Is reputation on Stack Overflow always a good indicator for users' expertise? No!

Shaowei Wang
University of Manitoba
Winnipeg, Canada
shaowei.wang@umanitoba.ca

Daniel M. German
University of Victoria
Victoria, Canada
dmg@uvic.ca

Tse-Hsun Chen
Concordia University
Montreal, Canada
peterc@encs.concordia.ca

Yuan Tian, Ahmed E. Hassan
Queen's University
Kingston, Canada
{y.tian,ahmed}@queensu.ca

2021 IEEE International Conference on Software Maintenance and Evolution (ICSME) | 978-1-6654-2882-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICSME52107.2021.00067

***Abstract*—Stack Overflow (SO) users are recognized by reputation points. The reputation points are often a great avenue for users to build their career profile and demonstrate their expertise in some domains. Prior studies used users' reputation as a proxy to estimate their experience and expertise. However, there are various ways for a user to earn reputation points that do not require much expertise, such as asking high-quality questions. Therefore, it is important to understand the meaning of a high-reputation point and if the reputation could be used as a good indicator for users' expertise and experience on Stack Overflow. In this study, we explore how users earn reputation points on Stack Overflow by mining their reputation-related activities (e.g., asking questions, answering questions, and editing posts). We study the reputation-related activities of 93,053 high-reputation users that have at least 1,000 reputation points. We find that 1) 13.8% of the studied users earn their majority reputation points through asking questions rather than answering questions. 2) In general, most of the posted answers received no or very few reputation points with users gaining their points from a very small proportion of highly-voted answers. 12% of users' entire reputation comes from one single answer. We suggest future research and Stack Overflow introduce a new metric (i.e., v-index) to evaluate the expertise of a user.**

***Index Terms*—Q&A websites, reputation system, developer expertise.**

I. Introduction

Technical question and answer (Q&A) platforms such as Stack Overflow have become more and more important for software developers to share knowledge and make contributions to communities. Stack Overflow (SO) users are recognized by reputation points. A user can earn reputation points through several activities, such as asking good questions and providing useful answers[1]. Such reputation points may reflect the level of expertise and skill that a user has to some extent. For example, if a user posts many high-quality answers that are related to a certain domain, he/she probably is an expert in this domain. Earning more reputation points helps strengthen users' profiles and may even bring more job opportunities[2]. In addition, prior research also uses reputation as a proxy to estimate the expertise and experience of developers [1]–[3]. In other words, reputation is considered an indicator of developers' expertise and skill.

However, is reputation really always a good indicator? For example, a high-reputation user with 1,926 points only asked three questions in Stack Overflow and answered one question. 92% of his/her reputation points come from one asked question[3]. Therefore, in this study, we are interested in understanding how users earn their reputation points on Stack Overflow, and whether reputation is a good indicator to estimate the expertise of a user on Stack Overflow.

In this paper, we conduct an empirical study on 93,053 high-reputation users (reputation point $>$ 1,000) on Stack Overflow (SO) and analyze their reputation-related activities (e.g., asking questions and answer questions). We find that: 1) 13.8% of the studied high-reputation users earn majority (more than 50%) of their reputation points through asking questions. 2) Users gain a significant portion of reputation points from a very small proportion of highly-voted answers. 12% of users' entire reputation comes from one single answer and 35% of their posted answers received no reputation points.

In summary, a notable number of high-reputation users gain their reputation points mostly from asking questions instead of answering questions. In addition, most of the answers that were posted by high-reputation users received no or very few reputation points. Based on the patterns of how users earn their reputation points, we suggest that Stack Overflow and future research probably need to introduce a new metric to measure evaluate the expertise of a user. We propose a metric similar to h-index [4] that measures a user has posted v answers each of which has been upvoted at least v times.

