Choosing an Appropriate Task to Start With in Open Source Software Communities: a Hard Task

Igor Steinmacher, Marco Aurélio Gerosa

Departamento de Computação – Federal University of Tecnology-Paraná (UTFPR)
Instituto de Matemática e Estatística – University of São Paulo (USP)
igorfs@utfpr.edu.br, gerosa@ime.usp.br

Abstract. Open Source Software (OSS) projects leverage the contribution of outsiders. Usually these communities do not coordinate the work of the newcomers, who go to the issue trackers and self-select a task to start with. We found that "finding a way to start" is recurrently reported both by the literature and by practitioners as a barrier to onboard to an OSS project. We conducted a qualitative analysis with data obtained from semi-structured interviews with 36 subjects from 14 different projects. We used procedures of Grounded Theory – open and axial coding – to analyze the data. We found that newcomers are not enough confident to choose their initial task and they need information about the tasks or direction from the community.

Keywords: Open Source Software, coordination, task selection, newcomers, onboarding, new developer, joining process

1 Introduction

Open Source Software projects rely on geographically distributed developers working as a team and incorporating their individual creations into a single, seamless body of source code [13]. Many OSS projects leverage contributions from volunteers and require a continuous influx of newcomers for their survival, long-term success, and continuity. According to Qureshi and Fang [10], it is essential to motivate, engage, and retain new developers to promote a sustainable number of developers in a project.

However, newcomers usually face many difficulties to make their first contribution to an open source project. Therefore, a major challenge for OSS projects is to provide ways to support newcomers' joining. Understanding developer motivation and project attractiveness are well-explored topics in the literature [7, 12, 21]. However, little is known about the barriers that newcomers face when onboarding to a project [20].

In a previous research, we identified 57 barriers faced by newcomers when onboarding to OSS projects from interviews with OSS practitioners. Among these barriers, a barrier called "*difficulty to find a task to start with*" called our attention. In addition, we also evidenced this barrier in other studies conducted [14, 16].

"Difficulty to find a task to start with" is a problem inherently related to the coordination of OSS projects. Members of OSS projects usually coordinate their tasks by using issue trackers such as bugzilla, Jira, Rapid Miner, etc. These issue trackers are sys-

tems where any user or developer is free to report, discuss, and choose a task to work on. When newcomers want to contribute, they usually access the issue tracker to choose a bug or a feature request that they can (or want to) handle. However, as reported by a newcomer "*it is a bit frustrating trying to find something I could do or fix.*"

Thus, in this paper, we focus on this barrier, aiming at better understanding it. To achieve this, we analyzed data from interviews conducted with 36 OSS practitioners from 14 different projects. We started our investigation aiming at answering a broad question ("What are the barriers that hinder newcomers' onboarding to OSS projects?") and, as the barrier called our attention, we focused on it by analyzing the existing data and conducting another round of interviews. We did a qualitative analysis based on procedures of grounded theory. The analysis resulted in a model containing 14 concepts that help explaining the barrier.

2 Related Work

Newcomers' onboarding is an issue faced in many online communities. Many studies in the literature deal with newcomers joining process in collective production communities, including studies on online communities [1, 4] and on OSS projects [3, 6, 15, 19].

Von Krogh et al. [6] analyzed interviews with developers, emails, source code repository, and documents and proposed a joining script for developers who want to take part in the project. Nakakoji et al. [8] studied four OSS projects to analyze the evolution of their communities. They presented eight possible roles for the community members and structured them into a model composed of concentric layers, like the layers of an onion, later called the onion patch.

Some researchers tried to understand the barriers that influence the retention of newcomers. Zhou and Mockus [22] worked on identifying the newcomers who are more likely to remain in the project in order to offer active support for them to become long-term contributors. Jensen et al. [5] analyzed if the emails sent by newcomers are quickly answered, if gender and nationality influence the kind of answer received, and if the reception of newcomers is different in users and developers lists. Steinmacher et al. [15] studied how reception influences the retention of newcomers in an OSS project. Park and Jensen [9] show that visualization tools support the first steps of newcomers in an OSS project, helping them to find information more quickly.

