

Knowledge Curation Discussions and Activity Dynamics in a Short Lived Social Q&A Community

Hengyi Fu
School of Information
Florida State University
hf13c@my.fsu.edu

Besiki Stvilia
School of Information
Florida State University
bstvilia@fsu.edu

ABSTRACT

Studying the dynamics and lifecycles of online knowledge curation communities is essential to identify and assemble community type specific repertoires of strategies, rules, and actions of community design, governance, content creation and curation. This paper examines the lifecycle of a short lived social Q&A community on Stack Exchange by performing the content analysis of the logs of member discussions and content curation actions.

Keywords

Online communities; knowledge curation; social Q&A; community lifecycle.

1. INTRODUCTION

Knowledge and data curation online communities have become essential parts of the Web's knowledge and information infrastructure providing free knowledge resources to both human and software agents (e.g., search engines). Although there is a significant body of literature on the design and management of online communities and data curation, there have been only a few studies of the dynamics and lifecycles of online knowledge curation communities in general, and social Q&A communities in particular. This poster presents part of a larger study which examines the work organization and dynamics pertaining to knowledge curation in Stack Exchange social Q&A communities. In particular the poster presents analysis of the work organization and dynamics of a short-lived social Q&A community which lasted only 30 weeks, from its inception to its closing. The study's findings can inform the design and management of similar knowledge creation and curation communities, as well as the definition and provision of infrastructure components, including policy and software tools to support their work and development at different stages of their lifecycles.

Online Social Q&A sites are powerful tools for users to seek knowledge by posing questions and getting answers from peer communities. Prior research on social Q&A communities falls into two main categories. The first category focused on analyzing the content of posts from social Q&A communities to identify and understand information needs and motivations for asking questions [1], evaluate and predict the quality of the questions and answers [2], and identify the experts in social Q&A communities [3]. The second category considered Q&A communities as sociotechnical systems focused on

understanding the social organization of Q&A systems, such as what factors enable the social Q&A communities to thrive, how people interact with each other in these sites, and how the community design (reputation and badge system, user groups, governance mechanisms, etc.) could facilitate user contributions and community growth (e.g., [4, 5]). This paper reports on early results of an exploratory study which extends that later literature by focusing on the dynamics and lifecycle aspects of a social Q&A community.

2. RESEARCH QUESTIONS

The study investigated the following research questions:

1. What are some of the actions used in community building, maintenance, and knowledge curation activities in social Q&A communities?
2. What are some of the strategies used in those activities?
3. What are the temporal distributions of those actions and related discussions throughout the community's lifecycle?

3. METHOD AND DATA

To address the above questions, the authors performed content analysis of the logs of member actions and discussions from one of the closed Q&A sites at Stack Exchange called *Economics* (<http://area51.stackexchange.com/proposals/1618/economics>). The site successfully passed the definition, commitment, and private beta phases defined by Stack Exchange management (<http://area51.stackexchange.com/faq>), but failed to be launched and was closed by Stack Exchange after staying in the public beta phase for 206 days. The log data from the Economics site spans 8 months, from October 2011 when the site entered private beta, to May 2012 when the site was closed. It contains data from both the main site (topic specific question-answering threads) and the meta-site (site management discussion threads). The authors used the data dump of the Economics site published by Stack Exchange and performed content analysis with open coding of 166 entries from the meta site (51 entries posted as questions, 115 posted as answers and comments). Next, the authors iteratively clustered codes into 13 categories and recoded the data within those categories. Finally, the authors generated descriptive time series statistics for 30 weeks for both the 13 meta discussion categories (417 instances) and 38 action type codes (3,284 instances) defined and assigned by Stack Exchange software to 1,007 question-answering threads in the main site. The action frequencies of each user on the Economics main site and the meta site were also calculated.

4. FINDINGS AND DISCUSSION

4.1 Community Building Actions

The analysis of the time series data of action type codes showed that members did a fair amount of quality assurance work by actively revising submitted questions and answers during the first 11 weeks of the site's existence (Fig. 1). In week 12, however, the frequency of that activity experienced a sharp

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

Copyright is held by the owner/author(s).

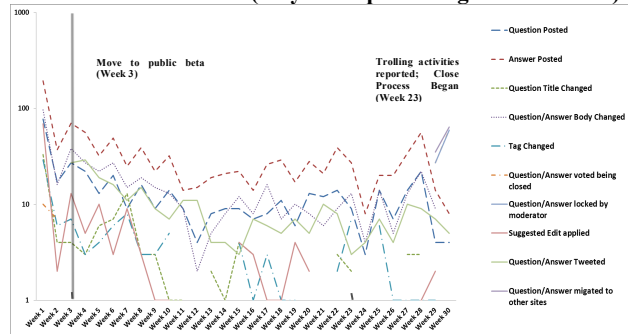
JCDL '16, June 19-23, 2016, Newark, NJ, USA

ACM 978-1-4503-4229-2/16/06.

