2014 28th International Conference on Advanced Information Networking and Applications Workshops

Reputation on Stack Exchange: Tag, You're It!

Laura MacLeod
Department of Computer Science
University of Victoria
Victoria, BC, Canada
Email:lmacleod@uvic.ca

***Abstract*—Question and Answer sites like the Stack Exchange network allow users to contribute knowledge in a variety of topics. Of particular interest are the habits of expert users, or users with high reputation scores. Expanding on previous work done by Bosu et al., we performed an exploratory analysis of the data to determine if there is a positive correlation between user reputation scores and the diversity of tags a user contributed to. For our analysis, we used data available from Stack Exchange to created a visual network of user tagging habits. We found that there is a correlation between these two measures, suggesting that expert users contribute to a wide variety of tags. Our research also shows poor community structures in the networks studied. This is consistent with the prevailing literature on Stack Exchange use and the power law. Our findings confirms one of the recommendations put forward by Bosu et al. through the use of a much larger population and put forward ideas for new areas of exploration on question and answer sites.**

***Keywords*-Stack Overflow, Community Detection, Tagging, Software Engineering ;**

I. Introduction

Question and answer (Q&A) sites like those found in the Stack Exchange network let users ask open-ended questions to a large population. In return, users receive answers from users with knowledge or experience in the respective field. Stack Exchange sites also use voting to rank answers for a question. By providing high-ranked answers, asking questions, and participating in discussions, users gain reputation points which represent their activity on the site. These in turn unlock new features on the site for them, such as being able to downvote answers.

Established in 2008, the Stack Exchange network is comprised of over 150 Q&A sites. They cover topics such as travel, software development, and English grammar. Stack Overflow, which is focused on programming questions, is by far the most popular site on the Stack Exchange network: users have posted over 6.1 million questions and over 12 million comments [1].

In 2009, the Stack Exchange network made its data freely available, including user names, locations, and artifact IDs. There has been some recent interest in mining this data in order to understand the habits of developers. For example, the 10th Working Conference on Mining Software Repositories specifically held a data mining competition on the Stack Overflow data dump [2].

II. Related Work

Bosu et al. [3] propose five recommendations as to how users can improve their reputation score on Stack Overflow. The authors conduct a case study of 10 "trusted users" with high reputation scores to discover what common habits these highly-ranked users share. Our work examines whether one of these recommendations holds true — namely that a user should comment on as many different types of tags as possible to improve their score. Contrary to Bosu et al., we examine data about almost 150,000 users of the StackExchange Network.

A. Experts on Q&A Sites

A number of studies of Q&A sites have focused on expert users. This typically refers to users who are highly active on the site or have demonstrated a certain level of expertise. Pal et al. [4] define the expert as an "answer person" who helps drive the community to evolve and change. Hanrahan et al. [5] take a more methodical approach and define the expert using three indicators: the reputation score of the user, a Z-score obtained using an algorithm by Zhang et al. [6], and finally a delta between up and down votes. Even then, these authors further divide experts into categories such as users who are expert question askers or those who specialize in answering.

In our work, we use the reputation score on Stack Exchange sites as a metric to measure a user's activity and the quality of that activity. Users earn reputation points by posting and answering questions as well as completing tasks such as filling out the biography of their profile. A user may have many profiles across a number of Stack Exchange sites, all with different reputation scores.

Pal et al. [7] created a model for predicting which users will become experts and identified three working patterns of experts. For Stack Overflow, Wang et al. [8] showed that only 8% of users answer more than 5 questions. This suggests that those who eventually become experts make up a small minority of a Q&A community. This finding is consistent with other studies that show that on Stack Overflow, participation follows the power law: a small number of users generate a large amount of the content [1].

978-1-4799-2652-7/14 $31.00 © 2014 IEEE
DOI 10.1109/WAINA.2014.108

IEEE computer society

III. Data and Approach

This section outlines the steps taken to collect and analyze our data. In an initial exploratory phase, we collected data from the Stack Exchange network and created a weighted directed graph of users from this data. From this, we were able to apply an exploratory data analysis technique and develop two research questions which we then statistically tested.

A. Data Selection

To explore the relationship between reputation scores and tagging on Stack Exchange sites, we collected data from a random sample taken from a Stack Exchange data dump. We used the Stack Exchange data dump from June 2013. This dump represents all activity on the network from early 2008 up until June of 2013. The dump exceeded 13 Terabytes of data.