Finding the appropriate task is usually classified as a problem because new developers have difficulty to find bugs or features that are of interest, that match their skill sets, are not duplicates, and are important for their future community [19]. Von Krogh et al. [6] found that members of the community encouraged the new participants to find their first tasks themselves. Park and Jensen [9] reported that "*... subjects expressed a need for information specific to newcomers, for instance,... what to contribute to…*"

Regarding support to deal with tasks, Čubranić et al. [2] presented a tool that recommend source code, emails messages, and bug reports to support newcomers. Wang and Sarma [19] presented a tool to enable newcomers to identify similar bugs through synonym-based search. These tools can help newcomers by increasing their knowledge about the tasks and their complexity.

From the literature analysis, we could observe that, from the communities perspective, newcomers should be able to find the most appropriate task themselves, as reported by von Krogh [6]. However, other researches argue that the projects should give special

attention to this issue and support the newcomers finding their tasks. The goal of this work is to understand this problem from both perspectives, to enable researchers and community to focus on creating strategies to support newcomers to OSS projects.

3 Research Method

We conducted a qualitative research to understand the difficulties to find an appropriated task and proposed a set of strategies to alleviate the problems caused by this barrier. Qualitative research produces results that cannot be achieved through statistical procedures or similar methods [17]. The results of this kind of approach are richer and more informative, helping to answer questions involving variables that are difficult to quantify, such as human characteristics like motivation and perception [11].

We used semi-structured interviews as data collection method. The participants were recruited primarily through mailing list and forums from 15 different projects and from weblogs postings. We also invited newcomers and project owners directly, identifying them in project pages and by mining and following projects' mailing lists and issue trackers.

We interviewed 36 OSS developers from 14 different projects, including 11 experienced members, 16 newcomers that succeeded, 6 dropout newcomers, and 3 newcomers that were still trying to place their first contributions. Some information about their profile is presented in Table 1. The interviews were conducted using textual based chat tools, like Google Talk. We chose this mean once the participants are used to this kind of tool for their professional and personal activities. Each interview was conducted individually and the data was logged in a local computer. Interviews began with a short and general explanation of the research, followed by some questions to profile the interviewees regarding their technical experience and main occupation.

The first step of the study consisted in a first round of interviews (*It1* in the last column of Table 1) to answer a broad question ("What are the barriers that hinder newcomers' onboarding to OSS projects?"). The interviews of this step focused on identifying the barriers from newcomers and experienced members' perspective. During the analysis of these interviews, we needed to clarify some doubts to better understand some information and conducted few other interviews with some participants (represented as *It2* in the last column of Table 1).

We qualitatively analyzed the data using procedures of Grounded Theory (GT) [17], which is based in three coding steps: open coding, when concepts are identified and their properties and dimensions are discovered in the data; axial coding, when connections between the codes are identified and grouped according to their properties to represent categories; and selective coding, when the core category is identified and described. Although the purpose of the GT method is the construction of substantive theories, its use does not necessarily restricted only to researches with this goal. According to Strauss and Corbin [17], the researcher may use only some of its procedures to meet one's research goals.

During open coding, we assigned codes to sentences, paragraphs, or revisions. This procedure overlapped the axial coding, in which connections between the categories were identified. In practice, open and axial coding were executed several times to refine the emerging codes and categories.

