Mining Stack Exchange: Expertise Is Evident from Initial Contributions

Daryl Posnett*, Eric Warburg*, Premkumar Devanbu*, Vladimir Filkov*
* Department of Computer Science
University of California, Davis
Davis, CA

{dpposnett, ptdevanbu, vfilkov}@ucdavis.edu

***Abstract*—`Stack Exchange` is a very popular Question & Answer internet community. Users can post questions on a wide variety of topics; other users provide answers, usually within minutes. Participants are not compensated for their services and anyone can freely gain value from the efforts of the users; `Stack Exchange` is therefore a gift economy. Users, however, do gain reputation points when other users "upvote" their questions and/or answers. `Stack Exchange` thus functions as a learning community with a strong reputation-seeking element that creates a valuable public good, *viz.,* the Q&A archive. The incentive structure of the community suggests that over time, the quality of the product (*viz..*, delivered answers) steadily improves, and furthermore, that any individual who durably participates in this community for an extended period also would enjoy an increase in the quality of their output (*viz.*, the answers they provide). We investigate the validity of these widely held beliefs in greater detail, using data downloaded from `Stack Exchange`. Our analysis indicates that these intuitions are actually not supported by the data; indeed the data suggests that overall answer scores decrease, and that people's tenure with the community is unrelated to the quality of their answers. Most interestingly, we show that *answering skill*, i.e. getting high average answer scores, which is different than reputation, is evident from the start and persists during one's tenure with the community. Conversely, people providing low rated answers are likely to have done so from the start.**

I. Introduction

It has been well-established in educational & training circles that fostering communities of learners [1] can facilitate the achievement of learning goals. In fields such as software engineering and medicine, learning is a life-time pursuit for professionals seeking to maintain competency. Thus, participating in and sustaining learning communities is a durable and valuable aspect of professional life. Online social networks clearly have a strong role to play here.

Websites such as `Stack Exchange`[1] have arisen to meet this need for learning. Users can post questions and other users post answers. Questions and answers are both rated, and authors accrue rating points. Over time, authors accumulate reputation scores; high reputations are coveted as badges [2] of honor. `Stack Exchange` is very active and diverse, with numerous questions posted daily on topics ranging from mobile application programming to philosophy. Table I shows some descriptive statistics of the largest ten sites under `Stack Exchange`. The numerical data include the number of posts (both questions and answers), distinct posters, questions, answers, and the Question/Answer ratio, in terms of the number of answers per question.

TABLE I: Descriptive statistics of the largest 10 sites, collected in early May 2012

Site	# Posts	# Posters	# Ques	# Ans	A/Q Ratio
StackOverflow	6372657	513436	1966272	4406301	2.24
ServerFault	261317	38201	79416	181894	2.29
SuperUser	260503	44218	85460	175031	2.05
Programmers	87155	13356	12954	74188	5.72
Meta StackOverflow	81060	9470	25715	55345	2.15
Gaming	24321	4304	8608	15688	1.82
Unix	17856	3598	5904	11946	2.02
Photo	14212	1987	3211	10997	3.42
Android	10384	2583	3809	6575	1.73

Questions on `Stack Exchange` are answered very quickly—the median time for question answering on the largest `Stack Exchange` site about programming, `Stack Overflow`, has been reported to be 11 minutes [3]. The site is ranked 2^{nd} among reference sites, 4^{th} among computer science sites, and 97^{th} overall among all websites[2]. This site thus has tremendous impact; understanding its function (and dysfunction) can be very helpful in making it more valuable, and potentially designing other, similar, or better services. In an interview[3] for Wired magazine Jeff Atwood, one of the creators of `Stack Exchange`, stated that the goal of this online community is not just to provide quality answers to the users, but also, when it pertains to software programmers, to trick them into becoming better communicators. Since programmers constitute most of its users, one of the original goals of `Stack Exchange` is to improve over time people's abilities to provide better answers. How can we tell if `Stack Exchange` is achieving this goal?