Name	Total Comments	Total Posts	Total Uses
android.stackexchange.com	43080	33888	27761
apple.stackexchange.com	87343	75644	38786
cstheory.stackexchange.com	31723	12527	11999
meta.android.stackexchange.com	2141	1382	2101
meta.diy.stackexchange.com	889	774	1058
meta.gis.stackexchange.com	1805	1077	2410
meta.superuser.com	12215	5033	16393
meta.travel.stackexchange.com	2142	1298	734
meta.unix.stackexchange.com	1691	1201	3180
meta.webmasters.stackexchange.com	1121	933	1872
stackapps.com	7350	3282	13407
webapps.stackexchange.com	22827	27276	30124

Table I
The sites randomly selected for our sample.

Because of the sheer size of the data and the probing nature of our investigation, it was not feasible to unzip and create databases for all of the Stack Exchange sites. Instead, we used the size of the zipped files as a proxy for our population. From this population, we calculated our sample size and randomly selected Stack Exchange sites [9].

The resulting sample was made up of seven meta sites and five traditional Q&A sites (cf. Table I). Meta sites are those dedicated to questions about other Stack Exchange sites. For example, a question about how reputation scores are calculated on Stack Overflow should be posted on the Meta Stack Overflow site.

B. Exploratory Data Analysis

To generate our research questions, we used an exploratory data analysis technique. Exploratory data analysis relies on

> *"visual displays to reveal vital information about the data,... the techniques [of exploratory data analysis] is one of searching, with stress placed up the use of alternative techniques to assess the same body of data." [10]*

In line with this philosophy we first explored the data before formulating our research questions. We first created a visual representation of the data we had collected. By exploring the data, we observed the relationships between our variables which we will compare to what we expected to find.

C. The Graph

After extracting data from databases made from the Stack Exchange data dumps, the data was converted into the Pajek [11] file format. Pajek uses a text based representation to describe graphs and can be used by a number of visualization programs. We used the freely available Gephi [12] tool to construct our graphs.

The resulting network was one where nodes represented users and tags. We created weighted, directed edges from users to tags based the number of times a user had posted a question, answer, or comment in a respective tag. We created such a graph for each Stack Exchange community in our sample.

Figure 1 shows an example for such a graph. User nodes are shaded based on their PageRank [13] in this network, as a simple to distingush the edges going from or two nodes. Nodes with a high number of outgoing edges, or tags commented on, represent users with a higher amount of activity on the site.

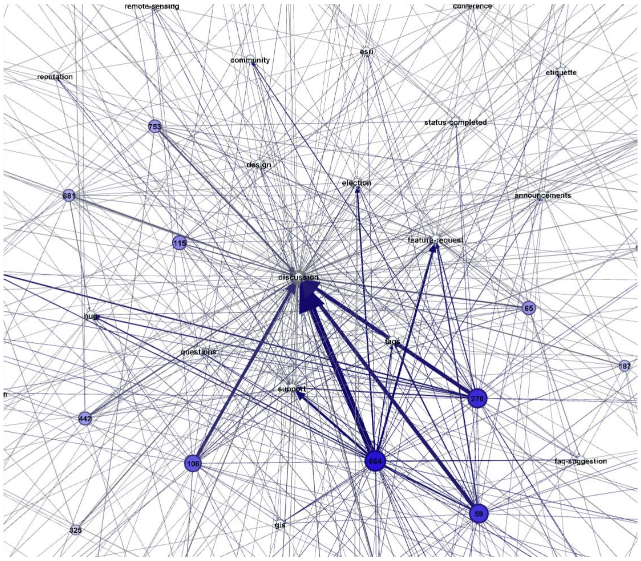

Figure 1. An example graph created using data from Stack Exchange

IV. Research Questions and Findings

Based on our exploration of the data, we formulated the following research questions.

RQ 1: Is there a correlation between a user's reputation score and the number of unique tags that they have commented on?

In our data analysis phase, we selected a small subset of user IDs who had a high Page Rank score. By manually exploring these users on the relevant Stack Exchange site, we observed that users with a high PageRank tended to have high reputation scores. Based on this observation, we wanted to statistically explore the link between reputation scores and tagging habits.

To explore this research question, we calculated the Pearson's correlation coefficient for each Stack Exchange site in our sample [14]. This does not take into account the weight of edges — i.e., the number of times a user has commented on an individual tag. The results from these calculations can be seen in Table II.

