6-15-2019

Role of Newcomers Supportive Strategies on Socio-Technical Performance of Open Source Projects

Shahab Bayati
Univeristy of Auckland, s.bayati@auckland.ac.nz

K Dharini Amitha Peiris
University of Auckland, New Zealand, a.peiris@auckland.ac.nz

Role of Newcomers Supportive Strategies on Socio-Technical Performance of Open Source Projects

Completed Research Paper

Shahab Bayati
University of Auckland
New Zealand
s.bayati@auckland.ac.nz

K. Dharini Amitha Peiris
University of Auckland
New Zealand
a.peiris@auckland.ac.nz

Abstract

The success of open source software (OSS) projects have been studied in previous research. This paper focused on the effect of newcomers' supportive strategies in OSS projects on the success level of the projects. Our research analyzes the socio-technical commitment to the project as a proxy for success. Data about 453 OSS projects from GitHub.com is collected and analyzed to empirically test the research model. We have applied a clustering technique to explore the dataset attributes. Results show the importance of newcomers' supportive strategies on the different socio-technical aspects of OSS projects' leading to success. Also, we have tested the effect of programming language diversity and project profile health on the success of projects. The outcome of this study has both managerial and practical implications.

Keywords: Open Source, Social Coding, Newcomers, Socio-Technical Success, Programming Language Diversity, Clustering

INTRODUCTION

Rapid growth and popularity of open source software (OSS) lead to the high attention of software engineering practitioners and information systems researchers. Studies in OSS are established in different aspects of OSS projects and community. A big set of studies focused on developers motivations in joining and committing to OSS projects (Hertel et al. 2003; Roberts et al. 2006; Von Krogh et al. 2003). Investigating OSS community as a network of developers and projects with social network theory and social capital theory is another set of studies (Grewal et al. 2006; Singh 2010). Some researchers studied the success antecedents of open source projects (Crowston et al. 2006; Midha and Palvia 2012). Newcomers' adaption and their challenges in OSS community is also a concern of few researchers in OSS studies (Bayati and Peiris 2018; Steinmacher and Gerosa 2014; Steinmacher et al. 2015). This study is contributed to the last two streams of OSS literature by analyzing the success of OSS projects based on their strategies for supporting newcomers.

The success of OSS projects are analyzed through different perspectives (Crowston et al. 2006; Ghapanchi et al. 2011; Grewal et al. 2006). The level of coding activities in a project is widely used as a measure of success in OSS literature (Ghapanchi et al. 2011; Grewal et al. 2006). However, as it is suggested by (Dabbish et al. 2012; Tsay et al. 2014) open source development is not just about the technical activities and software engineering is a socio-technical process. Then we have considered the level of social behaviour and discussion as another measure of success in OSS. Socio-Technical congruence theory depicts that a successful project needed to be aligned socially and technically. It is mentioned in the literature that developers are evaluating projects in a social coding environment with

both social and technical events surrounding a project community (Tsay et al. 2014). In OSS technical success literature some factors such as programming language, organizational sponsorship, number of releases, and response rate are considered as effective factors (Stewart et al. 2006; Subramaniam et al. 2009). We have argued for the effect of these factors on the social and behavioural success of OSS projects

Volunteers and newcomers role in OSS success is vital (Schilling et al. 2012). In this study, we have hypothesised the antecedents of newcomers' supportive strategies in the success of open source projects. Literature has shown newcomers face a challenging and tough pathway to be an active and reputable member in OSS community (Bayati and Peiris 2018; Steinmacher and Gerosa 2014). Stainmatcher et al. have categorized the main newcomers' barriers in OSS (Steinmacher et al. 2015). This study contributed to the newcomers' commitment to OSS community by testing the role of project owners' strategies in regarding supporting newcomers. Then the main research question of this study is "What is the effect of newcomers' supportive strategies in OSS project socio-technical success?"

For the purpose of this study, we have used a dataset of open source projects hosted on GitHub as the largest repository of OSS projects. GitHub is a social coding platform which provides social networking facilities for open source contributors among other coding common practices such as CVS and issue trackers. GitHub provides some regular statistics and indicators about project such as health and popularity. Currently, they have added and community profile health indicator which represents the availability of templates, guidelines and licences for each project. The health of OSS projects is evaluated with different metrics and techniques in the literature (Jansen 2014; Piggot and Amrit 2013). We have argued about the effect of documentation healthiness of projects effects on their success level.

