To Disengage or Not to Disengage: A Look at Contributor Disengagement in Open Source Software

Philip Gray
New College of Florida
Sarasota, Florida, USA
philip.gray24@ncf.edu

ABSTRACT

Contributors are vital to the sustainability of open source ecosystems, and disengagement threatens that sustainability. We seek to both strengthen and protect open source communities by creating a more robust way of defining and identifying contributor disengagement in these communities. To do this, we collected a large amount of grey literature relating to contributor disengagement and performed a qualitative analysis in order to better our understanding of why contributors disengage.

CCS CONCEPTS

• **Software and its engineering → Open source model**.

KEYWORDS

open source, disengagement, grey literature

ACM Reference Format:
Philip Gray. 2022. To Disengage or Not to Disengage: A Look at Contributor Disengagement in Open Source Software. In *44th International Conference on Software Engineering Companion (ICSE '22 Companion), May 21–29, 2022, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3510454.3522685

1 INTRODUCTION

Open source software is critical to our digital infrastructure. According to Github's Open Source Survey in 2017, nearly 94% of respondents reported using open source software in their professional work [8]. While a large number of people rely on open source, many of these projects are reliant on two or fewer core contributors. These core contributors are vital to a project's survival, and the disengagement of just one can oftentimes be disastrous [2].

Not only are people reliant on open source, much of open source is reliant on other open source projects as well. Should one project fail, many other projects may be at risk for failing in addition to the initial one. A notable example of of this cascading effect was leftpad, where one open source author deleted their project resulting in thousands of other projects being disrupted [1, 7].

Prior works looking at contributor disengagement have defined and identified disengagement by looking at contributions over time, and then built models to identify disengagement from that definition [5]. Rather than this blanketed approach of looking for x commits over a period of y time, we sought to first identify contributors that have publicly announced they have disengaged through grey literature.

Grey literature provides us the perfect opportunity for finding disengagement because it encompasses non-formal literature such as blog posts, tweets, and podcasts. This can provide us with the reasons for disengagement in the heat of the moment, and it prevents us from reopening prior wounds had we asked contributors directly.

Previous studies have not used grey literature, which presents us with a unique opportunity. This gives us an alternative direction in which to look for disengagement and may provide us with reasons different than what prior works have identified. This in-turn will better our understanding, enabling us to protect our digital infrastructure from failure by allowing us to craft more personalized and targeted solutions and prevent future problems down the road.

2 APPROACH

Before we began searching for grey literature, we first wanted to define disengagement in the context of our research. We define disengagement as either:

- Leaving a project
- Stepping down from a leadership position
- Taking a leave of absence

All of which can be attributed to a project, or to contributing completely. We began searching for grey literature on contributor disengagement with two different methods depending on the medium [3]. For the text-based, we first created a list of various terms that were related to our keywords "contributor disengagement." For example, we added words and phrases related to disengagement such as "farewell," "goodbye," and "so long" to our list.

After the compilation of our list, we began searching using combinations of the various terms. We also utilized various search engine operators as well such as "find this exact word or phrase..." We limited our results for each combination to the first 10 pages (100 results) due to the sheer volume of results. We then filtered any results that were unrelated, or duplicates. A number of our results were discussions or articles about a contributor who disengaged, and we attempted to (we limited it to a maximum of 3 different pages) navigate to the contributor's original post.

For the podcasts, we utilized the Apple Podcast API and searched for the first 100 podcasts related to open source. Afterwards, we collected all the episodes for these podcasts and searched through the episode descriptions with the above keywords. If a keyword was found, we then transcribed the audio.

ICSE '22 Companion, May 21–29, 2022, Pittsburgh, PA, USA

ACM ISBN 978-1-4503-9223-5/22/05…$15.00
https://doi.org/10.1145/3510454.3522685

Category	Code	Description	Count
Cultural Factors	Community Hostility	Disengagement due to community hostility whether it's directed specifically at them, or just an abusive community in general	13
Volunteering	Lack of Support	Disengagement due to the lack of monetary, and or physical support with the project	12
Cultural Factors	Project Direction	Disengagement due to the direction that a project is going in	11
External Factors	Not Enough Time	Where contributors would disengage because of external factors affecting the amount of time they could allocate to the project	8
Volunteering	Burnout	Disengagement because they're feeling burned out.	7

Table 1: The top five most common reasons for contributor disengagement.

After collection and filtering, we began the process of qualitative analysis. We started from scratch by cycling through each individual case of disengagement, and coded the various reasons for why they disengaged [6]. To ensure inter-rater reliability, we split the cases and distributed them equally (along with the codebook) to the other two researchers. There they coded the cases themselves, and we found that the agreement was 81.2% (percent agreement.)

After a number of cycles consisting of refining and combining, we ended up with 13 unique codes. We then began the process of axial coding. We found three overarching categories that each one of our codes belonged in:

Volunteering Related to the act of contributing in general.
Cultural Related to the culture of the project's community.
External Related to factors outside of the project.

We're continuing to analyze and code new cases, as we finish transcribing more and more audio. Additionally, we're looking to examine per-project contribution histories of these contributors to validate whether or not they actually disengaged after making one of these postings.