Due to the public availability of most of `Stack Exchange`'s historical data, this online learning community is providing unprecedented access into the quantity and quality of answers posted. Answers are up-voted and down-voted on

[1] See www.stackexchange.com

[2] See www.alexa.com

[3] http://www.wired.com/wiredenterprise/2012/07/stackoverflow-jeff-atwood/

978-0-7695-5015-2/12 $26.00 © 2012 IEEE
DOI 10.1109/SocialInformatics.2012.67

the Stack Exchange sites by the community based on their utility. This rating system allows the answers seen as most useful by the community to float to the top, and is similar technically to the up/down voting schemes on many popular blogging sites. This provides a valuable source of data for questions concerning the effect of time and person's tenure on answer quality.

An argument can be made for using an individual's answer score history as a measure of their ability to answer questions, or perhaps even of their domain expertise. A high scoring answer has survived the down-votes and has accrued a substantial number of upvotes; such upvotes are evidence of the utility of some answers to other participants.

Providing a useful answer necessitates that the answerer is both a good communicator and in possession of a non-trivial understanding of the question's domain. A history of answer scores that are consistently above the mean answer score betrays a certain expertise in a subject. Then, viewing expertise is a measure of output quality, we define the top echelon of 5% - 10% of answerers, based on their answer scores, as *experts*. These expert participants are different from participants with high Stack Exchange *reputation score*, which is primarily a function of the number of questions answered, a measure of *quantity* rather than quality.

The effect of tenure and experience has been a subject of study in the organizational behavior literature [4]. Unlike a typical bricks & mortar organization, Stack Exchange is in cyberspace; it also operates on very different time-scale, moving much faster, with typical answers coming within minutes. Thus, the effect of experience on Stack Exchange participants is an interesting subject of study. The effect of experience on quality of output has been studied in open-source software development, with results suggesting that focused experience does improve quality [5].

It is quite reasonable, as the Stack Exchange founder Jeff Atwood has expressed, to expect that experience in Stack Exchange has a salutary effect on performance or expertise; the longer an answerer is nurtured by the Stack Exchange community by asking questions or providing answers, the more we expect an observable increase in the utility of their posts. Of course, it is also possible that experience has negligible or even negative effect on quality (as suggested in the management literature [4]).

A. Answer Quality, Experience, and Expertise

In this paper, first, we consider the background effect of the entire community on answer scores over time. Stack Exchange has been active since August 2008. During this period, the size of the community has grown substantially. It is reasonable to expect that as more and more people join the community, the overall quality of community output will improve due to the "wisdom of crowds" effect." Such effects are claimed to exist in open-source projects [6]. This observation is often referred to as "Linus' Law"; specifically, it is claimed that "with many eyeballs, all bugs are shallow", suggesting that program bug-fixing efforts are facilitated by the participation of more users.

Among open-source projects, it well-known that more successful projects tend to have larger developer and bug-reporter communities. Thus one might expect a similar temporal phenomenon in Stack Exchange: as the community grows, the quality of the answers delivered will improve.

Research Question 1: Does the quality of answers change as the community grows? To what is the change attributable?

Next, we turn to the effect of experience on the ability of individuals to better answer questions. Individuals develop a long-term relationship with Stack Exchange; some participants stay for an extended period of time, answer hundreds of questions over a period of years. It is reasonable to expect that individuals gain from such experience, and improve both their innate knowledge and their question-answering skill. This was, indeed, one of the original goals for Stack Exchange (see above). Of course, individuals may answer questions at different *rates*. Some may be frequent respondents; others may let long periods lapse between answers. So we study the effect of a person's actual experience with the site, on their answer quality.

Research Question 2: Does experience, or time spend in the community, have an effect on a person's answer quality?

Likewise, most individuals have fairly low scores, but there are some with sustained high answer scores. These individuals are of great interest. How did they get that way? Was it talent, or experience or both? Do they burst forth into the Stack Exchange firmament as brilliant performers, and thus remain, or do they start more modestly and improve as they spend time in Stack Exchange, and learn from their betters. The converse is also interesting: do people who start off brilliantly decline with time?

Research Question 3: Are good answerers made or born?

