

Graduate Research Plan Statement

I'm interested in empirical software engineering research, specifically focused on software developers. In previous work, I have studied disengagement of contributors from open source software [1], as well as how developer teams are affected by remote work caused by COVID-19 [2]. I am interested in continuing along this path and further studying software developers, specifically within OSS ecosystems.

Within the realm of modern software, open source software (OSS) is commonly thought of as the infrastructure. OSS allows for rapid innovation by reducing overhead for building new projects. However, despite the utility afforded by many OSS projects, many still struggle to attract and retain enough contributors to survive. Coelho et al. performed a study on 104 popular depreciated projects on GitHub and found that 41% cited developer related issues, one of the most commonly cited ones being lack of time of main contributors [3]. Previous research has also found that developers leaving OSS projects can have a negative impact on project quality [8]. These findings highlight the need to understand developer engagement within OSS because of disengagement's impact on not just project quality and survival.

Studying OSS developers and their contributions is a complex and well researched area [4]. Previous research has used a vast number of different metrics to identify and measure factors that impact OSS engagement. There has been a great amount of work done studying developer onboarding [5], and factors that affect the likelihood that a developer contributes to or sticks with a project [6].

However, there has been less work studying the disengagement of open source contributors. The onion model is a popular model for the developer structure within open source projects [7], with a small number of 'core' contributors at the center with a large amount of knowledge and often responsibilities, surrounded by increasingly peripheral developers who contribute less frequently and regularly. The cost of disengagement in general to a project is high, however the impact is often different depending on the developer [8]. The disengagement of core contributors is very expensive, can cause issues like knowledge loss [9], and can decrease the overall probability of project survival [10]. In my first paper [1], I studied technical and occupational factors that affected core contributor disengagement.

While there are many different kinds of OSS contributors, during the first paper, we generalized them into two categories, paid and volunteer contributors. We chose this categorization because they by definition have fundamentally different motivations for participating in OSS.

One notable finding from the first paper was that contributors with different motivations (paid vs. volunteer) tend to disengage for different reasons. This showed we cannot create a one-size-fits-all understanding OSS developer disengagement. Instead, we need a more nuanced understanding of the differences behind the kinds contributors and their reasons for disengaging.

To accomplish this, I plan to use a mixed methods methodology, using open-ended interviews with both paid and volunteer core OSS contributors who have disengaged to see their reasons for disengaging in more depth. I also plan to use publicly available trace data from GHTorrent [11], which is a database of publicly available GitHub trace data to identify disengaged contributors, and also operationalize new factors that affect disengagement based on the contributor interviews.

From the first paper, we learned that the factor that had the strongest impact on the risk of disengagement was experiencing a recent educational or professional transition. Experiencing a transition increased the risk of disengagement by a factor of 2.48. This factor was operationalized using publicly available CV data. There are additional factors that may have a significant impact on the risk of disengagement for a contributor, that exist outside the space of GitHub like this one. In future work, I am also interested in operationalizing additional factors based on previous research and also our own future survey and

interview findings to help identify additional data-driven factors that are important components of the full perspective of core contributors.

This work will help us understand more specifically why volunteer core contributors disengage. By understanding the differences between the kinds of core contributors and their motivations for disengaging, we will be able to take steps towards attempting to understand and eventually prevent disengagement, thus helping increase the quality and survival of OSS projects and ecosystems. Thank you for your consideration.

References:

- [1] Miller, C., Widder, D. G., Kästner, C., & Vasilescu, B. (2019, May). Why do people give up flossing? A study of contributor disengagement in open source. In *IFIP International Conference on Open Source Systems* (pp. 116-129). Springer, Cham.
- [2] Miller, C., Rodeghero, P., Storey, MA., Ford, D., & Zimmermann, T. (2021) "How Was Your Weekend?" Software Development Teams Working From Home During COVID-19, *Under Review*
- [3] Coelho, J., & Valente, M. T. (2017, August). Why modern open source projects fail. In *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering* (pp. 186-196).
- [4] Roberts, J. A., Hann, I. H., & Slaughter, S. A. (2006). Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the Apache projects. *Management science*, 52(7), 984-999.
- [5] Casalnuovo, C., Vasilescu, B., Devanbu, P., & Filkov, V. (2015, August). Developer onboarding in GitHub: the role of prior social links and language experience. In *Proceedings of the 2015 10th joint meeting on foundations of software engineering* (pp. 817-828).
- [6] Steinmacher, I., Gerosa, M. A., & Redmiles, D. (2014, February). Attracting, onboarding, and retaining newcomer developers in open source software projects. In *Workshop on Global Software Development in a CSCW Perspective*.
- [7] Crowston, K., Annabi, H., Howison, J., & Masango, C. (2004, November). Effective work practices for software engineering: free/libre open source software development. In *Proceedings of the 2004 ACM workshop on Interdisciplinary software engineering research* (pp. 18-26).
- [8] Foucault, M., Palyart, M., Blanc, X., Murphy, G. C., & Falleri, J. R. (2015, August). Impact of developer turnover on quality in open-source software. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering* (pp. 829-841).
- [9] Izquierdo-Cortazar, D., Robles, G., Ortega, F., & Gonzalez-Barahona, J. M. (2009, January). Using software archaeology to measure knowledge loss in software projects due to developer turnover. In *2009 42nd Hawaii International Conference on System Sciences* (pp. 1-10). IEEE.
- [10] Foucault, M., Palyart, M., Blanc, X., Murphy, G. C., & Falleri, J. R. (2015, August). Impact of developer turnover on quality in open-source software. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering* (pp. 829-841)
- [11] Gousios, G. (2013, May). The GHTorrent dataset and tool suite. In *2013 10th Working Conference on Mining Software Repositories (MSR)* (pp. 233-236). IEEE.