

Predicting the Perceived Quality of Online Mathematics Contributions from Users' Reputations

Yla R. Tausczik, James W. Pennebaker
Department of Psychology
University of Texas at Austin
{tausczik, pennebaker}@mail.utexas.edu

ABSTRACT

There are two perspectives on the role of reputation in collaborative online projects such as Wikipedia or Yahoo! Answers. One, user reputation should be minimized in order to increase the number of contributions from a wide user base. Two, user reputation should be used as a heuristic to identify and promote high quality contributions. The current study examined how offline and online reputations of contributors affect perceived quality in MathOverflow, an online community with 3470 active users. On MathOverflow, users post high-level mathematics questions and answers. Community members also rate the quality of the questions and answers. This study is unique in being able to measure offline reputation of users. Both offline and online reputations were consistently and independently related to the perceived quality of authors submissions, and there was only a moderate correlation between established offline and newly developed online reputation.

Author Keywords

online reputation, community question-answering, information quality

ACM Classification Keywords

H.5.3 Information Interfaces and Presentation: Group and Organization Interfaces—*Computer-supported cooperative work*

General Terms

Experimentation, Human Factors

INTRODUCTION

MathOverflow (MO) is a small online community of question-answering comprised of 3470 active users. Users contribute 'research level math' questions and answers, and everyone is encouraged to participate: 'Remember, MathOverflow is run by you!' Questions like 'Can a positive binary quadratic form represent 14 consecutive numbers?' are posted and an-

swered by community members. The members also rate the quality of the questions and answers.

The site is an example of crowdsourcing, a strategy for achieving high quality content by pooling the knowledge and effort of many people, regardless of their expertise. Wikipedia, the best known example of crowdsourcing, has been found to have a comparable number of errors as traditional encyclopedias written by experts [2]. A comparison of question and answer sites revealed that sites with a wider user base as opposed to a set of experts produced higher quality answers [4].

Many crowdsourcing communities are designed to create a level playing field, in which previous experience and reputation does not hinder contributions. Specifically, by de-emphasizing established reputation, small insights from untrained or less trained users that would otherwise go undeveloped can contribute to the body of knowledge in addition to insights from experts. Many authors have found that even within these projects, there are still major differences between contributors. One study found that power users, those who made the most contributions, had greater influence than first-time users in discussions of which news articles should make it to Wikipedia's front page [7]. An attempt to employ online crowdsourcing to solve an open mathematics problem, a project known as "polymath1", resulted in a proof of a new mathematical result but in the end far fewer individuals than hoped, almost all experienced mathematicians, actually made substantial contributions [3].

Other researchers have taken the position that it may not matter that these communities are not egalitarian. They argue that communities may benefit from combining the advantages of a wide user base and a reputation system that identifies experts and promotes their contributions. Past user behavior on Yahoo! Answers (YA), such as the number of questions answered, was found to be a significant predictor of the quality of answers [9]. Jurczyk and Agichtein present evidence and argue that metrics like high network centrality in giving answers, could identify authors with higher expertise on YA [6].

The central question is: To what extent is reputation related to the quality of contributions? By answering this question, communities can be designed to place the optimal emphasis on reputation. On the one hand, minimizing reputation may give individuals with little or no reputation the opportunity to

Permission to make digital or hard copies of all or part 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. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$10.00.

produce high quality content. On the other hand, an author's reputation may be a good heuristic for the quality of their contributions and may make it easier to identify high quality content.

THE CURRENT STUDY

Collaborative crowdsourcing projects could be improved if there was a greater understanding of the degree to which reputation affected the perceived quality of contributions. We use behavior on MO to measure the importance of four reputation variables in predicting the perceived quality of questions and answers.

Launched in September, 2009, MO consists of 3470 active users (6280 registered accounts). About half of the users use their real names; a majority of these users are academic mathematicians. Most academic users have recognizable offline academic activity such as journal publications.

MO is unique in having a small online community that draws members from a specific offline community—academic mathematicians. The existence of an offline and online community allows us to contrast multiple forms of reputation and expertise on perceived quality of contributions. In addition, the content of the site, research level mathematics, requires extensive knowledge to make good contributions and there are quantifiable differences in the quality of the contributions. Accordingly, it is an ideal site to answer our research questions.

