Poster: Understanding Newcomers Success in Open Source Community

Shahab Bayati
ISOM Department, Business School, The University of Auckland
Auckland, New Zealand
s.bayati@auckland.ac.nz

ABSTRACT

Newcomers and volunteers contributions play an effective role the open source software (OSS) success. This role is confirmed through a rigor set of studies in software engineering discipline. As Open source projects are developed based on social and technical efforts, then it is very important for newcomers to empower their socio-technical skills. This paper focuses on newcomers' success in open source community by analyzing newcomers' reputation on their initial activities in a social coding environment such as GitHub. By applying mining software repositories (MSR) techniques on GitHub data we found the main projects' attributes where successful newcomers contributed to them. These attributes can help other newcomers to select the right project for their initial activities.

CCS Concepts

. Information systems → Open source software
. Software and its engineering → Collaboration in software development

KEYWORDS

Newcomers, Open Source, Reputation Analysis, Social Coding

1 INTRODUCTION

Open source software development relies on volunteers' commitment to the projects [1]. To have a successful and active project they need to have more contributors ranging from newcomers to rock-stars. However, many of these newcomers lost their motivations and stopped their jobs in OSS projects regarding various socio-technical barriers [2]. In this study, we explored the newcomers' activities and their reputation in OSS community as a success factor to find what makes a newcomer successful. We mainly focused on the type of project they select in their initial phases.

The main motivational factors of open source contributors are discussed in the literature [3]. These motivations are categorized as intrinsic, extrinsic and ideological factors. However, many of OSS newcomers are not successful in their joining process into OSS community and they lost their motivations. These issues are investigated in the OSS literature.

As newcomers are important for project survival, some projects provide mentorship programmes for newcomers [4]. Recommender systems help newcomers in finding an appropriate task to start [5], a right mentor to get initiative guides, [4] and documents and code to get familiar with project architecture [6]. But the main concept which is not studied well is what the successful newcomer has done during their joining process. This study contributes to OSS literature by focusing on newcomers' success factors by comparing their initial activities and project selection.

2 RELATED WORKS

The effective role of newcomers and their main issues and challenges in OSS community is studied in the literature with various perspectives [1, 2, 7]. The main newcomers' barriers in OSS is gathered in a comprehensive literature review [7]. Some technical studies designed artefacts to help a newcomer to find a task, source code or mentors to have a successful joining process [8]. Newcomers performance is related to socialization experiences they got in mentorship programmes and with community members [9]. The importance of initial behaviour in the success of newcomers is investigated in [10]. Newcomers' popularity in OSS community is investigated in this study to support them to reach to high status by selecting right projects.

3 DATA COLLECTION

In our analysis, we used GhTorrent [11] public archive of GitHub. To collect newcomers related data we have counted the number of followers they gained in a year. GitHub developer will follow each other when they want to be aware of their activity [12]. Based on social capital theory, gaining status in online community is one the major motives for contributors. We used following as a reputation factor. A developer with more followers considered as more successful and popular one in this study.

We categorized newcomers in four different classes according to the gained followers. We summarized their activities in all the projects they have contributed to them. In our analysis we have listed all the projects these newcomers have any type of collaboration on them including technical (Code Commits, Pull Requests, Issue reporting and issue resolution) or social which consist of any conversation, discussion, and comments they made on others technical events. Then we analysed these projects socio-technical activities and compared them to different categories.

ICSE '18 Companion, May 27-June 3, 2018, Gothenburg, Sweden

ACM ISBN 978-1-4503-5663-3/18/05.
https://doi.org/10.1145/3183440.3195073

We have sampled the list of newcomers on different categories. Our sample contains 473 newcomers which are categorized as 61 from Super Reputable (SR) category, 93 from High Reputable (HR), 133 from Low Reputable (LR) and 186 from Non-Reputable (NR). Results show that SR newcomers contributed in more projects than others. Also, this group follows more GitHub members than others. SR has more followees. The interesting point is that SR has less commitment than other groups which is aligned with the result of [12] which depicted commits are not an effective factor in developer popularity. Also, it can be understood that newcomers in SR and HR are more interested in coordination activities.

4 DATA ANALYSIS

We have captured a list of project characteristics that newcomers contributed to them for the very first of their joining period in GitHub and for each selected project we captured 3-year data. List of attributes abbreviation and description is illustrated in table 1. We have applied Random Forest to find the project feature importance. Random Forest provides mean decrease accuracy value which represents the decrease in model accuracy in the absence of the feature.