This paper is structured as follows. In the next section, the literature of OSS success factors and newcomers participation in OSS development is reviewed. Data collection and cleaning process is discussed later. Next, the exploratory data analysis is presented. Furthermore, data analysis and results are discussed.

RELATED WORKS

To define this study contribution and scope, this section reviews the related literature to open source project success and the role of newcomers in open source community.

Open Source Success

The success of OSS projects leads many researchers to investigate the antecedents of open source success. Some studies mapped the information system success (Delone and McLean 2003) model to open source projects (Crowston et al. 2006). The market success of open source projects are compared with technical success in several studies (Grewal et al. 2006; Midha and Palvia 2012; Subramaniam et al. 2009). The number of downloads and code contributions are widely used for these analyses. Project sponsorship, age, codebase size, members, licence type, programming language , audience, and code quality are found as effective factors to OSS success (Ghapanchi et al. 2011; Grewal et al. 2006; Midha and Palvia 2012; Stewart et al. 2006; Subramaniam et al. 2009). Most of these studies tested their hypotheses on the data collected from SourcForge.net however, GitHub as a social coding platform provides more data.

Social network theory is also applied to the studies related to success. Most of these studies formed a network of developers or projects based on their affiliation network. Role of project leaders and developers connections on projects success is approved in (Hahn et al. 2008; Singh et al. 2011). Effect of project embeddedness is studied in (Grewal et al. 2006). All of these studies focused on the importance of social ties on OSS success.In contrast with the body of literature, This study combines both social and technical activities success factors in open source projects as these factors are affecting developers and project managers' decision making process (Dabbish et al. 2012; Tsay et al. 2014).

Newcomers in Open Source Community

Newcomers in OSS projects play a vital role in project survival, however, the turnover and attrition among newcomers are very high in OSS community (Schilling 2012; Von Krogh et al. 2003). Newcomers facing many challenges in their initiative activities in OSS projects which discourage them

from sustained participation (Steinmacher et al. 2015). This forced open source leaders to provide supportive programs such as mentorship for newcomers (Canfora et al. 2012). Organizations such as OpenHatch support newcomers in their interaction with OSS community to facilitate their progress. The specific role of newcomers in open source project success is not mentioned in the OSS literature. We argued that a successful project requires new members to join. A successful project has a strategy for supporting newcomers and retain them. The role of organization (Project) strategies in newcomers' satisfaction and retention is studied in online community literature (Bauer and Erdogan 2011). Drawing on these studies we have extended this hypothesis to the open source projects.

RESERCH MODEL

The research model of this study is presented in figure 1. It shows that we are testing the effect of newcomers' supportive strategies, GitHub's community profile health, project responses to the community requests, and programming language diversity on the project success. In terms of success evaluation, following the literature, we have used coding activities and developers' interest. Also, we have contributed to success metrics by applying social success metrics.

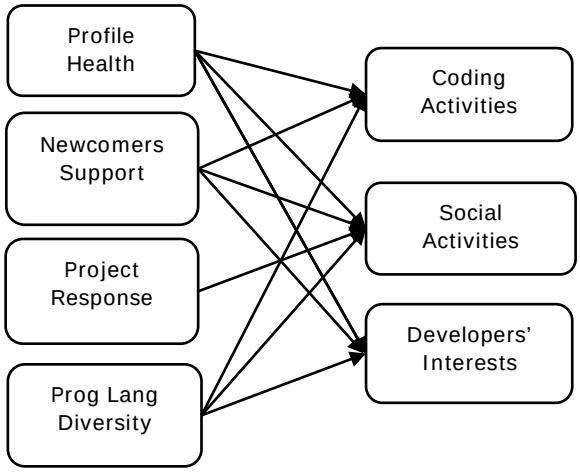

Figure 1. Research Model

Newcomers Support (NS) illustrates how a project supports newcomers to join and contribute to the projects. The literature of the online community has mentioned the role of organization orientation in newcomers' satisfaction (Bauer and Erdogan 2011). However, its effect on OSS project success is studied for the first time in this piece of research. It is a common strategy among OSS projects in GitHub to label the issues submitted to the bug tracker. Some of these labels indicates the suitability of a task for newcomers including the terms such as "Easy", "Low Hanging Fruits", "Starter", "Beginner", "Easy Fix", "Newbies", "Junior", "Good First", "First Timer Only", "Beginner Friendly", "Newcomers" and some more similar terms. We have approached to the software engineering forums such as StackOverflow.com and listed these terms to find out which projects are using these terms. We used a binary measure to show that a project using these labels or not. We have argued having supportive strategies lead to more newcomers' involvement and more socio-technical commitment in the project.