Both offline and online reputations reflect people's standing in the community. In this study, **offline reputation** in the academic mathematics community was measured by the number of publications the contributors had. Online reputation was measured in three ways: **MathOverflow Points (MO Points)**, **authoritativeness**, and **social connectedness**. MO Points are a feature of the software running MO and are based on the ratings of a user's past submissions; it is displayed next to the user's name. The other two measures were selected because they have been used in previous studies and are indirect measures of reputation. Authoritativeness measures how confidently a user worded his or her submission. Network betweenness centrality was measured; it is a social network analysis metric that measures social connectedness.

The one offline and three online reputation variables were regressed against two measures of perceived quality of submissions: the **net number of votes a question received** and the **net number of votes an answer received**. Registered users are encouraged to up vote a question or answer if they find it valuable and to down vote it if they do not. Next to each submission appears a box which gives the total score that it has received. In the context of research level mathematics questions it is difficult to measure objective quality because there are few judges qualified enough to evaluate the content. The perceived quality is a good approximation of objective quality because the MO users voting on the content are themselves some of the few judges who are qualified to evaluate the objective quality.

First, we examined the relationships between the four reputation variables, to test whether these measurements represent four fundamentally different kinds of reputation or if they are actually measuring the same construct. Many question-answering sites, like MO, have begun to implement online reputation systems. The impact of these systems and the relationship between newly created online and established offline reputation is unclear.

Second, we compared the different measures of reputation to see which was the best predictor of perceived quality. We compared different measures of online reputation. MO Points is a direct measure of online reputation and is a record of past achievements. There is some evidence that indirect methods of measuring reputation, like authoritativeness and social connectedness, may be better than direct methods, like MO Points, because they record social processes rather than achievements and are more proximate. For example, in analyzing communication between Enron employees it was found that social connectedness was the best predictor of actually leading projects [5].

We also compared online and offline reputation. Studies of reputation have focused on reputation created on the site. There is some doubt whether those with higher online reputation actually make better contributions; an analysis of edits to Wikipedia made by regular and one-time users showed that both could make high quality edits [1]. Few studies have assessed the importance of offline reputation directly because most communities involve anonymous users and so it is difficult to measure the level of reputation outside of the online community.

Finally and most importantly, we addressed the central research question by assessing the total contribution of reputation in predicting the perceived quality of contributions.

METHOD

Information about registered users and matched user information was collected with records of posting behavior on the MO site. MO moderators publicly post dumps from MO servers with user and post data monthly, allowing access to this publicly available data.

Perceived Quality of Questions and Answers

All posts on MO from May 1, 2010 to June 1, 2010 were downloaded. In all, 953 questions made by a single author and 1213 answers to multi-answer questions (from 422 questions) were collected. There were an average of 32 questions posted per day, each question received an average of 225 views, 3.74 votes, 2.70 comments, and 1.61 answers. The two primary perceived quality measures focused on the rated quality of questions and of answers. Users can vote on the quality of a question based on its novelty, insight, or interest. The more votes, the higher the perceived quality. Users can also vote on the quality of an answer.

User Reputation

For each author of a question or answer, four different measures of reputation were collected.

Table 1. Correlations between author reputation measures.

	(2)	(3)	(4)
offline reputation (1)	0.38**	0.06	-0.08*
MathOverflow Points (2)		0.15***	-0.06
authoritativeness (3)			0.00
social connectedness (4)			

Signif. codes: '.' 0.05, '**' 0.01, '***' 0.001*

Offline reputation Sixty-two percent of users filled out personal details when they registered such as their real name, an academic or personal website, a short self-description, and a location, that allowed us to collect information about their offline reputation. Offline reputation was operationalized as the number of papers a user published in peer-reviewed mathematics journals. The number of publications for each user was collected from their academic website if provided or by searching MathSciNet (a comprehensive database of mathematics articles) using the user's name if provided. Data was analyzed using two approaches by either imputing offline reputation assuming no papers or by excluding missing values. In both cases, offline reputation was significant. Users who did not identify themselves on the site displayed different behavior: they had lower online reputation and did not participate as much. Because there was not a substantial difference in the results from using either approach and in order to preserve information on users with low identity information, the imputed measure of offline reputation was used and an additional dichotomous variable of whether offline reputation was identifiable was also included.