Table 1. Data Collection Summary

	Time spent in OSS per week	*First Project?*	*Profile*	*Country*	*Years in the project*	*Data Collection*
P1	less than 5 hours	N	member	France	8	I1
P2	from 5 to 10 hours	Y	member	Germany	3	I1
P3	from 10 to 20 hours	N	member	Germany	3	I1, I2
P4	from 5 to 10 hours	N	member	Canada	10	I1
P5	from 5 to 10 hours	N	member	Germany	15	I1, I3
P6	more than 20 hours	N	member	Hungary	10	I1, I2
P7	more than 20 hours	N	member	Australia	5	I1
P8	more than 20 hours	N	member	Brazil	5	I1
P9	more than 20 hours	N	member	Turkey	8	I1, I3
P10	from 5 to 10 hours	N	member	Brazil	15	I1
P11	less than 5 hours	N	member	Brazil	7	I1
P12	less than 5 hours	Y	newcomer	Germany	0	I1, I3
P13	less than 5 hours	Y	newcomer	Brazil	0	I1
P14	from 5 to 10 hours	Y	newcomer	India	1	I1
P15	from 5 to 10 hours	Y	newcomer	India	0	I1
P16	less than 5 hours	Y	newcomer	Germany	0	I1
P17	less than 5 hours	N	newcomer	USA	0	I1, I2
P18	less than 5 hours	Y	newcomer	USA	0	I1
P19	more than 20 hours	Y	newcomer	Greece	0	I1
P20	less than 5 hours	Y	newcomer	Brazil	0	I1
P21	less than 5 hours	Y	newcomer	Brazil	0	I1, I2
P22	less than 5 hours	Y	newcomer	Brazil	0	I1
P23	N/I	N	newcomer	UK	-	I1
P24	from 10 to 20 hours	N	newcomer	Brazil	1	I1, I2, I3
P25	from 5 to 10 hours	Y	newcomer	Brazil	1	I1
P26	N/I	Y	newcomer	France	0	I1
P27	from 5 to 10 hours	N	newcomer	Germany	0	I1
P28	from 5 to 10 hours	N	dropout	USA	0	I1
P29	less than 5 hours	Y	dropout	India	0	I1
P30	less than 5 hours	N	dropout	Germany	0	I1
P31	less than 5 hours	Y	dropout	Brazil	0	I1
P32	less than 5 hours	Y	dropout	India	0	I1
P33	less than 5 hours	Y	dropout	India	0	I1
P34	less than 5 hours	N	onboarding	China	0	I1
P35	from 10 to 20 hours	Y	onboarding	India	0	I1, I2
P36	less than 5 hours	Y	onboarding	Greece	0	I1, I3

During the analysis of the first set of interviews, we were able to identify 57 different barriers, grouped in six categories, some of them split in subcategories. As our work progressed, we decided to focus on a recurrent and specific category of barriers highlighted in the interviews: *find a way to start*. More specifically, we were interested in better understanding the barrier called *difficulty finding a task to start with*, once the barrier was recurrently evidenced, appearing in 9 interviews. To investigate this specific barrier, we conducted an in-depth analysis of the existing data, and conducted another iteration of interviews with five participants (*It3* in the last column of Table 1). The results of this analysis are presented in the following section.

4 Difficulty to Find a Task to Start With

Our analysis resulted in the emergence of concepts that enabled us to better understand this barrier. These concepts represent a more detailed view on the reasons why newcomers understand the choosing of an appropriate task as a barrier to onboard to the project. The resulting model comprising the concepts and relationships among them is presented as a network in Fig. 1.

