How Are Programming Questions from Women Received on Stack Overflow? A Case Study of Peer Parity

Savannah Morgan
Centre College
Danville, Kentucky, USA
savannah.morgan@centre.edu

Abstract

Stack Overflow is a useful Q&A online community for programmers. However, it has not been as popular among programmers who are women. Prior studies have shown low participation of women programmers on Stack Overflow, but have not demonstrated how the presence of their peers may promote activity. In this work, we understand how the presence of women, through peer parity, can encourage women to participate more on Stack Overflow. We extracted 3,2209,817 posts from the most recent Stack Overflow data dump. By modifying a popular gender computing tool, we took the first names of online users and identified the gender of 2,502,438 users from the 5,987,284 list of current users. We randomly selected peer parity and non peer parity questions asked by women and identify differences in participation on those posts. We found that women become more active after they have taken part in a peer parity post. With further analysis of peer parity through eye tracking and tone detection tools we can determine ways to foster a more welcoming community that would encourage women to participate.

CCS Concepts • **Social and professional topics** → **Women**; • **Human-centered computing** → *Empirical studies in HCI*; • **Software and its engineering** → *Collaboration in software development*;

Keywords Stack Overflow, peer parity, gender, women in computing

ACM Reference Format:
Savannah Morgan. 2017. How Are Programming Questions from Women Received on Stack Overflow? A Case Study of Peer Parity. In *Proceedings of 2017 ACM SIGPLAN International Conference on Systems, Programming, Languages, and Applications: Software for Humanity (SPLASH Companion'17).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3135932.3135952

SPLASH Companion'17, October 22–27, 2017, Vancouver, Canada

ACM ISBN 978-1-4503-5514-8/17/10…$15.00
https://doi.org/10.1145/3135932.3135952

1 Research Problem and Motivation

On Stack Overflow, programmers are able to gain privileges by asking questions, commenting on questions, and answering questions. Women programmers make up only 7.6% of the Stack Overflow community [3]. The programming community is missing potential insight due to the lack of participation from women on Stack Overflow. Studies suggest that peer parity encourages minorities to participate in communities so we investigate the stack overflow community by analyzing posts with peer parity. Peer parity is the concept that multiple individuals of the same group are present in a community where they are considered the minority [1]. In our work, we classify peer parity posts as posts where more than one woman is present and non peer parity posts as posts with only one woman. To narrow our focus, we only observe questions asked by women. Our study is motivated by the following research question:

How is the community responding to questions asked by women?

2 Background and Related Work

On Stack Overflow, men are more active and gain higher reputations than women users [5]. The lower rates have been attributed to various barriers that deter women more than men. *Paradise Unplugged* determined the barriers discouraging women more than men to be intimidating community size, perception of slacking, stranger discomfort, fear of negative feedback and feeling qualified enough to participate [2]. People like to work with individuals who are similar to them so we expect that the presence of women peers in the Stack Overflow community will increase women participation [7]. If women were to find peers on a post then we expect less fear of negativity, an increase in confidence and a more inclusive community. In addition, increasing women participation would not only benefit women programmers, but the community as a whole. Studies have shown that gender diversity enhances team productivity [6]. Therefore, understanding

the reasons behind the absence of women could provide insight into solutions that would improve women developers prospects and strengthen the overall community.

3 Approach and Uniqueness

To investigate peer parity, we extracted 32,209,817 posts from a Stack Overflow data dump including tags, badges, and display names. We then modified a popular gender computing tool to use first names of online users and identified the gender of 2,502,438 users from the 5,987,284 list of current users. Previous studies used a gender computing tool that predicted the gender of first names based on an algorithm, but our modified tool uses direct comparisons to determine the gender of users [4]. To ensure we had the ground truth, we modified the tool so that it only identified the user's gender if the name is found in our lists of common names with inherent genders. We compared the differences in the computed genders of previous tools with our own and found 819,823 differences. More specifically, 41,826 male users were determined as women users by the previous gender computing tool. Next, we gathered all question posts asked by a user that was determined to be a woman. We separated the 510,915 posts into posts with multiple unique women answering/commenting on the question and posts with one unique woman answering/commenting on the question. We then randomly selected 114 peer parity posts and 114 non peer parity posts asked by women. We compared differences in participation among peer parity and non peer parity posts in the sample, by identifying factors affecting women participation on Stack Overflow which are listed in Figure 1. After gathering the information on the 228 posts we analyzed the data to address our research question.