This question has two aspects: first, we ask whether it is possible to spot high-performers, the pillars of the Stack Exchange community, and also their direct opposites, very early in their Stack Exchange careers. Are they observable right from the start? Secondly, we ask, (if they stay with Stack Exchange) are their trajectories within Stack Exchange different from those of the lower-performing majority that use and visit Stack Exchange?

B. Background & Theory

Stack Exchange is a relatively new phenomenon, but has attracted a fair amount of early interest from researchers. Stack Exchange can be viewed as an instance of an online learning community, where individuals help each other gain expertise, both in terms of domain knowledge, *and* in answering questions well and to the point. It has long been argued that people learn better in co-operating groups than on their own [1]. The question naturally arises, how can online platforms be designed to facilitate learning goals. Vassileva [7] presented 3 important design goals of learning communities: help users find *relevant content*; help users find *people with relevant knowledge* and *motivate people to learn*. In a study of the factors that influence the great success of `Stack Exchange`, Mamykina *et al.* [3] find that the phenomenal success of `Stack Exchange` arises not just from initial design quality, but also from the continuous engagement of the designers in improving the site. They also investigate how often users play both answering and questioning roles, and find that users that play mostly answering roles account for most of the answers on `Stack Exchange`. Gyöngyi *et al.* [8] study Yahoo! Answers, and investigate the breakdown of users across 3 categories, those who just ask questions, those who just answer, and those who do both. They find that this mix varies across knowledge categories. They also investigate the 3-mode network of users, questions, and answers, to study the relationship of network-based reputation measures to actual community-vote based scores in Yahoo! Answers. The find that network-based based measures, specifically, HITS hub & authority scores [9], are better indicators of final scores than the volume of delivered answers. For example, connections in the 3-mode network with other reputable individuals are better indicators of high Yahoo! Answers scores than simply the volume of answers.

Other researchers have studied the content of Q&A and other knowledge-sharing websites. Programmers, specifically, are in an eternal struggle with knowledge deficits; these deficits arise from the immense and growing complexity of development platforms such as Android. Such APIs (application programming interfaces) are rich & powerful, but dauntingly complex and often inadequately documented. Parnin & Treude [10] find that the availability of such documentation is, in general, very uneven; some parts of the API are well and prolifically documented, while others languish in obscurity. Jiau & Yang [11] find that similar inequalities exist in `Stack Exchange`; however, they give an optimistic spin to this unfair situation, arguing that the strongly reuse-oriented structure of modern APIs allows poorly documented elements to gain “trickle-down” benefits from similar elements that are better-endowed with documentation.

We are interested in one specific aspect of `Stack Exchange`: *the quality of the information delivered to users and viewers*. How does information quality change with time and experience?

II. Methodology

We mined the publicly available historical `Stack Exchange` data. The data contains a rich set of details on every post such as the user id of the poster, the time a question or an answer was posted, the post’s net score, the number of views it received, and what post it was in response to. From this we can construct the *posting history* for each user over time.

To answer our research questions we use boxplots to visually assess the validity of our hypotheses, as well as an advanced statistical measure, the stickiness coefficient, to validate hypotheses over longitudinal data.

A. Measures

An *answer score* is the difference between the number of *up votes* and number of *down votes* for that post. A person’s *tenure* with `Stack Exchange` is the time between their first and last post on the sites. During their tenure we count their *total posts* and distinguish between those that are answers and those that are questions. A person’s *posting history* is an ordered collection of posts and their scores, marked with the time of each post.

B. Stickiness Coefficient

To validate hypotheses in our longitudinal data of `Stack Exchange` users’ answer scores, we use a summary statistic for time-course and longitudinal data called the stickiness coefficient, recently introduced by Gottlieb and Müller [12]. This coefficient takes as input population measurements over time, where a set of measurements for an individual constitute their *trajectory*, with the assumption that the data are generated by a smooth underlying stochastic process. It then computes a single numeric measure of the relationship between measurements at two different time points.