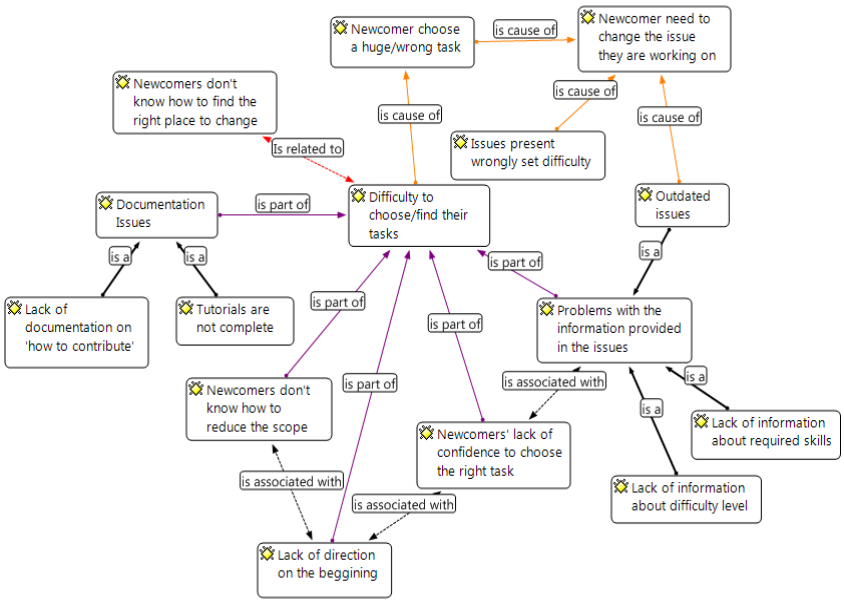

Fig. 1. Result of the qualitative analysis represented as a network

The phenomenon we were trying to understand in this case was the "Difficulty to choose/find their tasks." Five issues emerged as part of the barrier under investigation, and three problems that are consequences of this barrier.

Some developers reported the **lack of direction at the beginning** as part of the explanation why it is difficult to find a task to work on. We found that some newcomers expect that someone indicate them what are the tasks they should start with. To explain this, one of them reported: *"We feel more secure when someone is guiding us. I think this is related to the experience with large projects and Open Source… Less experienced developers always stand on the back foot."*

As it is possible to observe in the previous quote, this issue is related to the **lack of confidence** that newcomers have when choosing a task. The newcomer said, "*I really didn't know which one to pick, which one I was 'authorized' to pick... whether they are important*." It shows he was not sure about what he could do, what would be enough and important. Another newcomer told us a case of a close friend that tried to join the project: "*A college colleague gave up… He did not find a task that he felt confidence to try.*" This occurs because newcomers, even when they are aware they can choose a task, they are not sure about what are easy or not, and what they can touch: "*it is frustrating, there is a bunch of issues, but I don't have the proper knowledge to judge what exactly I can touch…*"

On the other side, experienced members reported that **newcomers do not know how to reduce the scope**. However, none of them put it as a newcomer's fault. One said: "*sometimes they want to contribute but don't know how to reduce the scope when starting*." Another practitioner [P6] reported: *"the task chosen [by the newcomer] makes some sense, but is huge, and the newbie thinks she'll be able to implement it in a few days."* The same participant reported the association among the inability of reducing the

scope and lack of direction when asked if he had seen some case in which a newcomer gave up without finishing assignment task: "*When they choose a task that is reasonably sized, just we don't give the necessary help.*"

Some interviewees also mentioned **documentation issues** as part of the difficulty to find a task to work on. We could identify two related issues from the interviews: **lack of documentation on how to contribute** and **incomplete/outdated tutorials.** These issues were mainly reported by newcomers that recently joined a project. One of them summarized the first one when he said that "*there is no good guide for starter*" [P34]. Two other newcomers, who were onboarding to the project when we conducted the interviews, reported a hard time with the incomplete tutorials. One of them reported: *"A proper up-to-date guide with tutorials would have really helped. As in when I began, I was totally new to open source, but the tutorials skipped over certain instructions and I had no idea what I was doing wrong." [P33]*

Problems with the information provided in the issues were recurrently reported by newcomers and experienced members. We could clearly distinguish three types of issues: **outdated issues**; **lack of information about required skills**; and **lack of information about difficulty level.** An experienced member [P9] reported that the lack of information frustrated him when he was trying to contribute to a project: *"the issues that I can contribute were not clearly defined so it was my job as a newcomer to find out how to contribute there were no easy or junior bugs as in some projects."* However, defining a task or issue as good for newcomers is sometimes not enough. A newcomer also evidenced that the skills needed to accomplish that task need to be clear: *"the issues should indicate the area of knowledge, like C++, build, shell script" [P12].*