Table 1: Project attributes

IMP	Attributes	Description
1	MemCommitters	Number of project committers who are the member
2	Age	Days passed after creation date
3	Members	Number of project members
4	Committers	Number of project committers
5	Commits	Number of commits applied to the project
6	Lang	Programming language category
7	Watchers	Number of project watcher
8	Issues	Number of reported issues
9	Owner	Owner Type
10	PullReq	Number of submitted Pull Requests
11	Forks	Number of times project forked
12	IssueCmnt	Number of comments on issues
13	PRIssueCmnt	Number of comments on issues related to Pull Requests
14	CommitCmnt	Number of comments on project commits
15	PRCmnt	Number of comments on Pull Requests

5 DISCUSSION AND CONCLUSION

This study focused on the OSS newcomers and investigates their success way by mining historical data of GitHub newcomers. The analysis shows that to be successful newcomers in terms of popularity, it is not just a technical commitment that can help a newcomer, however, they need to be more social. They need to participate in reviewing process and follow other members. It is recommended to contribute just not in one project. More projects they contribute more members they know and more socio-technical skills they gain. Also, we have mined the projects newcomers contributed to them and found that items like project age and number of committers from project core members can help newcomers in selecting a project. Popular newcomers joined to the younger set of projects.

To the best of our knowledge, this is the first study looks for newcomers' contribution in GitHub to find out selecting which project types can make them popular in OSS. However, similar to the most of the studies, it suffers from some limitations. Focusing on the limited number of newcomers and projects is the main issue of this study although we have categorized our data based on a large sample distribution, it can be improved by more data from different platforms. We have counted all the commits in with same value, however in the real world commits are different. Same for following. Having a reputable follower is different from the general member but we consider all the contributors with same influence value. Longitudinal data of developer activities over time can give a better understanding rather the case of cross-sectional data. In addition, this study only focused on the followers' count as a factor of success which is biased on the social side of social coding environment, however, other metrics such as expertise can be considered as developer success measures.

REFERENCES

[1] G. Von Krogh, S. Spaeth, and K. R. Lakhani, "Community, joining, and specialization in open source software innovation: a case study," Research Policy, vol. 32, pp. 1217-1241, 2003.

[2] I. Steinmacher, I. S. Wiese, T. Conte, M. A. Gerosa, and D. Redmiles, "The hard life of open source software project newcomers," in Proceedings of the 7th international workshop on cooperative and human aspects of software engineering, 2014, pp. 72-78.

[3] G. Von Krogh and E. Von Hippel, "The promise of research on open source software," Management science, vol. 52, pp. 975-983, 2006.

[4] I. Steinmacher, I. S. Wiese, and M. A. Gerosa, "Recommending mentors to software project newcomers," in Proceedings of the Third International Workshop on Recommendation Systems for Software Engineering, 2012, pp. 63-67.

[5] J. Wang and A. Sarma, "Which bug should I fix: helping new developers onboard a new project," in Proceedings of the 4th International Workshop on Cooperative and Human Aspects of Software Engineering, 2011, pp. 76-79.

[6] I. Steinmacher and M. A. Gerosa, "How to support newcomers onboarding to open source software projects," in IFIP International Conference on Open Source Systems, 2014, pp. 199-201.

[7] I. Steinmacher, M. A. G. Silva, M. A. Gerosa, and D. F. Redmiles, "A systematic literature review on the barriers faced by newcomers to open source software projects," Information and Software Technology, vol. 59, pp. 67-85, 2015.

[8] Y. Malheiros, A. Moraes, C. Trindade, and S. Meira, "A source code recommender system to support newcomers," in Computer Software and Applications Conference (COMPSAC), 2012 IEEE 36th Annual, 2012, pp. 19-24.

[9] K. Carillo, S. Huff, and B. Chawner, "What makes a good contributor? Understanding contributor behavior within large Free/Open Source Software projects–A socialization perspective," The Journal of Strategic Information Systems, 2017.

[10] M. Zhou and A. Mockus, "Who will stay in the floss community? modeling participant's initial behavior," IEEE Transactions on Software Engineering, vol. 41, pp. 82-99, 2015.

[11] G. Gousios, "The GHTorent dataset and tool suite," in Proceedings of the 10th Working Conference on Mining Software Repositories, 2013, pp. 233-236.

[12] K. Blincoe, J. Sheoran, S. Goggins, E. Petakovic, and D. Damian, "Understanding the popular users: Following, affiliation influence and leadership on GitHub," Information and Software Technology, vol. 70, pp. 30-39, 2016.