The stickiness coefficient S_X is a normalized measure that takes values on the range $[0, 1]$ and captures the tendency for individual trajectories to stay on one side or the other of the population mean trajectory. $S_X = 1$ indicates perfect stickiness implying that trajectories remain on one side of the mean or the other, whereas $S_X = 0$ indicates no statistically significant tendency for a trajectory to remain on either side of the mean.

Here we are interested in understanding how participants’ mean answer scores vary over time (*i.e.* over their answer histories). Are the deviations from the mean meaningful in some sense with respect to understanding how participants may increase their expertise by answering the questions of others?

III. Results and Discussion

To motivate our hypotheses we use data from the second largest `Stack Exchange` site, serverfault.com. [4] On this site, as of the time of our data collection in early May in 2012, there were 261317 posts, of which 79416 were questions and 181894 were answers. The number of posters was 38201. Of these, more than 35000 posters had posted more than 5 answers each. The average score per answer for those posters was 1.6, with a median of 1.4. The maximal answer score was 35.

[4] We didn’t use stackoverflow.com due to its size proving computationally too expensive. We performed the calculations on all of the smaller sites, and the results presented here are typical.

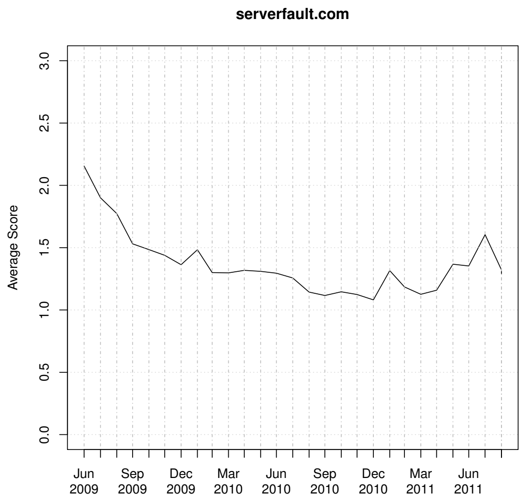

Fig. 1: Average answer scores for all answerers, aggregated weekly, for the serverfault.com site.

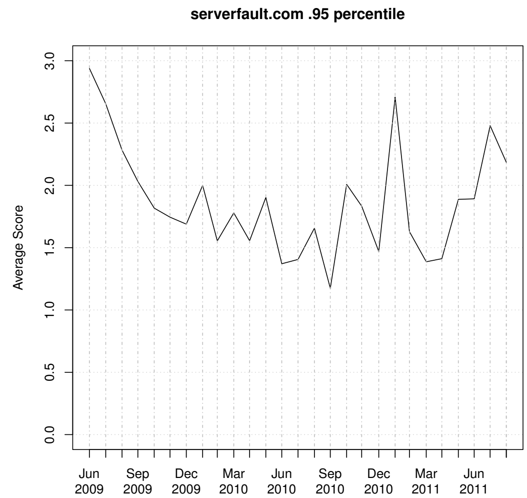

Fig. 2: Average answer scores for the top 95% of scorers, aggregated weekly, for the serverfault.com site. The scores are notably much higher than those for all answerers in Fig. 1.

A. RQ1: Does the quality of answers change as the community grows?

The `Stack Exchange` community grows as more and more people post questions and many answers are provided by the community. In our first study, we sought to quantify how the quality of answers changes with time, as the community size increases. In Fig. 1 we show the average answer score (sum of all answer scores divided by the total number of answers) aggregated in weekly windows of time, over the whole history of the serverfault.com site. The plot shows a general downward trend (with an interesting spike most recently). In other words higher scores are becoming rarer, modulo the last few months. This result is arguably a consequence of a combination of several different competing pressures. As more people join the community, and they are answering questions in pursuit of higher reputation and not necessarily quality, the number of lower scores overwhelms the total. Also, and perhaps as a response to the above, the community is making it more difficult to obtain higher scores on average, *i.e.* it filters out low importance answers quicker.

This trend of decreasing scores is also seen on the plot of the scores for the top 5% of scorers, given in Fig. 2. Also evident is that the top 5% of scores are higher than the remainder as shown in Fig. 1.