II. Background

Reputation is an incentive system that is used by Stack Exchange websites (i.e., a set of Q&A websites, including Stack Overflow). The current implementation of the incentive system on Stack Exchange websites is designed to encourage users to perform desirable activities by awarding them reputation points. A user can earn reputation points in several ways, such as asking good questions or providing useful answers[4]. Table I summarizes the ways to gain or lose reputation points. In this study, we investigate the activities that users do to receive reputation points on Stack Overflow. For instance, a

[1]http://stackoverflow.com/help/whats-reputation

[2]https://www.quora.com/Is-it-okay-to-mention-Stack-Overflow-reputation-points-in-my-resume

[3]https://stackoverflow.com/questions/32635704

[4]http://stackoverflow.com/help/whats-reputation

2576-3148/21/$31.00 ©2021 IEEE
DOI 10.1109/ICSME52107.2021.00067

Authorized licensed use limited to: Queen's University. Downloaded on July 08,2023 at 00:39:07 UTC from IEEE Xplore. Restrictions apply.

TABLE I: The details of how users gain or lose reputation points.

A user gains reputation points when	**#reputation points**
their question is voted up (i.e., Question upvoted)	+5
their answer is voted up (i.e, Answer upvoted)	+10
their answer is marked as "accepted" (i.e., Answer accepted)	+15
the user mark an answer as an accepted answer (i.e., Accept answer)	+2
their suggested edit is accepted (i.e., Edit accepted)	+2
a bounty is awarded to their answer (i.e., Bounty reward)	+bounty amount
site association bonus (i.e., Site association)	+100
A user loses reputation points when	**#reputation points**
their question is voted down (i.e., Question downvoted)	-2
their answer is voted down (i.e., Answer downvoted)	-2
the user votes down an answer (i.e., Downvote answer)	-1
the user places a bounty on a question (i.e., Bounty proposal)	-bounty amount
one of their posts receives 6 spam or offensive flags (i.e., Spam)	-100

user that asks a question can earn reputation points whenever the question is upvoted (+5) but might lose reputation when it is downvoted (-2). The person that answers a question can earn or lose points in three ways: if the answer is accepted by the user who asks it (+15—some questions might also award a bounty offered by users) or it is upvoted (+10 each time), but might lose points if the answer is downvoted (-2). For example, the accepted answer to this question[5] has yielded more than 27,928 upvotes, which would translate to at least 279,280 points (takes 75% of the user's entire reputation points); the person asking the question has gained at least 21,091 upvotes, which would translate to more than 105,000 points. Note that only 105 users that have more than 279,280 reputation points on Stack Overflow.

III. Research Questions & Data Collection

A. Research questions

Many prior studies used users' reputation points as a proxy to estimate their expertise on SO [1]–[3]. However, little is known about how users obtain their reputation and whether a high reputation of a user suggests his/her high level of expertise? For example, a user has a high reputation, while all his/her reputation points were made through asking questions rather than answering others' questions. In such a case, the reputation is not a good proxy for his/her expertise. Therefore, in this section, we investigate where users' reputation points come from on Stack Overflow. For example, do users earn their majority reputation points by answering questions or ask questions? Intuitively, answering questions is the most direct way to reflect the expertise of some domains. In academia, the h-index is used to measure the impact of researchers' output in terms of both quality and quantity [4]. We also are interested in investigating how users earn their reputation points from their posted answers. In the current reputation system, it only considers the total reputation points received by posted answers of a user, but lacks the consideration of the quality of all posted answers. One very highly-voted answers can make a remarkable points for a user and bias the user's overall reputation. Do the majority of posted answers receive zero and very few upvotes? Do the majority reputation points of a user come from one or two posted answers? By knowing this, we can know what a high-reputation point indicates and whether the reputation is a good indicator for the level of skill and expertise. Hence, we formulate our study by answering the following two research questions:

- RQ1: How do users earn their reputation points through different activities?
- RQ2: How do users earn their reputation points from their posted answers?

B. Data collection

We downloaded the data dump which is published by Stack Overflow in March 2017[6]. To study how users earn reputation points, we collect our studied users based on the following criteria: 1) users that have at least 1,000 reputation points; 2) users that have registered in Stack Overflow for more than 1 year. We choose such criteria to ensure that the studied users are active on Stack Overflow and have enough time to participate in the various activities to earn reputation points (e.g., asking and answering questions, and editing posts). We ended up with 93,053 users and 24,085,374 posts that were associated with these users. After collecting the user data, we extract the activities that are related to earning reputation points (Table I) and use the data for further analysis. The dataset is available online[7].