MathOverflow Points MO was designed to display a score for each user based on the number of votes all his or her submissions had received. This score was recorded just prior to the period in which questions and answers were collected, April 29th, 2010.

Authoritativeness The text of each question and answer were processed linguistically using the computer program Linguistic Inquiry and Word Count (LIWC) which counts the rate of word use in psychologically meaningful categories [8]. Three categories of word use were recorded based on their ability to predict existing and manipulated status in past research: total word count, first person plural, and the inverse of first person singular [11]. A composite of these was created by taking the sum of the z-score for each variable.

Social connectedness Social network analysis was performed on users' history of posting to question and answer threads just prior to the period in which questions and answers were collected, April 29th, 2010 [12]. That is an edge was created from Person A to B if B answered a question given by A, edges were weighted by the number of answers provided to questions by A. For each user, betweenness centrality was measured in this network. This has been used as a measure of social connectedness in online communities [10]. The measure of network betweenness centrality was corrected to control for the user's level of participation to separate it from MO Points.

Table 2. The relationship between author reputation and the perceived quality of questions and answers.

	number of votes per question (β)	number of votes per answer (β)
intercept	0.08	-0.01
offline reputation	0.22**	0.12***
MathOverflow Points	0.26***	0.12***
authoritativeness	0.18***	0.00
social connectedness	0.07*	0.03
missing identity	-0.24**	-0.01
lifetime of post	0.01	0.06**
lifetime of user account	0.02	-0.02
amount of math notation	-0.02	0.05*
variance explained by author reputation measures	0.40	0.07

Signif. codes: '.' 0.05, '**' 0.01, '***' 0.001*

Statistical Approach

Separate analyses were conducted on which characteristics of the author predicted highly rated questions and answers. The authors varied greatly in their levels of offline and online reputation. Those that submitted answers had higher reputation scores than those who submitted questions (median values 233 versus 11 MO Points and 4 versus 2 publications).

Multi-level logistic and linear regression models were calculated to estimate the degree to which the status variables predicted the perceived quality of a question or answer. Questions were nested within authors and answers were nested within questions. A few other variables were included in the model as covariates: whether offline reputation was missing, days since the submission was posted, the number of days between the first and last day the author logged onto MO, and the amount of mathematical notation. In the model predicting the number of votes an answer received, the average number of votes per answer to the question was also included to control for the popularity of the question. Measures that were right skewed were log-transformed to correct for deviations in normality. All measures have been standardized so the magnitudes of their coefficients can be compared. Variance explained was calculated for the addition of the reputation variables to the model with covariates.

RESULTS AND DISCUSSION

Relationship between author reputation measures

There was moderate to small overlap in the four reputation variables. The strongest relationship was between offline reputation and MO Points; these two variables were moderately correlated (Table 1). Among measures of online reputation, MO Points and authoritativeness were slightly related, but neither was related to social connectedness. These results suggest that the four measurements of reputation are measuring different types of reputation.

The best predictors of perceived quality

All four reputation variables partially predicted the perceived quality of submissions (Table 2). Offline reputation and MO

Points were the best predictors. MO Points was a better predictor of perceived quality of submissions than the two other measures of online reputation, authoritativeness and social connectedness. The indirect measures were significant predictors for the perceived quality of questions only and they had lower magnitudes of effect. Of the two indirect measures, authoritativeness was more important than social connectedness. These results show that in the case of MO, direct measures of past achievements are the best predictors of the perceived quality. Achievements may be more important than social factors, like how authoritative the submission is worded or how socially connected the user is, because contributions in mathematics are more objectively quantifiable.