B. RQ2: Does time spend in the community have an effect on a person's answer quality?

We divide a person's tenure in two periods of equal length, *viz.*, the former (first half) and latter (second half), and use their answer scores in each period as two distinct populations. In Fig. 3 we show the boxplots of the two populations for the top 5%, left two, and the bottom 95% of the population. The distributions of the scores are virtually identical for the two parts of people's tenure. Thus, we find that there is no observable difference between the scores a person receives in the former and latter part of their career with stackoverflow. This is true for both the top scorers and everyone else. We note that the outliers in the top part of the boxplots for the top 5% are due to these distributions being left-skewed and with long tails.

C. RQ3: Are answerers good/bad from the start?

Here we ask: do people who's initial answers receive higher scores continue to post answers that score well throughout their tenure? And, do people who initially receive low scores continue to provide lower quality answers?

We provide a two-part answer to these question. In the first part, we visualize the differences between the answer score distributions for different categories of answerers (top, middle, and bottom scorers). In the second part we use the stickiness coefficient to demonstrate that early deviations from the mean in the answer scores are sticky, *i.e.* tend to be indicative of performance throughout tenure.

For the first part, we select only those people who have provided more than 10 total answers over their tenure. For each person, we divide all answers they have provided during their tenure into *the first* 5 *answers* and *the rest*, *i.e.* all the other answers. We use the average of answer scores of the first 5 answers. to identify the top 5%, middle 90% and bottom 5% of answerers. For the remainder of the answers we sample 5 answers from each user across their tenure. This three-way split generates three distributions of average answer scores of identical size. Fig. 4 shows the resulting boxplots. There, on the y-axis are the average answer scores, and on x are the distributions for the top 5% of answerers using their first 5 answers (*tf*), middle 90% of answerers using their first 5 answers (*mf*), and bottom 5% of answerers using

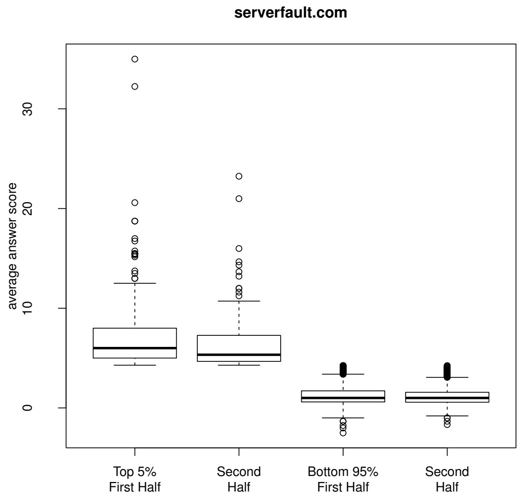

Fig. 3: Box plots of answer scores in the first half of one's tenure with the community (first and third plots) and second half (second and fourth plots), for the top 5% (two left plots) and the bottom 95% (two right plots) of average answer scores. The plots show that the avg scores are indistinguishable between the two halves of one's tenure.

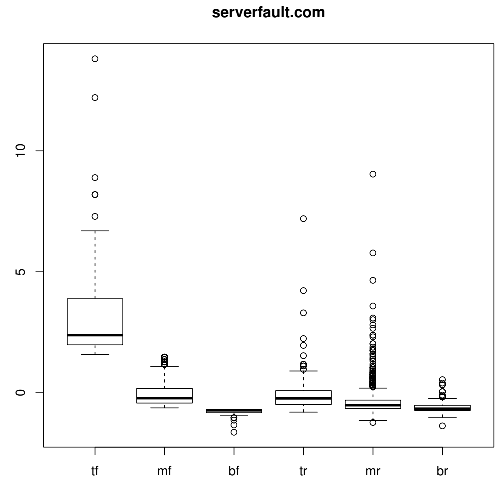

Fig. 4: Comparison of distribution boxplots of the top 5% (t), middle 90% (m), and bottom 5% (b) of scorers based on the average of their first 5 answer scores (f) and the rest of their answer scores (r). Thus, mf indicates the distribution of answer scores for the middle scorers based on their first 5 answer score average.

their first 5 answers (*bf*). The tr, mr, br are the corresponding distributions for the same 3-way split of the top, middle, and bottom answerers; but the average answer scores are calculated using the rest of their answers (all but the first five).