IV. Results

A. Approach & Results of RQ1

Approach: To understand how users earn reputation points from different reputation-related activities. We first calculate the distribution of reputation earned by each studied user, and calculate the percentage of these users that earned their major reputation points across different activities. We next investigate the relationship between the reputation points of users and the proportion of their reputation points that come from answer upvotes.

Results: 13.8% of the studied high-reputation users earn the majority (i.e., more than 50%) of their reputation points through receiving upvotes from their posted questions. Figure 1 presents the percentage of users that earn the majority (more than 50%) of their reputation points in a specific way. We observe that 78.5% of the users earn their majority reputation through receiving upvotes from their posted answers. Although the majority of the studied users gain their reputation points from answering questions, which may reflect their expertise on some domains; 13.8% of the high-reputation users mostly gain their reputation points from

[5]https://stackoverflow.com/questions/11227809/11227902

[6]https://archive.org/details/stackexchange

[7]https://zenodo.org/record/5162250

Authorized licensed use limited to: Queen's University. Downloaded on July 08,2023 at 00:39:07 UTC from IEEE Xplore. Restrictions apply.

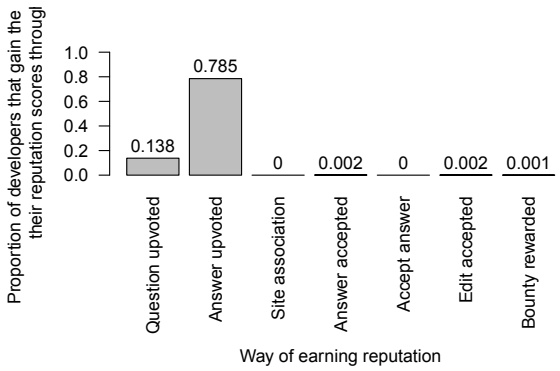


Fig. 1: The percentage of users that earns the majority (more than 50%) of their reputation points in a specific way. Note that 7.2% of users earn reputation points evenly from different activities, so that they do not fall in any of the specific ways.

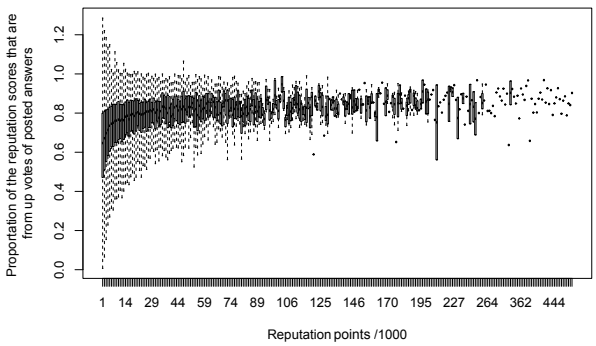


Fig. 2: The proportion of the reputation of users that are earned from the upvotes of posted answers. In some case, the proportion is larger than 1. This is because such users also lost reputation points, which leads to the total reputation points is less than the reputation points that are earned from the received upvotes of posted answers.

asking questions. For example, a user who has a reputation of 38,202 as of Sep 2018, gains more than 90% of the user's reputation points from asking questions. More specifically, 89% of the user's entire reputation points come from only one single question which received 6,808 upvotes makes him 34,040 points. We rarely observe other common ways for users to earn reputation points other than posting questions and answers. For example, we only see 0.2% (166 out of 94,503 users) of users earn their majority of reputation through editing posts (Stack Overflow limits the points a user can receive from editing operations up to 1,000). Another example is bounty; only a user with more than 50 reputation points is eligible to propose a bounty. Such rules limit the number of questions with bounties on Stack Overflow. According to the data of March 2017, there are around 160,000 questions with bounties, taking only 1% of all questions on Stack Overflow.

Users with more reputation points are more likely to rely on receiving upvotes from posted answers to earn reputation. Figure 2 presents the relationship between the total reputation points that a user has and the proportion of the reputation points that are earned from received upvotes of his/her posted answers. We notice that the proportion increases as the reputation points increase from 1,000 to around 40,000 and keeps stable (with a small fluctuation) afterward. For example, the first user whose reputation passed 1 million, Jon Skeet obtained almost all his reputation points by answering 34,077 questions and receiving 374,173 upvotes from these answers[8]. Such finding indicates that it is more reliable to use the reputation points of high-reputation users as a proxy of the expertise for users with higher reputation compared with lower reputation users, since such high-reputation users are more likely to gain their reputation through solving questions.