Both offline reputation and MO Points significantly and independently predicted the perceived quality of questions and answers. The two forms of reputation had nearly equal magnitudes of effect. MO Points was a slightly better predictor for the perceived quality of questions. The other two forms of online reputation were worse predictors than offline reputation at predicting the perceived quality of submissions. These results demonstrate that online and offline reputation are measuring different forms of reputation and one form of online reputation and offline reputation are nearly equal in contributing to the perceived quality of submissions.

Without some mathematical training it unlikely a user will be able to make a significant contribution to MO. The interesting comparison is between the first year graduate student who will have no mathematics papers and the Fields medalist who may have a hundred. These results show that some of the inexperienced students are able to develop reputation on MO and exert influence; they also show that newly developed reputation on the site does not erase the importance of long-standing reputation—the Fields medalist’s submissions are still prominent. Although not visible, it is likely that the same relationship between online and offline reputation generalizes to other communities of collective action. Future research should investigate whether communities of collective action should be designed to display offline reputation as well as online reputation. Designers are faced with the problem of delicately balancing advertising user’s offline reputation with the possible consequence of suppressing valuable contributions by low reputation individuals against the possible benefit of promoting valuable expertise.

Overall importance of author reputation

Overall, reputation explained a significant, but moderate percentage of the variance. Reputation variables explained about 40% percent of variance in the perceived quality of questions and only 7% of variance in the perceived quality of answers. Not surprisingly reputation seems to be important in making highly rated submissions, but the majority of variance in perceived quality must be attributable to other variables.

Counterintuitively, reputation variables were better predictors of perceived quality in questions than answers. There are two likely explanations. One, readers may be more discriminating in which questions they pay attention to and find interesting. There are far more questions to choose between

than answers to a single question. Two, there is a high threshold for submitting answers. The average online and offline reputations for users submitting answers was far higher than for questions. Reputation may explain the difference between submitting and not submitting an answer, but the difference between a good and a bad answer may have more to do with a single insightful idea than overall expertise and reputation.

CONCLUSION

This study shows that multiple forms of reputation are related to the perceived quality of contributions. Which suggests that both offline and online reputation could be effectively used to identify high quality submissions, although counter intuitively reputation may be better at predicting high quality submissions when there is low barrier for entry. Two issues should be addressed in future research. One, whether there is a systematic effect of displaying an authors reputation information on the quality of content on the site. Two, separating the degree to which reputation effects actual quality of submissions and perceptions of quality above actual quality.

REFERENCES

1. Anthony, D., Smith, S. W., and Williamson, T. Reputation and reliability in collective goods: The case of the online encyclopedia Wikipedia. *Rationality and Society* 21 (2009), 283–306.
2. Giles, J. Internet encyclopaedias go head to head. *Nature* 438 (2005), 900–901.
3. Gowars, T., and Nielsen, M. Massively collaborative mathematics. *Nature* 461 (2009), 879–881.
4. Harper, F. M., Raban, D., Rafaeli, S., and Konstan, J. A. Predictors of answer quality in online Q&A sites. In *Proc. CHI 2008*, ACM Press, 865–874.
5. Hossain, L., Wu, A., and Chung, K. K. S. Actor centrality correlates to project based coordination. In *Proc. CSCW 2006*, ACM Press, 363 – 372.
6. Jurczyk, P., and Agichtein, E. Discovering authorities in question answer communities by using link analysis. In *Proc. CIKM 2007*, ACM Press, 919–923.
7. Keegan, B., and Gergle, D. Egalitarians at the gate : One-sided gatekeeping practices in social media. In *Proc. CSCW 2010*, ACM Press, 131–134.
8. Pennebaker, J. W., Booth, R. J., and Francis, M. E. Linguistic Inquiry and Word Count: LIWC.
9. Shah, C., and Pomerantz, J. Evaluating and predicting answer quality in community QA. In *Proc. SIGIR 2010*, 411–418.
10. Sylvan, E. Predicting influence in an online community of creators. In *Proc. CHI 2010*, ACM Press, 1913–1916.
11. Tausczik, Y. R., and Pennebaker, J. W. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29 (2010), 24–54.
12. Zhang, J., Ackerman, M. S., and Adamic, L. Expertise networks in online communities: Structure and algorithms. In *Proc. WWW 2007*, ACM Press, 221–230.