Project Response (PR) depicts how the project responds to a non-member contribution. The importance of this construct is already mentioned in project popularity and online communities health (Jarczyk et al. 2014; Wagner et al. 2014). In this study, we measured this construct regarding two major events run by non-members in GitHub including pull requests and issues (Bug report or feature request). We measure it with the rate of closed/merged pull requests and the rate of issues closed on one side. It is clear that closing more issues and pull request requires more technical efforts and also has been already tested in literature (Zhang et al. 2013). Then we argued more responses from project to the issues are related with the social commitment to the project.

Project Health (PH) is an independent factor. The Health value is collected for each project from its GitHub community profile. GitHub recently added a new parameter on the project profile with the health value in percentage. We used it as the independent variable in our analysis. Based on GitHub advice community health profile is calculated based on the checklist of availability of project description, required documentation, license and template formats. We argued that more structured projects in terms of documentation and licensing are more active socially and technically. Also, they are more interesting for developers to participate.

Programming Language Diversity (LD) illustrates the number of different programming languages used in a project. We argued that if a project uses a different type of programming languages in their code base then it may bring the more attention from the larger set of community members and more contribution is expected. Cultural diversity and language diversity was already discussed in the literature of open source project success factors.

Socio-Technical Activities (SA and TA) these two factors are considered as dependent factors in this research models. Socio-technical congruence and Conwey's Law in OSS literature are mentioned as a factor of OSS project health performance (Amrit and Van Hillegersberg 2010; Bolici et al. 2009). For the technical factors following the literature (Midha and Palvia 2012; Subramaniam et al. 2009), we applied the number of commit to the Git repository of the project. For the social and behavioural activities, we considered the number of comments GitHub members provided on each issue submitted to the project issue tracker. Discussion and comments are used in social coding literature with other perspectives (Daniel and Stewart 2016).

Developers' Interests (DI) is the dependent factor which is measured by the number of distinct committers in a project. This factor reveals how interesting and welcoming is a project to developers for their technical commitment. This factor is already used in open source literature (Ghapanchi et al. 2011; Subramaniam et al. 2009).

DATA COLLECTION

We have collected our data from GitHub as the largest open source project community. By the June 2018, the number of user accounts created on GitHub reached to 40 million and more than 100 Million repositories were created. GitHub provides a public API to get the projects and users data. In Addition, GhTorrent archives the information about projects and users activities in GitHub (Gousios 2013). We have used the GhTorrent data dump and GitHub public API together to capture all the required data about the projects that we need for this analysis. In this study our unit of analysis is a project and we emphasize all activity types project-wise.

As it is mentioned in (Kalliamvakou et al. 2014), it is required to apply a robust data cleaning process on GitHub data. A large set of repositories hosted on GitHub are forked, personal or inactive projects. We have to remove them from our list. As this is a preliminary study, we have collected a random set of 453 projects which are created in 2012. This year is selected as GitHub reached the high maturity level in OSS community in 2011-12. Then we have aggregated all the social and technical activities on each project for a period of 2 years after the creation date. The summary statistics about this preliminary dataset is presented in table 1. IsSupportive variable describes that the project uses the task labelling for newcomers or not in Boolean format. 46 projects out of 453 are newcomers' supportive. We have converted this variable to a dummy variable. As table1 shows, we need to apply a transformation method (logarithmic transformation) on our data as it has skewness in most of attributes.

Table 1. Summary Statistics

Attribute	Definition	Min	Max	Median	Mean	Std
Health	Project Health Level	0	100	16	23.02	22.02
LngCnt	#Programming Languages used in the project	0	23	2	3.27	3.25
PRClosedCnt	#Pull Request Closed	0	2424	0	93.46	312.06
IssueClosedCnt	#Issues Closed	0	1434	0	54.84	164.06
IssueComment	#Comments on Issues	0	10472	0	243.6	980.79
Commits	#Code Commits	0	10676	93	514.3	1233.11
Committers	#Committers	0	364	6	12.6	27.93

Correlation matrix of the variables in our dataset is listed in table 2. All the correlation values are significant and we do not have a very strong values (>0.8) in this matrix. Then, there would not be a multi-collinearity issue and it is safe to apply regression analysis for confirmatory analysis.