[8]https://stackoverflow.blog/2018/01/15/thanks-million-jon-skeet/?utm_source=so-owned&utm_medium=hero&utm_campaign=jon-skeet-milestone

78.5% of the users earn major of their reputation points through receiving upvotes from their posted answers, while 13.8% of the studied high-reputation users earn the majority of their reputation points through asking questions. Users with more reputation points are more likely to rely on receiving upvotes from their posted answers to earn reputation points.

B. Approach & Results of RQ2

Approach: To understand whether a user receives the most of their reputation points from a small portion of answers, we examine the distribution of the reputation points of answers (e.g., is the distribution peaked or flat). If all answers of a user contribute a similar amount of reputation points, then the distribution is flat; otherwise, if some answers contribute much more reputation points than the average, the distribution is peaked. We use Kurtosis [5] to measure the peakedness of a distribution. The Gaussian distribution has a Kurtosis of 3. A Kurtosis of higher than 3 means that the distribution has a higher peak than the Gaussian distribution, while a Kurtosis lower than 3 means that the distribution is flatter. A large Kurtosis indicates that the user obtains his/her reputation from a small portion of answers; otherwise, it indicates that the user receives reputation points evenly from each answer. To better understand the proportion of reputation points contributed by different answers for each user, we compute the proportion of answers that get zero reputation points over all answers and the proportion of reputation points contributed only by the top 5 most scored answers.

Results: The reputation points that are received by answers are very skewed. Figure 3 presents the distribution of Kurtosis values that are calculated based on the reputation points of posted answers for each user. We notice that 97.7% of the studied users have a peaked distribution of reputation points

Authorized licensed use limited to: Queen's University. Downloaded on July 08,2023 at 00:39:07 UTC from IEEE Xplore. Restrictions apply.

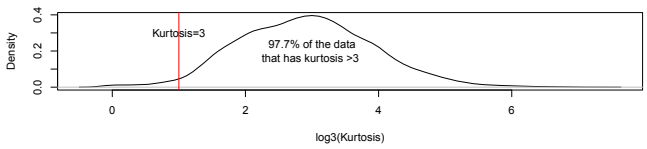
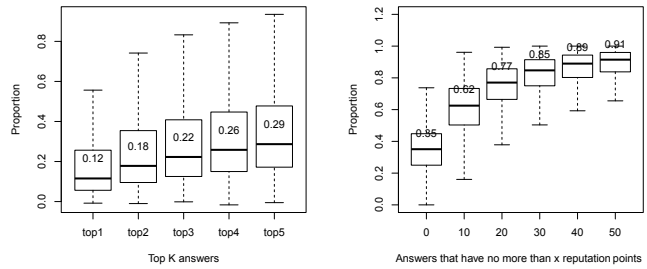


Fig. 3: Distribution of Kurtosis values that are computed based on the distribution of reputation points made by posted answers of each user.



(a) (b)

Fig. 4: The proportion of the reputation points that are obtained from the top k answers over all reputation points of each user (a). The proportion of answers that gain less than x reputation points per answer of each user (b).

that are obtained from answer upvotes, which suggests that in general, the reputation points are not evenly contributed among a user's answers.

35% (median value) of the answers posted by users contribute nothing to their total reputation. 12% of users' entire reputation is made by one single answer. We examine the proportion of the posts that receive very few reputation points in Figure 4 (b). 35% and 62% (median value) of the answers contribute nothing and no more than 10 reputation points to their total reputation pool. Figure 4 (a) presents the median proportion of the reputation points that are obtained from the top k answers over all reputation points of users. The top one answer contributes 12% of the entire reputation points of his/her user and that the top 5 answers contribute 29%. Note that the median number of answers that were posted by the studied users is 60. For example, a user[9] has 1,156 reputation points as of Feb. 21st, 2018, and 72% of his reputation points are earned through one answer[10], which received 84 upvotes. The user posted 9 answers; however, three of them received no upvote and two of them received only one upvote. These observations suggest that most of the answers posted by users do not contribute much to their reputation. Only a small number of "golden" answers (i.e., the highly-voted ones) contribute a significant proportion of reputation points.