From the plots, it is visually apparent that, based on their first 5 answers, participants in the top 5% of scorers (*tf*) continue to receive higher scores for the remainder of their tenure (tr); while at the same time, the middle and bottom performers perform similarly to their initial scores for the remainder of their tenure. Some regression to the mean is apparent for the top scorers, attributable perhaps to the above noted phenomenon of higher scores becoming increasingly rare.

To formally establish these results, we apply the stickiness coefficient to our longitudinal data of answer scores. In particular, we ask it there is any trending in the trajectory (values over time) of the answer scores of answerers versus the trajectory of the population mean. The stickiness coefficient evaluates the tendency in the answering process for deviations from the mean curve to co-vary over time. That is, if a person receives high scores to their initial answers, will they tend to achieve high answer scores throughout their tenure/life with `Stack Exchange`?

We used the Matlab implementation of the stickiness coefficient by by Gottlieb and Müller [12]. We used answer score histories for each answerer as their trajectory. For each user we use the time of their first post to realign all of their answers to a zero based reference by simply subtracting this minimum time value from each of their post times.

Using the entire population of answerers and their answer scores proved to be impossible practically due to computational limitations of the Matlab code. Instead, we used 200 population samples each of 100 randomly selected answerers. This resulted in a stickiness coefficient of 0.49. The confidence bounds at 95% were $[0.47, 0.51]$.

The high value for this coefficient indicates that an initial deviation from the mean carries throughout one's tenure; that is, the mean quality of a participants initial answers is likely to persist throughout their tenure. Those who answer well initially will continue to do so and those who answer poorly, are unlikely to see dramatic improvement.

Discussion Our findings have two implications. On the negative side, it appears that founder Jeff Atwood's vision of people developing into good answerers through exposure to the benign influence of `Stack Exchange` is challenged by the reality that people's initial performance tends to "stick" by and large; more specifically, the very top 5% and the middle 90% appear to largely perform in the same range through their entire careers. The bottom 5% do show some improvement, although their performance continues to lag behind the rest. On the other hand, the observed "stickiness" of the top 5% suggests a strategy that `Stack Exchange` organizers could exploit: good answerers can be identified reasonably well almost immediately upon their entry into the community. Efforts at this point to groom and retain them could well prove to be beneficial, as our data suggests that these are precisely the people who continue to do really well throughout their tenure with `Stack Exchange`.

IV. Related Work

The balance of questioners and answerers across the two-sided network defined by Q/A sites has received attention from several different perspectives. Kumar *et al.* modeled the evolution of two-sided markets using attachment curves and applied their model to the `Stack Exchange` data treating questioners and answerers as each side of the market [13]. Their results show that `Stack Exchange` exhibits asymmetric network effects where the questioners find greater benefit in the presence of large number of answerers while answerers do not display such a pattern of dependence on the number of questioners. They further show that questioners grow at a quadratic rate while answerers grow only linearly.

Using data from the TurboTax live community with manually labeled expertise, Pal and Konstan explore question selection bias and its relation to status among answerers. They hypothesize that experts will prefer questions with few answers while non-experts will gravitate towards questions that have already received some attention [14]. Their results show that experts are more selective in picking questions with zero existing value. They further demonstrate that selection bias can be used to identify users with potential expertise.

Liu *et al.* estimate the relative expertise of questioners and answerers using a two-player competition model. They conclude that their method outperforms link-based analysis, and further, that answer ranking is an important factor in predicting expertise [15].

Hanrahan *et al.* present a summary perspective of the distribution of reputation in `Stack Exchange` and report that more than half of the users have a reputation score of less than 10 and answer fewer questions than they ask [16]. Using answer latency as a proxy for question difficulty, no significant relationship was found between user expertise and question duration, however, particularly difficult questions are often answered by the questioner to a significant degree.