Table 2. Correlation Matrix

Attribute	1	2	3	4	5	6	7	8
1-Health	1	0.30	0.47	0.49	0.45	0.45	0.54	0.39
2-LngCnt		1	0.27	0.28	0.23	0.24	0.43	0.27
3-PRClosedCnt			1	0.71	0.37	0.72	0.77	0.46
4-IssueClosedCnt				1	0.34	0.66	0.64	0.32
5-IsSupportive					1	0.46	0.47	0.35
6-IssueComment						1	0.67	0.39
7-Commits							1	0.62
8-Committers								1

EXPLORATORY DATA ANALYSIS

In this section, we explore data set regarding our research perspective to have a better understanding of our dataset. Firstly we have gone through the programming languages of each project to see which programming language community are more supportive for newcomers. Table 3, shows the most popular languages in our dataset. JavaScript, Python and C# have the highest rate regarding supportive strategies for OSS newcomers. JavaScript community compared to Java is more than three times supportive for newcomers.

Table 3. Supportive strategies in different programming languages.

PL	Supportive	Total	Rate
JavaScript	14	73	0.19
Java	4	70	0.06
PHP	5	58	0.08
Python	9	53	0.17
Ruby	2	43	0.05
C	1	23	0.04
C++	1	16	0.06
Shell	0	14	0
Objective-C	0	12	0
C#	1	10	0.1

Next, we analyse the OSS licences to see the newcomers' supportive projects are using which licences. We find that projects with permissive licences such as MIT, Apache and BSD are more supportive to newcomers rather than restrictive ones. Moderate protection such as LGPL licences is fitting in between in terms of newcomers supportive. We can argue that project owners who care about newcomers are less restrictive in terms of licence choice. This requires confirmatory analysis.

Table 4. Supportive strategies in different OSS licences

Licence	Supportive	Total	Rate
MIT	10	37	0.27
Apache	3	26	0.12
BSD 3	7	21	0.33
LGP V3.0	1	11	0.09
GPL V2.0	0	9	0

We investigate the relationship between health levels of project profile with the newcomers' supportive strategies and find that supportive project are generally healthier. It also shows project owners who care

about the newcomer's task selection also cares about the preparing contribution document, templates and licences. Figure 2. Shows the gap in project health level in supportive OSS projects with others.

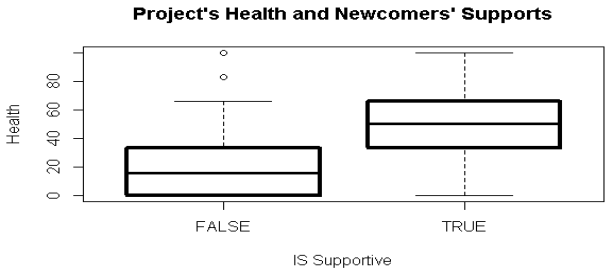

Figure 2. Project Profile Health Level regarding newcomers' supportive strategies

In our dataset the most common labels for the tasks related to newcomers are "Easy", "Beginner", "Good First Bug", "Low Hanging Fruit", "New contributor", "Starter", and "Newbies". This can help project coordinators to use a consistent terminology regarding supporting newcomers in social coding platforms. More comparison on IsSupportive attributes regarding the main socio-technical activities are shown in figure 3. The boxplots clearly shows the differences in the level of socio-technical activities in the newcomers' supportive projects. These plots can explain our main arguments in this study regarding the role of newcomers' supportive strategies in OSS projects.