[9]https://stackoverflow.com/users/3439/marcus-ericsson?tab=profile
[10]https://stackoverflow.com/questions/1809484#1973304

Academia	correlation	Stack Overflow	correlation
h-index vs. #paper	0.52	v-index vs. #answers	0.63
h-index vs. #citation	0.93	v-index vs. #upvotes	0.80

TABLE II: The correlation between h-index and number of papers and number of citations of research in academia and the correlation between v-index and number of answers and upvotes of users in Stack Overflow.

In general, most of the posted answers received no or very few reputation points and users earn their reputation points from a very small proportion of highly-voted answers. 35% (median value) of the answers posted by users contribute nothing to their total reputation pool. On median, 12% of users' entire reputation points come from one answer.

V. Discussion

In Section IV-A, we observe that 13.8% of the studied users earn their majority reputation points by asking questions. In such cases, even the reputation of a user is high, it does not indicate that she/he has high-level expertise in certain domains. Even for the users getting their points mainly from answering questions, we observe that most of their answers do not receive any reputation points. In general, the majority of reputation points of a user are earned from a small proportion of golden answers, even the number of such answers is small. In other words, the reputation of a user can be biased by a very small portion of highly-voted answers.

Given such a situation in Stack Overflow, h-index [4] which is used in academia probably could be applied to the users of Stack Overflow. For example, Stack Overflow probably could introduce a metric called ***v-index*** that measures a user has posted v answers each of which has been upvoted at least v times, which is similar to the h-index that is used to measure the impact of a researcher in academia. To test our assumption, we download the real-world academic dataset[11] from ArnetMiner [6]. We aim to examine whether the patterns between the h-index and the number of citations and papers are similar to those between the v-index and the number of answers and upvotes. patterns We present the relationship between h-index and the number of papers and citations in real-world academic and the relationship between v-index and the number of the posted answers and upvotes in Stack Overflow, respectively, in Figure 5. We also compute the Spearman correlation for them and the results are listed in Table II. We observe that the h-index and v-index have similar characteristics. Therefore, **Stack Overflow and future research probably should consider introducing a new metric, which is similar to h-index, called v-index to evaluate the expertise of a user.** Furthermore, Stack Overflow also encourages users to contribute via other activities (e.g., editing posts), therefore future research should also consider other activities when estimating user's expertise and experience.

[11]The dataset is publicly available at http://arnetminer.org/citation

Authorized licensed use limited to: Queen's University. Downloaded on July 08,2023 at 00:39:07 UTC from IEEE Xplore. Restrictions apply.

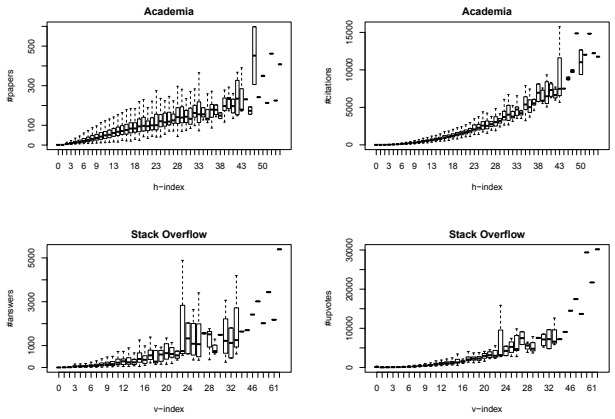


Fig. 5: The relationship between the h-index and the number of papers and citations in real-world academic. The relationship between the v-index and the number of posted answers and upvotes in Stack Overflow.

VI. Threats to Validity

Internal Validity In this study, we only look at the high-reputation users that have at least 1,000 reputation points. It is not clear whether our findings still hold for the low-reputation users. To alleviate this issue, we randomly sample 50,000 users whose reputation point ranges from 100 to 1,000. We apply the same analysis that we describe in Section IV-A. I find that more users (22%) earn their majority reputation from the questions they asked compared with the users having at least 1,000 reputation points.

External Validity One external threat is that it is not clear whether our findings still hold on other Q&A websites. Another threat is regarding to the factors that we considered, since there might be additional factors that could be more relevant to the lasting time of the value of an answer. However, our results show that the explanatory power of our models is very high when using the studied factors. Future studies should investigate more Q&A websites and consider more factors.

