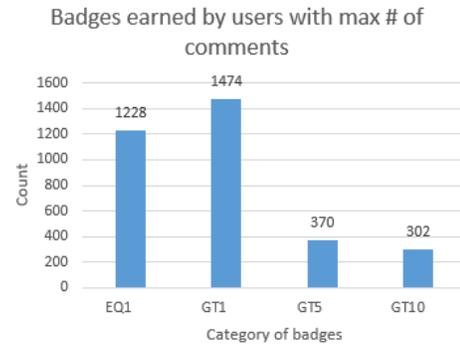


Figure 5. Badges earned based on number of comments

The answer posts with the highest number of comments were posted by users that have earned between 1 and 5 answer badges. Answer posts of users who have earned over 10 badges had fewer comments. This result could mean that the high achievers do not engage in collaboration as much as the users with fewer comments.

3.3 Discussion

This paper aims at studying collaboration in a typical Q&A social network using *edits* and *comments*. From the results of our analysis, we can conclude the following:

Since collaboration results in better answers [2], there is a need to encourage collaboration Q&A in social networks. In Stack

Overflow new persuasive strategies to encourage active collaboration through edits and comments may be required. For instance, badges or rewards that encourage editing and commenting could increase collaboration through these means.

Using comments as a form of collaboration to improve answers posted on the network, we concluded that the best answer posts usually had more comments compared to other posts. Hence networks should encourage users to participate actively in commenting as this could lead to better quality answers.

4. CONCLUSION & FUTURE WORK

In this short paper, we examined the use of *edits* and *comments* as forms of collaboration in a successful Q&A social network, Stack Overflow. Since high-quality answers to users' questions is one of the sustaining factors of a Q&A social network, it is important to understand who the collaborators are that improve existing answers.

We analyzed over 40,000 answer posts and their corresponding edits and comments. Our study showed that the collaborators are not the highest achievers in terms of badges earned, since users with more than 10 answer badges carried out only a few edits to answers that earned the "best answer" reward. This study also revealed that most edits associated with answers are posted by users who do not have answer badges. Of the users who have earned answer badges and have edited "best answer" posts, only a few of them have earned over 10 badges while about 60% of them have just one badge. Hence, collaboration in answer posts is not a function of achievement earned in terms of badges.

To investigate collaboration using *comments*, we analyzed the comments of over 20,000 answer posts. The result of our analysis showed that posts that earned best answer reward had more comments than those that did not. We also discovered that the answer posts with the highest number of comments were posted by users that have earned between 1 and 5 answer badges.

This study is still work in progress, hence further analysis has to be carried out to validate our findings. For example, there is a need to analyze the comments in order to exclude comments that do not contribute to the answers posted.

Although the study showed that the high achievers in the network who have over 10 answer badges posted answers with fewer comments, this could mean that their answers could not be improved any further based on their level of expertise. Hence we need to extend this study by analyzing the content of the posts and comments.

5. References

- [1] L. Camarinha-Matos and H. Afsarmanesh, "Collaborative Networks. Value creation in a knowledge society," *International Federation for Information Processing (IFIP)*, vol. 207, pp. 26-40, 2006.
- [2] J.-M. Dalle, M. Devillers and M. den Besten, "Answer editing in StackOverflow," *Proceedings of 2nd Thematic Conference on Knowledge Commons: Governing Pooled Knowledge Resources*, 2014.
- [3] G. Dror, D. Pelleg, O. Rokhlenko and I. Szpektor, "Churn prediction in new users of yahoo! answers,," *Proceedings of the 21st International Conference Companion on World Wide Web*, pp. 829-834, 2014.
- [4] L. Dall Asta, M. marsili and P. Pin, "Collaboration in social networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 12, pp. 4395-4400, 2012.
- [5] D. McDonald, "Recommending Collaboration with Social Networks: A Comparative Evaluation," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. Pages 593-600, 2003.
- [6] K. Panciera, M. Masli, L. Terven. "How should I go from ___ to ___ without getting killed?": motivation and benefits in open collaboration," *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11)*. ACM, pp. 183-192, 2011.
- [7] S. Wang, D. Lo, L. Jiang. "An empirical study on developer interactions in StackOverflow," *Proceedings of the 28th Annual ACM Symposium on Applied Computing (SAC '13)*. ACM, pp. 1019-1024, 2013.