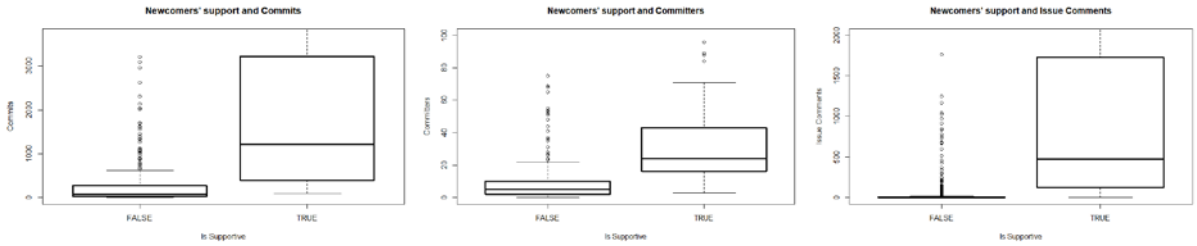

Figure 3. Comparing projects employed newcomers' supportive strategies with others (Commits, Committers, Issue Comments)

Clustering

To understand the data set and relationship among variables, we have applied K-Means (Witten et al. 1999) clustering to our dataset. Unsupervised technique such as clustering can help to have a better idea about the similarities across projects. Applying cross validation technique through Expectation Maximization clustering technique in Weka, suggests 7 as a suitable number of clusters. Then we have used Weka to run K-Means with 7 clusters (K=7). In this model we have used Euclidean distance as a distance function. Clusters are labelled from 0 to 6. The largest cluster is Cluster 0 with 135 (30%) data points and Cluster 1 is the smallest one with 18 (4%) data points. All the newcomers' supportive projects are positioned in Cluster 6 with 46 (10%) members. Figure 4 compares these cluster distribution based on the main socio-technical activities studied in this paper.

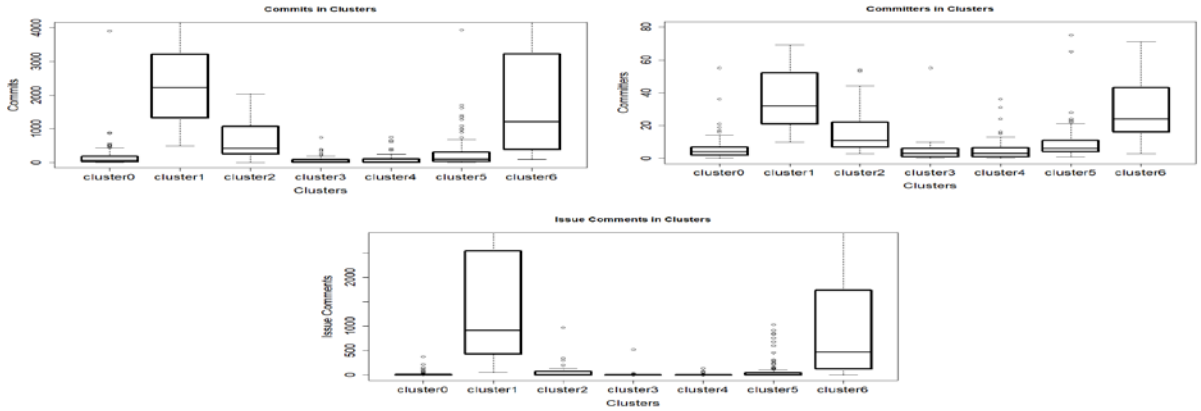

Figure 4. Cluster Analysis on the dataset (Commits, Committers, Issue Comments)

According to the outcome of visualization process and based on box-plot analysis we can figure out the cluster 6 (Newcomers' Supportive) and cluster 1 are the most socially and technically active projects. To realize the importance of our variable sets to define projects cluster we have employed Random Forest machine learning technique (Bayati and Peiris 2018). Figure 5 represents the level of importance for each variable. Project profile health, diversity in programming languages used in a project and newcomers supports through labelling are the most important factors to define project clusters. Figure 5 is divided into two parts. Mean Decrease Accuracy define the importance of factor of on the accuracy of predictive model in the absence of the factor. Mean Decrease Gini reveal the decrease level on the decision trees node in the absence of each variable (Ghasemkhani et al. 2015).

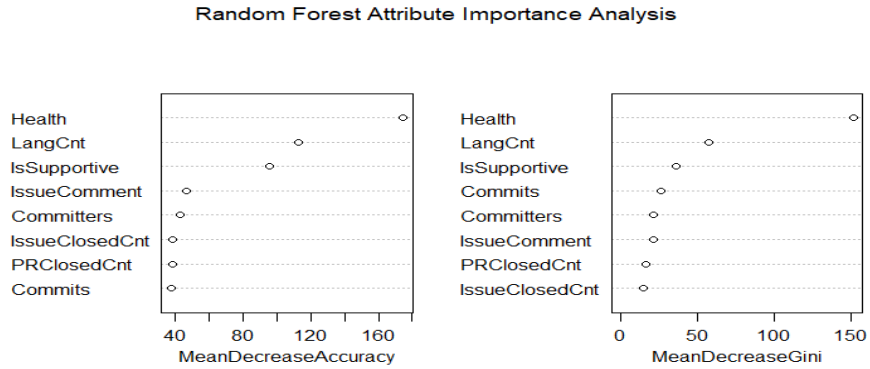

Figure 5. Varible importance in cluster analysis based on Random Forest Technique