VII. Related Work

Prior research studied how users interact with the incentive system of Q&A websites. Bosu et al. investigates how to build reputation through answering questions on Stack Overflow and provided suggestions for new contributors [7]. Movshovitz-Attias et al. analyzed the participation patterns between high and low reputation users on SO [8]. Wang investigated the reputation between men and women on SO [9]. Different from these studies, we focus on examining whether users' reputation is a good indicator for their expertise. A number of studies focus on studying other incentive system (e.g., bounties and badges) of Q&A websites [10]–[15]. For instance, Anderson et al. studied how user behavior is steered by the badges on Stack Overflow [10]. Cavusoglu et al. provided evidence to confirm the value of the incentive system to simulate voluntary participation [11]. Zhou et al. investigated the impact of bounties on question answering [14].

VIII. Conclusion

In this paper, we conduct an empirical study on 93,053 high-reputation users (i.e., the ones with more than 1,000 reputation points) and analyze their reputation-related activities. We find that 1) 13.8% of the studied users earn their majority reputation points through asking questions rather than answering questions. 2) In general, most of the posted answers received no or very few reputation points with users gaining their reputation points from a very small proportion of highly-voted answers. We suggest future research and Stack Overflow introduce a new metric to measure the expertise of a user on Stack Overflow, such as v-index we proposed.

References

[1] H. Zhang, S. Wang, H. Li, T.-H. P. Chen, and A. E. Hassan, "A study of c/c++ code weaknesses on stack overflow," *IEEE Transactions on Software Engineering*, 2021.

[2] 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*, 2012, pp. 91–94.

[3] P. Morrison and E. Murphy-Hill, "Is programming knowledge related to age? an exploration of stack overflow," in *10th Working Conference on Mining Software Repositories (MSR)*. IEEE, 2013, pp. 69–72.

[4] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National academy of Sciences*, vol. 102, no. 46, pp. 16 569–16 572, 2005.

[5] L. T. Decarlo, "On the meaning and use of kurtosis," *Psychological Methods*, pp. 292–307, 1997.

[6] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, "Arnetminer: Extraction and mining of academic social networks," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '08, 2008, pp. 990–998.

[7] A. Bosu, C. S. Corley, D. Heaton, D. Chatterji, J. C. Carver, and N. A. Kraft, "Building reputation in stackoverflow: an empirical investigation," in *10th Working Conference on Mining Software Repositories (MSR)*. IEEE, 2013, pp. 89–92.

[8] D. Movshovitz-Attias, Y. Movshovitz-Attias, P. Steenkiste, and C. Faloutsos, "Analysis of the reputation system and user contributions on a question answering website: Stackoverflow," in *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE, 2013, pp. 886–893.

[9] Y. Wang, "Understanding the reputation differences between women and men on stack overflow," in *25th Asia-Pacific Software Engineering Conference (APSEC)*. IEEE, 2018, pp. 436–444.

[10] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Steering user behavior with badges," in *Proceedings of the 22nd International Conference on World Wide Web*, ser. WWW '13, 2013, pp. 95–106.

[11] H. Cavusoglu, Z. Li, and K.-W. Huang, "Can gamification motivate voluntary contributions?: The case of stackoverflow Q&A community," in *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, 2015, pp. 171–174.

[12] B. Vasilescu, A. Serebrenik, P. Devanbu, and V. Filkov, "How social Q&A sites are changing knowledge sharing in open source software communities," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW '14, 2014, pp. 342–354.

[13] G. Hsieh, R. E. Kraut, and S. E. Hudson, "Why pay?: Exploring how financial incentives are used for question & answer," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '10, 2010, pp. 305–314.

[14] J. Zhou, S. Wang, C.-P. Bezemer, and A. E. Hassan, "Bounties on technical q&a sites: a case study of stack overflow bounties," *Empirical Software Engineering*, vol. 25, no. 1, pp. 139–177, 2020.

[15] S. Wang, T.-H. Chen, and A. E. Hassan, "How do users revise answers on technical q&a websites? a case study on stack overflow," *IEEE Transactions on Software Engineering*, pp. 1024–1038, 2018.

Authorized licensed use limited to: Queen's University. Downloaded on July 08,2023 at 00:39:07 UTC from IEEE Xplore. Restrictions apply.