V. Conclusion

The `Stack Exchange` Q&A community is a rapidly growing social structure that depends on the expertise of select members. Here we studied empirically tenure of posters and quality of answers in this community. We found that as the community grows in numbers, the overall quality of answers provided decreases in general, indicating perhaps that answers are more and more given by non-experts in pursuit of reputation, or that higher scores are increasingly more difficult to obtain.

When looking at how one's answers depend on one's tenure, or experience with the community we found, somewhat unexpectedly, that the number of prior posts do not increase one's answers' quality. This is consistent with the notion that expertise is present from the beginning, and doesn't increase with time spent with the community.

We also show more direct evidence that expertise is apparent from the onset of a participants tenure in the community; in other words, experts join the community as experts, and provide good answers immediately. Conversely, participants who provide low quality answers tend to so so from their earliest interactions.

These are preliminary but engaging studies. Deeper analysis of the relationship between the pillars of the community, *i.e.* top posters and the rest would be informative, especially if their actual behavior overall is different than the rest. One difference could be their focus on topics: is it narrower than the rest of the people? We also are looking at a more robust measure of answer quality, in terms perhaps of the answers ranking.

Acknowledgments

We thank Andrea Gottlieb and Hans-Georg Müller for kindly providing their stickiness coefficient Matlab code and for thoughtful comments on an earlier version of the manuscript. We thank Mehrdad Afshari for comments, discussion, and help with setting up an earlier draft of the manuscript. All authors gratefully acknowledge support from the Air Force Office of Scientific Research, award FA955-11-1-0246.

References

[1] E. Wenger, "Communities of practice: Learning as a social system," *Systems thinker*, vol. 9, no. 5, pp. 2–3, 1998.

[2] J. Antin and E. Churchill, "Badges in social media: A social psychological perspective," *Human Factors*, pp. 1–4, 2011.

[3] 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 2011 annual conference on Human factors in computing systems*. ACM, 2011, pp. 2857–2866.

[4] M. Sturman, "Searching for the inverted u-shaped relationship between time and performance: Meta-analyses of the experience/performance, tenure/performance, and age/performance relationships," *Journal of Management*, vol. 29, no. 5, pp. 609–640, 2003.

[5] F. Rahman and P. Devanbu, "Ownership, experience and defects: a fine-grained study of authorship," in *Proceeding of the 33rd international conference on Software engineering*. ACM, 2011, pp. 491–500.

[6] D. Posnett, A. Hindle, and P. Devanbu, "Got issues? do new features and code improvements affect defects?" in *Reverse Engineering (WCRE), 2011 18th Working Conference on*. IEEE, 2011, pp. 211–215.

[7] J. Vassileva, "Toward social learning environments," *Learning Technologies, IEEE Transactions on*, vol. 1, no. 4, pp. 199–214, 2008.

[8] Z. Gyongyi, G. Koutrika, J. Pedersen, and H. Garcia-Molina, "Questioning yahoo! answers," Stanford InfoLab, Tech. Rep., 2007.

[9] J. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[10] C. Parnin and C. Treude, "Measuring api documentation on the web," in *Proceedings of the 2nd international workshop on Web*, vol. 2, 2011, pp. 25–30.

[11] H. Jiau and F. Yang, "Facing up to the inequality of crowdsourced api documentation," *ACM SIGSOFT Software Engineering Notes*, vol. 37, no. 1, pp. 1–9, 2012.

[12] A. Gottlieb and H.-G. Müller, "A stickiness coefficient for longitudinal data," *Computational Statistics & Data Analysis*, Apr. 2012. [Online]. Available: http://dx.doi.org/10.1016/j.csda.2012.03.009

[13] R. Kumar, Y. Lifshits, and A. Tomkins, "Evolution of two-sided markets," in *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 2010, pp. 311–320.

[14] A. Pal and J. Konstan, "Expert identification in community question answering: exploring question selection bias," in *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 2010, pp. 1505–1508.

[15] J. Liu, Y. Song, and C. Lin, "Competition-based user expertise score estimation," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information*. ACM, 2011, pp. 425–434.

[16] B. 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*. ACM, 2012, pp. 91–94.