Results and Discussion

On this section, we have presented the outcome of data analysis on the extracted dataset of GitHub projects. We used R Programming for our multiple regression analysis. We have applied a logarithmic transformation on the skewed distributed attributes. We ran three regression models based on the logarithmically transformed values of code commits, issue comments, developer's interests.

Table 5. Regression Results for Technical Commitment

	Estimate(Std Error)	P-Value (Sig)
Intercept	-376.797 (73.28)	0.000 (***)
IsSupportive	1015.148 (2.31)	0.000 (***)
Health	19.518 (14.56)	0.000 (***)
LngCnt	102.253 (165.44)	0.000 (***)
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 944.4 Multiple R-squared: 0.4185, Adjusted R-squared: 0.4146 F-statistic: 107.5, p-value: < 2.2e-16		

Table 5 shows the result of the regression model for technical activities. The outcome supports all of our arguments about technical activities. The R2 value for the first model is 0.41 which shows the fitness of model. All the coefficient values are positive and significant. It approves that having a strategy to support newcomers in open source projects will improve the amount of coding activities compare to other projects. This fact can reveal that while project coordinators make the process of task finding easier for a newcomer they will have more chance to succeed by on-boarding new members. To the best of our knowledge, this finding is a new finding in the field of newcomers' engagement in OSS and online communities. If the codes base of a project consists of different languages it will have more coding activities. The health profile of a project is significantly associated with the level of technical activities.

Table 6. Regression Results for Social Commitment

	Estimate(Std Error)	P-Value (Sig)
Intercept	-42.347 (49.82)	0.396
IsSupportive	662.713 (112.19)	0.000 (***)
Health	0.182 (1.67)	0.843
PRClosedCnt	1.385 (0.13)	0.000 (***)
IssueClosedCnt	1.664 (0.26)	0.000 (***)
LngCnt	-1.949 (9.82)	0.913
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 629 Multiple R-squared: 0.5941, Adjusted R-squared: 0.5896 F-statistic: 130.6, p-value: < 2.2e-16		

Table 6 presents the result of empirical modelling regarding social activities. Although we reached to the acceptable R2 value of 0.59 in the case of social activities, it does not supports our hypothesis about the effect of programming language diversity and project health in the amount of social behavior activities. Our analysis shows a non-significant effect of having more languages on the level of discussions and comments inside the project. The reason could be that variety of programming languages in a project requires more knowledge in contributors about different languages for discussions. Similar to the human interactions while people are speaking in different languages there would be fewer communications among all. Developers with different programming languages capabilities may prefer to discuss in their own small communities. Formatting a project healthier in terms of profile documentation and license also does not show a significant correlation with social behavior in a project. A reason behind that could be discussed as more structured and template based contribution can restrict the freedom of volunteers to discuss the issues. Also, some of the questions may already have discussed in the available documents. This part requires more investigation. The role of newcomers' supportive strategies in social activities in OSS projects is significantly approved in this study. Also, it shows a significant effect of responses in a social discussion about OSS issues.

Table 7. Regression Results for Developers' Interest

	Estimate(Std Error)	P-Value (Sig)
Intercept	-0.962 (1.93)	
IsSupportive	19.172 (4.37)	0.000 (***)
Health	0.317 (0.06)	0.000 (***)
LngCnt	1.30 (0.38)	0.000 (***)
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 24.92 Multiple R-squared: 0.2111, Adjusted R-squared: 0.2059 F-statistic: 39.97, p-value: < 2.2e-16		

Table 7 illustrates the developers' interest level in OSS projects. IT also approve all of our hypotheses. The model reveals while we have the supportive strategies for newcomer inside a community, we may

gain more participants. Programming language diversity may lead to attracting more developers with different skills to the project. Also, developers are more interested in the healthier projects.