3 PRELIMINARY RESULTS

Our results are based on 41 cases that have been fully completed (more are still in the process of being transcribed and coded) we have identified 110 reasons for disengagement. They fell into the following higher-level categories, 50% were volunteering related, 32% cultural related, and 18% external related. Table 1 highlights the five most common reasons for contributor disengagement.

The most common reason for disengagement was community hostility, with 13 contributors citing that as one of the reasons for their disengagement. Efforts to combat this kind of hostility have been researched in the past, with some proposed solutions such as code of conducts. Unfortunately, while trying to combat hostility, these code of conducts often became the reason for the disengagement with 7 contributors citing the policies they implemented.

Lack of support was another prevalent issue, with 12 contributors citing that as the reason for their disengagement. Many of which complained about being the only person contributing to the development of the project, or the lack of financial incentive to contribute to the project.

Some of these results were also prevalent in traditional turnover studies[4] as well, such as burnout. However, burnout was typically not the sole reason contributors chose to disengage, instead it was paired with another reason such as a lack of interest, not enough time, or a lack of support.

Compared to prior works on disengagement [5], we found that most of our codes were similar. However, community hostility was nowhere to be found despite being prevalent in our analysis, which suggests that the differences in definition and identification may help us identify more forms of contributor disengagement.

4 CONTRIBUTIONS

Our findings can help researchers, contributors, and open source communities. With a bettered understanding of disengagement, more specific and targeted help can be provided to these communities to prevent contributor disengagement. For example, by understanding what kind of potential issues come into play when you're contributing, contributors can better prepare themselves for participating in open source. While communities can protect contributors by looking for ways to prevent community hostility, or policy disagreements. Researchers will have additional ways to identify and model disengagement as well, leading to an even stronger understanding of why contributors disengage.

We aim to eventually look the associations between the codes and see if there are trends between the various codes such as contributors who burnt out also noted a lack of support, etc. Eventually we hope that this research leads to better systems and or policies that protect and nurture the open source ecosystem.

It's important to note that due to the nature of how we searched for disengagement, we're only looking at contributors who publicly announced that they were disengaging. We also limited ourselves to grey literature in English.

In addition, we began working on a wiki[1] that serves to highlight all the various reasons why contributors disengage, and what some of their proposed solutions were. The idea behind this is to draw attention to the issue of contributor disengagement, and provide a place for contributors who might be on the verge of quitting to see that they're not alone, and perhaps see what steps other contributors used to remedy their issues.

5 ACKNOWLEDGEMENTS

This work was supported through Carnegie Mellon University's REU in Software Engineering (NSF Award 1852260), and the Sloan Foundation.

[1] https://disengagement-diaries.github.io/

REFERENCES

[1] Rabe Abdalkareem, Olivier Nourry, Sultan Wehaibi, Suhaib Mujahid, and Emad Shihab. 2017. Why do developers use trivial packages? an empirical case study on npm. In *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering*. ACM, Paderborn Germany, 385–395. https://doi.org/10.1145/3106237.3106267

[2] Guilherme Avelino, Leonardo Passos, Andre Hora, and Marco Tulio Valente. 2016. A Novel Approach for Estimating Truck Factors. *2016 IEEE 24th International Conference on Program Comprehension (ICPC)* (May 2016), 1–10. https://doi.org/10.1109/ICPC.2016.7503718 arXiv: 1604.06766.

[3] Vahid Garousi, Michael Felderer, and Mika V. Mäntylä. 2019. Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and Software Technology* 106 (Feb. 2019), 101–121. https://doi.org/10.1016/j.infsof.2018.09.006

[4] Dirk Homscheid and Mario Schaarschmidt. 2016. Between Organization and Community: Investigating Turnover Intention Factors of Firm-Sponsored Open Source Software Developers. In *Proceedings of the 8th ACM Conference on Web Science* (Hannover, Germany) *(WebSci '16)*. Association for Computing Machinery, New York, NY, USA, 336–337. https://doi.org/10.1145/2908131.2908200

[5] Courtney Miller, David Gray Widder, Christian Kästner, and Bogdan Vasilescu. 2019. Why Do People Give Up FLOSSing? A Study of Contributor Disengagement in Open Source. In *Open Source Systems*, Francis Bordeleau, Alberto Sillitti, Paulo Meirelles, and Valentina Lenarduzzi (Eds.). Vol. 556. Springer International Publishing, Cham, 116–129. https://doi.org/10.1007/978-3-030-20883-7_11 Series Title: IFIP Advances in Information and Communication Technology.

[6] Johnny Saldaña, Matthew Miles, and Michael Huberman. 2019. *Qualitative Data Analysis: A Methods Sourcebook*. Sage publications.

[7] Marat Valiev, Bogdan Vasilescu, and James Herbsleb. 2018. Ecosystem-level determinants of sustained activity in open-source projects: a case study of the PyPI ecosystem. In *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. ACM, Lake Buena Vista FL USA, 644–655. https://doi.org/10.1145/3236024.3236062

[8] Frances Zlotnick. 2017. Open Source Survey. https://opensourcesurvey.org/2017/