In addition to the issues that helped us understanding the difficulties newcomers face to find an appropriate task, we also evidenced other issues caused by this barrier. We found that, due to these difficulties, **newcomers choose a wrong/huge task**, as per this quote from an experienced member: "*most of them [newcomers] do not know how to start... what can they do first and can choose the wrong task*" [P20]. A newcomer that is contributing for one year to an OSS project reported this problem and accredited the problem to a mistake when the issue was classified as easy: *"Sometimes, I tried to work on a task that was classified as easy hacks and it was too complex that only experienced members could find the solution… the developers who registered these easy hacks, sometimes made mistakes when classifying the difficulty" [P24]*

Another newcomer reported his difficulties finding a task. The coordination mechanism used by the project failed to support his choice. He reported that he had to change tasks twice, but for different reasons: *"when I finally did have something to do, it basically completely changed two times … one task was too hard and the other was a feature that was already implemented, but the task was not updated" [P19]*

When searching for an issue to work on, newcomers want to be aware of some details about the issues; mainly they want to know whether they will be able or not to handle the task. This links this kind of problem with the **lack of confidence to choose a task**. The way they do this is looking at the issue tracker, aiming at coordinating their contribution with the community. Therefore, newcomers need up-to-date information to support their decision, to make it easier and clearer for them to be aware of what they can do, what they are able to do, and where they can look for support. The main point now is to identify potential strategies to improve this coordination mechanism in a way the newcomers can feel more comfortable in choosing their first task.

5 Threats to Validity

We are aware that each project has its singularities, that the OSS universe is huge, so, the level of support and the barriers can differ according to the project or the ecosystem. Our strategy to consider different projects and different profiles of developers aimed to mitigate this limitation, identifying recurrent barriers from multiple perspectives.

There is also a threat related to the sampling method of the interviews. We sent out invitations to specific development lists and directly to newcomers and project core members, what could introduce some bias.

Another threat to the validity of the results is the subjectivity of the data classification. We used Grounded Theory procedures to mitigate this threat, given that GT requires the entire analysis to be grounded on the data collected. Additionally, the analysis process was discussed along with two other researchers, to encourage a better validation of the interpretations through mutual agreement.

6 Conclusions

Although this study focused on OSS communities, better supporting newcomers is an important issue in many other communities. Many virtual communities count on volunteer contributions. These volunteers can easily leave, since they have no formal relationship with these communities [1]. Moreover, the impoverished awareness information, lack of trust, and the relatively weak interpersonal ties between members in many online groups make it more problematic to attract and retain people than in face-to-face groups [18]. Therefore, studying newcomers and the problems faced by them in virtual communities is a contemporary problem that still needs to be further investigated.

In this paper, we qualitatively analyzed data collected from newcomers, dropouts, and members of OSS projects, aiming at understanding the barriers faced by newcomers to OSS projects. The results of the analysis helped us understanding the "difficulty to find a task to start with" faced by newcomers. It was possible to evidence some of the actual issues that explain this barrier. Lack of confidence to choose a task appeared as a relevant concept in our analysis. Newcomers need the project to provide enough information about the tasks to support their decision about which task is more suitable for them.

The next step of this research encompasses finding suitable solutions and conducting experiments to assess the effectiveness of them in supporting newcomers onboarding to OSS projects.

7 References

[1] Choi, B., Alexander, K., Kraut, R.E. and Levine, J.M. 2010. Socialization Tactics in Wikipedia and Their Effects. Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (2010), 107–116.

[2] Cubranic, D., Murphy, G.C., Singer, J. and Booth, K.S. 2005. Hipikat: a project memory for software development. IEEE Transactions on Software Engineering. 31, 6 (2005), 446–465.

[3] Ducheneaut, N. 2005. Socialization in an Open Source Software Community: A Socio-Technical Analysis. Computer Supported Cooperative Work (CSCW). 14 4 (2005), 323-368.