The outcome of this analysis can be helpful for OSS project organizers and sponsors. As it is vital for OSS projects to reach success both socially and technically, it is recommended to have a solid supportive strategy for newcomers such as technical guideline and task labelling. As it has been already mentioned in the literature (Steinmacher and Gerosa 2014; Steinmacher et al. 2015) one of the main challenges for newcomer is finding a right task. This study prove that it will significantly improve the level activities in project.

Limitations

Despite interesting results of this study regarding the effect of newcomers supportive strategies in OSS field, we are aware of some limitations in our study. We have chosen a small random set of projects created on GitHub. We need to apply the same analysis on a bigger range of projects from other years as well. Although GitHub is known as the largest repository of OSS projects, however, BitBucket and OpenHub (Ohloh) also may provide more information. As, GitHub and GhTorrent provide a timestamped data about project events in GitHub we can run a longitudinal data analysis to have more robust result over time. In terms of social activities, we only focused on discussions on issues, along with this developers can follow each other on GitHub and there is an ability to subscribe on the users' favorite repositories. Technical activities in GitHub is not just coding. Code review and bug finding also require technical activity. In this study, we have considered all the code commits with same value however, in reality depending on the task and commit size the importance level of commit may be different.

CONCLUSION

This study investigates the relationship of newcomers' supportive strategies in a social coding environment with the level of project success. This study defines the project success in terms of amount of socio-technical activities and developer technical attention to the projects. Using GitHub data we have found a strong positive relationship with task labelling for newcomers and socio-technical activities in OSS community. In addition, this study contributes to the literature by considering the project health and measuring success with socio-technical activities. Exploratory data analysis shows permissive licence choice on projects is more aligned with newcomers supportive strategies. Results of this study have implications for managers and OSS researchers.

For future, we have the plan to apply the same study with a longitudinal dataset. We can also apply other success factors of the literature by gaining more data about each project.

REFERENCES

Amrit, C., and Van Hillegersberg, J. 2010. "Exploring the Impact of Socio-Technical Core-Periphery Structures in Open Source Software Development," journal of information technology (25:2), pp. 216-229.

Bauer, T. N., and Erdogan, B. 2011. "Organizational Socialization: The Effective Onboarding of New Employees," APA handbook of industrial and organizational psychology (3), pp. 51-64.

Bayati, S., and Peiris, K. 2018. "Road to Success: How Newcomers Gain Reputation in Open Source Community," in: PACIS 2018. Japan: AIS.

Bolici, F., Howison, J., and Crowston, K. 2009. "Coordination without Discussion? Socio-Technical Congruence and Stigmergy in Free and Open Source Software Projects," Socio-Technical Congruence Workshop in conj Intl Conf on Software Engineering, Vancouver, Canada.

Canfora, G., Di Penta, M., Oliveto, R., and Panichella, S. 2012. "Who Is Going to Mentor Newcomers in Open Source Projects?," Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering: ACM, p. 44.

Crowston, K., Howison, J., and Annabi, H. 2006. "Information Systems Success in Free and Open Source Software Development: Theory and Measures," Software Process: Improvement and Practice (11:2), pp. 123-148.

Dabbish, L., Stuart, C., Tsay, J., and Herbsleb, J. 2012. "Social Coding in Github: Transparency and Collaboration in an Open Software Repository," Proceedings of the ACM 2012 conference on computer supported cooperative work: ACM, pp. 1277-1286.

Daniel, S., and Stewart, K. 2016. "Open Source Project Success: Resource Access, Flow, and Integration," The Journal of Strategic Information Systems (25:3), pp. 159-176.

Delone, W. H., and McLean, E. R. 2003. "The Delone and Mclean Model of Information Systems Success: A Ten-Year Update," Journal of management information systems (19:4), pp. 9-30.

Ghapanchi, A. H., Aurum, A., and Low, G. 2011. "A Taxonomy for Measuring the Success of Open Source Software Projects," First Monday (16:8).

Ghasemkhani, H., Reichman, S., and Westerman, G. 2015. "Using Predictive Analytics to Reduce Uncertainty in Enterprise Risk Management,").

Gousios, G. 2013. "The Ghtorent Dataset and Tool Suite," Proceedings of the 10th Working Conference on Mining Software Repositories: IEEE Press, pp. 233-236.

Grewal, R., Lilien, G. L., and Mallapragada, G. 2006. "Location, Location, Location: How Network Embeddedness Affects Project Success in Open Source Systems," Management Science (52:7), pp. 1043-1056.