Site	Correlation
android.stackexchange.com	0.82
apple.stackexchange.com	0.83
cstheory.stackexchange.com	0.79
meta.android.stackexchange.com	0.90
meta.diy.stackexchange.com	0.63
meta.gis.stackexchange.com	0.70
meta.superuser.com	0.43
meta.travel.stackexchange.com	0.80
meta.unix.stackexchange.com	0.56
meta.webmasters.stackexchange.com	0.54
stackapps.com	0.77
webapps.stackexchange.com	0.84

Table II
RESULTS FOR REPUTATION AND TAGGING CORRELATION.

RQ2: What is the modularity between users' reputation scores and tagging?

Initially in our study we wanted to identify experts in small sub-communities based on tags. In manually observing the data, it appeared that experts were not confided to sub-communities, nor where there distinct sub communities to be observed. This lead to the development of our second research question.

To answer this research question, we used the community detection algorithm by Blondel et al. [15]. The algorithm breaks connections between nodes into partitions and configures them to get to the best point where connections within a group are dense, but sparse leading out of the group. This allowed us to calculate the modularity score of the respective networks. A modularity score of 0.3 to 0.7 denotes good modularity within a network [16].

V. DICSUSSION

The correlation results from our first research question show that there is a strong positive correlation between having a high reputation score and commenting on a diverse number of tags. A Pearson's correlation score of 0.4 typically denotes a strong positive correlation, and a score of 0.7 or above denotes a very strong relationship. Our findings provide support for the claims made by Bosu et al. [3].

Site	Modularity
android.stackexchange.com	0.22
apple.stackexchange.com	0.20
cstheory.stackexchange.com	0.23
meta.android.stackexchange.com	0.22
meta.diy.stackexchange.com	0.23
meta.gis.stackexchange.com	0.19
meta.superuser.com	0.17
meta.travel.stackexchange.com	0.16
meta.unix.stackexchange.com	0.28
meta.webmasters.stackexchange.com	0.24
stackapps.com	0.31
webapps.stackexchange.com	0.36

Table III
RESULTS FOR MODULARITY SCORE CALCULATIONS.

The meta sites have a lower correlation score on average than the other Stack Exchange sites. This could be due to the nature of meta sites. Since they serve more of an administrative function there may not be as much activity as on regular sites or they may use a more limited set of tags. Besides the meta sites, all of the regular sites have a very high minimum correlation value of .77. We suggest using these findings to generate a hypothesis which could be further tested on not only Stack Overflow, but other larger Stack Exchange sites.

The findings of our second research question are consistent with previous work on Stack Overflow sites. We found that the sites in our sample had very poor modularity. Only two of the communities had a score above 0.3. These were both larger sites in our sample and neither were meta sites.

The work by Mamykina et al. shows that only a handful of users actively participate on these sites [1]. Therefore we can expect to have large amounts of individuals who do not belong to a subset of the graph or a community in this case. Because so few people interact on a high level, the modularity of the graph is low. This is encouraging, as it suggests that other Stack Exchange sites—which, compared to the Stack Overflow site, are smaller in activity—still have similar properties to the results found by Mamykina et al. and Bosu et al. More research would be needed to confirm these findings.

VI. THREATS TO VALIDITY

Our data set was limited to a number of the smaller Stack Exchange sites and should not be generalizabled to other sites. In particular, we did not include the most popular site on the network, Stack Overflow, in our random sample. Because Stack Overflow is the largest and most popular site in the Stack Exchange network, it would be especially interesting to see the results of our investigation carried out again on that site.

The construction of our graph relied on two types of data: users and the tags they had been active in. While our findings show a positive correlation between tagging and reputation scores, there could be other factors that cause or influence this correlation. For example, we did not look at the total number of tags in each site and whether this impacted the distribution of reputation scores. We also did not take into account the amount of time a user had been active on any one site.

Using an exploratory data analysis technique means that we relied on observations of our data to develop research questions. Therefore, there may be bias in how our questions were constructed. We would recommended to use a variety of other techniques to confirm our findings.

VII. Future Work

This research was of an exploratory nature and the questions were limited in their scope. Regardless, the results of our study have drawn our attention to a number of hypotheses that we would like to explore in the future.

First, our correlation scores show a positive correlation between tagging and reputation score. In the future, we hope to use this work as basis for a hypothesis that we can then test on other sites in the Stack Exchange Network. More work could be done as well to see if these findings carry over to other Q&A sites with a similar setup, and if so, explore the common elements between sites that contribute to any convergent or divergent findings.

Second, since we have now explored one of the recommendations made by Bosu et al. we would be interested in applying other recommendations made by the authors to a larger population sample. For example Bosu et al. also recommend that users be the first answer to a question in order to improve their reputation score. We may also look into generating our own recommendations for users based on further research.