[4] Halfaker, A., Geiger, R.S., Morgan, J. and Riedl, J. 2013. The Rise and Decline of an Open Collaboration System: How Wikipedia's reaction to sudden popularity is causing its decline. American Behavioral Scientist. 57, (2013).
[5] Jensen, C., King, S. and Kuechler, V. 2011. Joining Free/Open Source Software Communities: An Analysis of Newbies' First Interactions on Project Mailing Lists. 44th Hawaii Intl. Conf. on System Sciences (2011), 1–10.
[6] Von Krogh, G., Spaeth, S. and Lakhani, K.R. 2003. Community, joining, and specialization in open source software innovation: a case study. Research Policy. 32, 7 (2003), 1217-1241.
[7] Meirelles, P., Santos, C., Miranda, J., Kon, F., Terceiro, A. and Chavez, C. 2010. A study of the relationships between source code metrics and attractiveness in free software projects. 2010 Brazilian Symposium on Software Engineering (SBES), (2010), 11–20.
[8] Nakakoji, K., Yamamoto, Y., Nishinaka, Y., Kishida, K. and Ye, Y. 2002. Evolution Patterns of Open-source Software Systems and Communities. Proceedings of the International Workshop on Principles of Software Evolution (2002), 76–85.
[9] Park, Y. and Jensen, C. 2009. Beyond pretty pictures: Examining the benefits of code visualization for Open Source newcomers. 5th Intl. Workshop on Visualizing Software for Understanding and Analysis (2009), 3–10.
[10] Qureshi, I. and Fang, Y. 2011. Socialization in Open Source Software Projects: A Growth Mixture Modeling Approach. Org. Res. Methods. 14, 1 (2011), 208–238.
[11] Seaman, C.B. 1999. Qualitative methods in empirical studies of software engineering. Software Engineering, IEEE Transactions on. 25, 4 (Jul. 1999), 557–572.
[12] Shah, S.K. 2006. Motivation, Governance, and the Viability of Hybrid Forms in Open Source Software Development. Manage. Sci. 52, 7 (2006), 1000–1014.
[13] Singh, P.V. 2010. The Small-world Effect: The Influence of Macro-level Properties of Developer Collaboration Networks on Open-source Project Success. ACM Trans. Softw. Eng. Methodol. 20, 2 (Sep. 2010), 6:1–6:27.
[14] Steinmacher, I., Silva, M.A.G. and Gerosa, M.A. 2014. Systematic review on problems faced by newcomers to open source projects. 10th International Conference on Open Source Software (2014), 10pp.
[15] Steinmacher, I., Wiese, I., Chaves, A.P. and Gerosa, M.A. 2013. Why do newcomers abandon open source software projects? International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE) (2013), 25–32.
[16] Steinmacher, I., Wiese, I.S., Conte, T., Gerosa, M.A. and Redmiles, D. 2014. The Hard Life of Open Source Software Project Newcomers. International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE 2014) (2014).
[17] Strauss, A. and Corbin, J.M. 1998. Basics of Qualitative Research : Techniques and Procedures for Developing Grounded Theory. SAGE Publications.
[18] Tidwell, L.C. and Walther, J.B. 2002. Computer-Mediated Communication Effects on Disclosure, Impressions, and Interpersonal Evaluations: Getting to Know One Another a Bit at a Time. Human Communication Research. 28, 3 (2002), 317–348.
[19] Wang, J. and Sarma, A. 2011. Which bug should I fix: helping new developers onboard a new project. Proceedings of the 4th International Workshop on Cooperative and Human Aspects of Software Engineering (Waikiki, Honolulu, HI, USA, 2011), 76–79.
[20] Wolff-Marting, V., Hannebauer, C. and Gruhn, V. 2013. Patterns for tearing down contribution barriers to FLOSS projects. 12th Intl. Conf. on Intelligent Software Methodologies, Tools and Techniques (2013), 9–14.
[21] Ye, Y. and Kishida, K. 2003. Toward an Understanding of the Motivation Open Source Software Developers. Proceedings of the 25th International Conference on Software Engineering (Portland, Oregon, 2003), 419–429.
[22] Zhou, M. and Mockus, A. 2012. What make long term contributors: Willingness and opportunity in OSS community. Software Engineering (ICSE), 2012 34th International Conference on (Jun. 2012), 518–528.