Hahn, J., Moon, J. Y., and Zhang, C. 2008. "Emergence of New Project Teams from Open Source Software Developer Networks: Impact of Prior Collaboration Ties," Information Systems Research (19:3), pp. 369-391.

Hertel, G., Niedner, S., and Herrmann, S. 2003. "Motivation of Software Developers in Open Source Projects: An Internet-Based Survey of Contributors to the Linux Kernel," Research policy (32:7), pp. 1159-1177.

Jansen, S. 2014. "Measuring the Health of Open Source Software Ecosystems: Beyond the Scope of Project Health," Information and Software Technology (56:11), pp. 1508-1519.

Jarczyk, O., Gruszka, B., Jaroszewicz, S., Bukowski, L., and Wierzbicki, A. 2014. "Github Projects. Quality Analysis of Open-Source Software," International Conference on Social Informatics: Springer, pp. 80-94.

Kalliamvakou, E., Gousios, G., Blincoe, K., Singer, L., German, D. M., and Damian, D. 2014. "The Promises and Perils of Mining Github," Proceedings of the 11th working conference on mining software repositories: ACM, pp. 92-101.

Midha, V., and Palvia, P. 2012. "Factors Affecting the Success of Open Source Software," Journal of Systems and Software (85:4), pp. 895-905.

Piggot, J., and Amrit, C. 2013. "How Healthy Is My Project? Open Source Project Attributes as Indicators of Success," IFIP International Conference on Open Source Systems: Springer, pp. 30-44.

Roberts, J. A., Hann, I.-H., and 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), pp. 984-999.

Schilling, A. 2012. "Links to the Source-a Multidimensional View of Social Ties for the Retention of Floss Developers," Proceedings of the 50th annual conference on Computers and People Research: ACM, pp. 103-108.

Schilling, A., Laumer, S., and Weitzel, T. 2012. "Who Will Remain? An Evaluation of Actual Person-Job and Person-Team Fit to Predict Developer Retention in Floss Projects," System Science (HICSS), 2012 45th Hawaii International Conference on: IEEE, pp. 3446-3455.

Singh, P. V. 2010. "The Small-World Effect: The Influence of Macro-Level Properties of Developer Collaboration Networks on Open-Source Project Success," ACM Transactions on Software Engineering and Methodology (TOSEM) (20:2), p. 6.

Singh, P. V., Tan, Y., and Mookerjee, V. 2011. "Network Effects: The Influence of Structural Capital on Open Source Project Success," Mis Quarterly), pp. 813-829.

Steinmacher, I., and Gerosa, M. A. 2014. "How to Support Newcomers Onboarding to Open Source Software Projects," IFIP International Conference on Open Source Systems: Springer, pp. 199-201.

Steinmacher, I., Silva, M. A. G., Gerosa, M. A., and Redmiles, D. F. 2015. "A Systematic Literature Review on the Barriers Faced by Newcomers to Open Source Software Projects," Information and Software Technology (59), pp. 67-85.

Stewart, K. J., Ammeter, A. P., and Maruping, L. M. 2006. "Impacts of License Choice and Organizational Sponsorship on User Interest and Development Activity in Open Source Software Projects," Information Systems Research (17:2), pp. 126-144.

Subramaniam, C., Sen, R., and Nelson, M. L. 2009. "Determinants of Open Source Software Project Success: A Longitudinal Study," Decision Support Systems (46:2), pp. 576-585.

Tsay, J., Dabbish, L., and Herbsleb, J. 2014. "Influence of Social and Technical Factors for Evaluating Contribution in Github," Proceedings of the 36th international conference on Software engineering: ACM, pp. 356-366.

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), pp. 1217-1241.

Wagner, D., Richter, A., Trier, M., and Wagner, H.-T. 2014. "Towards a Conceptualization of Online Community Health,").

Witten, I. H., Frank, E., Trigg, L. E., Hall, M. A., Holmes, G., and Cunningham, S. J. 1999. "Weka: Practical Machine Learning Tools and Techniques with Java Implementations,").

Zhang, C., Hahn, J., and De, P. 2013. "Research Note—Continued Participation in Online Innovation Communities: Does Community Response Matter Equally for Everyone?," Information Systems Research (24:4), pp. 1112-1130.