Third, by showing a positive correlation between reputation scores and tagging, it would now be interesting to investigate whether this relationship is a causation. This would involve a larger experiment and adding qualitative data. That study could aim to find those users who are experts in a smaller community, meaning that they have a high level of knowledge but do not post to a large number of tags. Based on the work by Bosu et al. we know that such users exist, but that they seem to be in the minority. This could also be an indication of a way in which the ranking system of Stack Exchange guides the behaviour of users. For example, it has been shown that some companies use sites like Stack Exchange to assess potential new hires [17]. Based on our findings there may be a way to game the ranking system for personal gain. Both of these ideas would be highly interesting to explore and could have an impact on the ways in which community based question and answer sites are designed.

Finally, users on one Stack Exchange site consistently have profiles on other Stack Exchange sites. Another possible continuation of the work presented in this paper would be to explore the reputation scores and habits of users across these communities.

The data that we have collected in this experiment has alerted us to a number of different questions that could to be explored about the Stack Exchange network of sites and how people use Q&A sites. We intend to explore these questions in our future work.

VIII. Conclusion

We reported on an exploratory data analysis study. Q&A sites are of particular interest to research because they bring together people from diverse backgrounds to share knowledge on an unlimited number of topics. Sites like Stack Exchange have fundamentally shifted how programmers problem solve and distribute knowledge. Building on the previous of work of Bosu et al., we statistically explored the relationship between reputation scores and tagging habits. We found that there is a positive correlation between reputation scores and contributing to a diverse number of tags. Our findings also supported previous work on the power law and contribution on Stack Exchange. We observed poor sub-community structure on the majority of sites analyzed. Finally our work has revealed to use a number of interesting ways in which we can further add to the knowledge of Q&A sites.

References

[1] L. Mamykina, B. Manoim, M. Mittal, G. Hripcsak, and B. Hartmann, "Design lessons from the fastest q&a site in the west," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2011, pp. 2857–2866.

[2] (2013) Mining challenge. [Online]. Available: http://2013.msrconf.org/challenge.php

[3] A. Bosu, C. S. Corley, D. Heaton, D. Chatterji, J. C. Carver, and N. A. Kraft, "Building reputation in stackoverflow: an empirical investigation," in *Proceedings of the Tenth International Workshop on Mining Software Repositories*. IEEE Press, 2013, pp. 89–92.

[4] A. Pal, F. M. Harper, and J. A. Konstan, "Exploring question selection bias to identify experts and potential experts in community question answering," *ACM Transactions on Information Systems (TOIS)*, vol. 30, no. 2, p. 10, 2012.

[5] B. V. Hanrahan, G. Convertino, and L. Nelson, "Modeling problem difficulty and expertise in stackoverflow," in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion*, ser. CSCW '12. New York, NY, USA: ACM, 2012, pp. 91–94.

[6] J. Zhang, M. S. Ackerman, and L. Adamic, "Expertise networks in online communities: structure and algorithms," in *Proceedings of the 16th international conference on World Wide Web*. ACM, 2007, pp. 221–230.

[7] A. Pal, S. Chang, and J. A. Konstan, "Evolution of experts in question answering communities." in *ICWSM*, 2012.

[8] S. Wang, D. Lo, and L. Jiang, "An empirical study on developer interactions in stackoverflow," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*. ACM, 2013, pp. 1019–1024.

[9] Raosoft. (2004) Sample size calculator. [Online]. Available: http://www.raosoft.com/samplesize.html

[10] F. Hartwig and B. Dearing, *Exploratory Data Analysis*, ser. 07. SAGE Publications, 1979. [Online]. Available: http://books.google.ca/books?id=jF8QC-BkhvQC

[11] V. Batagelj and A. Mrvar, "Pajek-program for large network analysis," *Connections*, vol. 21, no. 2, pp. 47–57, 1998.

[12] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: an open source software for exploring and manipulating networks." in *ICWSM*, 2009.

[13] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: bringing order to the web." 1999.

[14] D. LeBlanc, *Statistics: Concepts and Applications for Science*, ser. Statistics: Concepts and Applications for Science. Jones and Bartlett, 2004, no. v. 2. [Online]. Available: http://books.google.ca/books?id=gtawVU0oZFMC

[15] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, 2008.

[16] A. Clauset, M. E. Newman, and C. Moore, "Finding community structure in very large networks," *Physical review E*, vol. 70, no. 6, p. 066111, 2004.

[17] A. Capiluppi, A. Serebrenik, and L. Singer, "Assessing technical candidates on the social web," 2013.