DOI: <http://dx.doi.org/10.1145/2910896.2925432>

decline, relative to the number of questions asked and answered in the same time period. Another decline in the question and answer revision activity occurred in week 21, right before a trolling attack was discovered and a decision was made by Stack Exchange to close the site. A total of 303 unique users have been identified in the main site, while the unique user number in the meta site was 49. The numbers of user action/contributions in both the main and meta sites followed power law distributions: a small group of community members did a large amount of work and site managing discussions.

Figure 1. Temporal distribution of community building actions in the main site (only the top 10 categories included)



4.2 Community Building Strategies

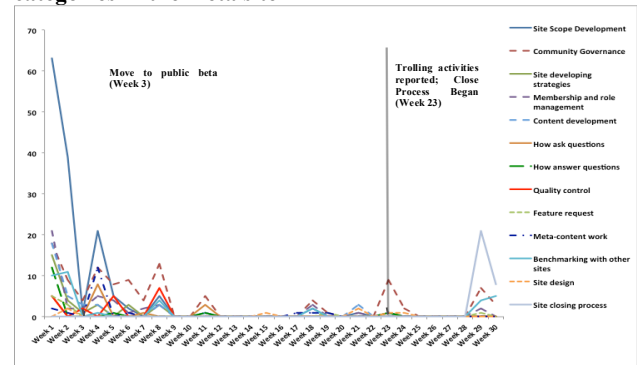
The most frequently occurring discussion categories on the meta site were discussions about the site's scope definition, community governance, and work organization. The community struggled with the following: defining an optimal scope definition for the site; developing strategies to recruit new members and to increase member participation; to generate content; to promote the site; to intervene effectively to ensure the quality of questions and answers; to educate and enculturate new members and users; and to fight off malicious agents (e.g., trolls). The time series data of meta discussions showed ongoing discussions both in the private beta and the public beta phases on whether to restrict the scope of the site to high level, graduate questions (to recruit and retain experts) or allow lay, normative, and homework type questions as well (to increase the traffic to the site; see Fig. 2). Identifying an optimal scope for a community's site at a particular phase of the community's lifecycle is essential to successfully jumpstart and/or grow the community. In the private beta phase the founders of the *Economics* site chose a narrow scope for the site limiting it to advanced, graduate level questions, while more general, basic level questions could be allowed once the site entered public beta. They were concerned that a broader or ambiguous scope would encourage low quality and/or simple questions and discourage experts from joining the community during the community's early building stage.

This strategy was also partially influenced by Stack Exchange's policy which required a site in its private beta phase to have a certain number of highly rated questions as well as a certain number of members who had achieved high reputation levels at Stack Exchange.

The literature shows that while proving access to a large collection of high quality content can help to jumpstart an online peer curation community, setting its scope to a narrowly defined topic or population may hamper its growth. Some of the existing or potential members may not buy into it and/or may not see their place within that narrow scope. Knowledge curation work is a complex set of activities. In addition to experts, who can provide high quality content, online peer knowledge curation

communities need members who are willing to perform routine but important quality assurance activities (e.g., copyediting, content migration, resolving conflicts, and, fighting spammers and trolls) [1,7]. Experts might not be interested in performing those types of work.

Figure 2. Temporal distribution of member discussion categories in the meta site



5. CONCLUSION AND FUTURE WORK

The preliminary findings of the study point to the importance of setting an appropriate scope and continuous quality assurance (e.g., evaluating, deleting, revising questions and answers, fighting trolls) for a Q&A community to succeed. The findings point to difficulties in striking a balance between making the content of a Q&A site interesting and challenging enough to lure in and retain experts and in making the site's scope general enough to accrue a critical mass of participants and traffic. Immediate future work will include expanding this study to a successful Q&A community and comparing and contrasting its repertoires of activity strategies, patterns and dynamics to the findings from the examination of the *Economics* social Q&A community.

6. REFERENCES

- [1] Cunningham, S. J., and Hinze, A. 2014. Social, religious information behavior: An analysis of Yahoo! Answers queries about belief. *Advances in the Study of Information and Religion*, 4, 1, 3.
- [2] Harper, F. M., Raban, D., Rafaeli, S., and Konstan, J. A. 2008. Predictors of answer quality in online Q&A sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 865-874.
- [3] Pal, A., Harper, F. M., and Konstan, J. A. 2012. Exploring question selection bias to identify experts and potential experts in community question answering. *ACM Transactions on Information Systems*, 30, 2, 10.
- [4] Mamykina, L., Manoim, B., Mittal, M., Hripsak, G., and Hartmann, B. 2011. Design lessons from the fastest q&a site in the west. In *Proceedings of the SIGCHI conference on Human factors in computing systems*.
- [5] Tausczik, Y. R., Kittur, A., and Kraut, R. E. 2014. Collaborative problem solving: a study of mathoverflow. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 355-367.
- [6] Resnick, P., Konstan, J., Chen, Y., and Kraut, R. E. 2012. Starting new online communities. In *Building Successful Online Communities: Evidence-based Social Design*, 231-280.
- [7] Stvilia, B., and Les Gasser. 2008. An activity theoretic model for information quality change. *First Monday*, 13, 4.