FACTORS	MOTIVATIONS
User Reputation	To discover the type of women users that are more likely to ask questions
Reputation of women on the post	To discover if women with lower reputation are drawn to peer parity posts
The difference in the number of questions asked by the user after the post in comparison to before the post	To analyze how peer parity affects women participation
The score of the question	To discover how the community responds to questions asked by women
The score of the accepted answer	To discover how the community responds to posts made by women
Tags	To analyze how the topic affects women participation
The number of women and men active on the post	To analyze how the gender breakdown affects activity and peer parity
The number of comments, answers and users on the post.	To analyze how the community responds to questions asked by women

Figure 1. The factors being analyzed and the motivations behind selecting these factors.

4 Results and Contributions

From our data set, we found that 97.8% of posts with women are non-peer parity posts. Therefore, our results show that it is uncommon for a woman to receive an answer or comment from another woman on a post. The low percentage of peer parity justifies our interest in peer parity and its influence on user activity. One interpretation of the results could be that the the low rates of participation can be attributed to the lack of peer parity on posts. 76.2% of posts where peer parity did exist, the question was asked by a woman. In contrast, only 20.2% of the non peer parity posts were asked by a woman. Our results signify that there are more women on posts initiated by women. Therefore, it is reasonable to assume that encouraging women to ask more questions would increase the presence of women on stack overflow. Once we established that there was a low percentage of peer parity posts we wanted to see the impact the peer parity posts had on the women users. 52.6% of women asked more questions after interacting with other women users, but there were also women who asked less questions after encountering other women on the post. This leads us to believe that there may be unique features about the posts following the question being examined that could potentially discourage women from returning.

5 Future Work

Our results are preliminary, providing us with a groundwork to build upon in future studies. To further this study we plan to increase our sample size and conduct a deeper analysis of the posts. We plan to analyze the type of posts that the user is active on after experiencing peer parity so we can have a better understanding of what causes women to stop being active and how peer parity affects user participation. Furthermore, we believe there to be additional factors affecting the presence of peer parity such as the tone of the comments and answers. Studies have shown women are more susceptible to fear negative feedback therefore, the number of users providing helpful input on the post and the scores could encourage or deter women from returning. A tone detection tool will be used on a sample of peer parity posts and non peer parity posts to see if posts with harsher tones have less women on the post. It would also be useful to use eye tracking tools on both male and female subjects in future studies to determine if users look at display names when participating on a post. Conducting a study with eye tracking tools would improve the validity of our results and determine how often users seek the presence of their peers when they consider posting on Stack Overflow. These future studies will help us determine if peer parity is beneficial to women on stack overflow.

Acknowledgments

I would like to thank Denae Ford and Dr. Chris Parnin for supervising this work. This material is based upon work supported by the National Science Foundation under Grants No.1559593.

References

[1] Denae Ford, Alisse Harkins, and Chris Parnin. 2017. Someone Like Me: How Does Peer Parity Influence Participation of Women on Stack Overflow?. In *IEEE Symposium on Visual Languages and Human Centric Computing*. IEEE, to appear.

[2] Denae Ford, Justin Smith, Philip J. Guo, and Chris Parnin. 2016. Paradise Unplugged: Identifying Barriers for Female Participation on Stack Overflow. In *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering (FSE 2016)*. ACM, New York, NY, USA, 846–857. https://doi.org/10.1145/2950290.2950331

[3] Stack Overflow. 2017. 2017 Stack Overflow Developer Survey. https://stackoverflow.com/insights/survey/2017

[4] Bogdan Vasilescu, Andrea Capiluppi, and Alexander Serebrenik. 2012. Gender, representation and online participation: A quantitative study of StackOverflow. In *International Conference on Solcial Informatics*. IEEE, 332–338.

[5] Bogdan Vasilescu, Andrea Capiluppi, and Alexander Serebrenik. 2013. Gender, representation and online participation: A quantitative study. *Interacting with Computers* 26, 5 (2013), 488–511.

[6] Bogdan Vasilescu, Daryl Posnett, Baishakhi Ray, Mark G. J. van den Brand, Alexander Serebrenik, Premkumar Devanbu, and Vladimir Filkov. 2015. Gender and tenure diversity in GitHub teams. In *CHI Conference on Human Factors in Computing Systems (CHI)*. ACM.

[7] KY Williams and CA OâĂŹReilly III. 1998. Demography and Diversity in Organisations: A review of 40 years of research in BM Staw and LL Cummings (eds) Research in Organisational Behaviour Vol. 20. *Jai Pres, Connecticut* (1998).