View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by AIS Electronic Library (AISeL)

Association for Information Systems

AIS Electronic Library (AISeL)

AMCIS 2020 Proceedings

Virtual Communities and Collaboration (VCC)

Aug 10th, 12:00 AM

Welcome? Investigating the reception of new contributors to organizational-communal open source software projects

Kevin Lumbard
University of Nebraska at Omaha, klumbard@unomaha.edu

Anna Buhman
University of Nebraska at Omaha, abuhman@unomaha.edu

Gabrielle E. Wethor
University of Nebraska at Omaha, gwethor@unomaha.edu

Matthew Hale
University of Nebraska at Omaha, mlhale@unomaha.edu

Sean Goggins
University of Missouri, s@goggins.com

See next page for additional authors

Follow this and additional works at: https://aisel.aisnet.org/amcis2020

Lumbard, Kevin; Buhman, Anna; Wethor, Gabrielle E.; Hale, Matthew; Goggins, Sean; and Germonprez, Matt, "Welcome? Investigating the reception of new contributors to organizational-communal open source software projects" (2020). *AMCIS 2020 Proceedings*. 14.
https://aisel.aisnet.org/amcis2020/virtual_communities/virtual_communities/14

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2020 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Presenter Information

Kevin Lumbard, Anna Buhman, Gabrielle E. Wethor, Matthew Hale, Sean Goggins, and Matt Germonprez

This event is available at AIS Electronic Library (AISeL): https://aisel.aisnet.org/amcis2020/virtual_communities/virtual_communities/14

Welcome? Investigating the reception of new contributors to organizational-communal open source software projects

Completed Research Full Paper

Kevin Lumbard
University of Nebraska at Omaha
klumbard@unomaha.edu

Anna Buhman
University of Nebraska at Omaha
abuhman@unomaha.edu

Gabrielle Wethor
University of Nebraska at Omaha
gwethor@unomaha.edu

Matthew Hale
University of Nebraska at Omaha
mlhale@unomaha.edu

Sean Goggins
University of Missouri
s@goggins.com

Matt Germonprez
University of Nebraska at Omaha
germonprez@gmail.com

Abstract

The way new contributors are received by the established contributors in an open source project is a factor in whether they will become more regular contributors. This research examines the reception of new contributors in three open source projects to discover whether there are differences in how established contributors respond to new contributors, and if so, what those differences are. Through statistical analysis of time to first response and sentiment analysis of that response to a new contributor's issue, we found that there is a difference in both the speed and content of responses to new contributors' issues as opposed to those of established contributors. This difference suggests that the open source projects we observed are attentive to whether an issue was created by a new contributor and may make an effort to respond in a welcoming manner.

Keywords

Open source, reception, new contributors

Introduction

In open source projects, reception of new contributors by established contributors is a significant predictor of future contribution. Reception of new contributors can be measured through timeliness and sentiment of responses to new contributors by established contributors (Steinmacher et al. 2015). This reception perceived by new contributors on a spectrum from inviting to unwelcome shapes their perceptions of the project and their willingness to engage further (Cleary et al. 2013; Fogel 2005). For these reasons, reception is an important indicator of open source project sustainability.

Given the transparency of community commentary associated with open development (Dabbish et al. 2012) and visible workflows (Howison and Crowston 2014), reception can affect more than just new contributors and, when perceived as unwelcoming, may proactively ward off others from involvement (Cleary et al. 2013; Fogel 2005; Steinmacher et al. 2015). Problematic reception has been shown to affect many open source projects (Steinmacher et al. 2015). Whereas many online communities provide anonymity, contributors to open source projects are often aware of one another (Gutwin et al. 2004). This visibility and ability for contributors to seek out information about one another to inform their interactions (Marlow et al. 2013) means that bad behavior (i.e. impolite or uncivil discourse), when it occurs, is on display for all to see and may adversely affect perceptions of a project.

As corporate engagement in open source increases, the reception of new contributors becomes a managed concern. Open source projects with high levels of corporate engagement are referred to as organizational-communal projects. Encouraging and recruiting new contributors is a key requirement for creating healthy, sustainable design streams (O'Mahony and Ferraro 2004) demanded by organizational-communal projects (Germonprez et al. 2018; Link and Germonprez 2018). This study examines the reception of new contributors in organizational-communal projects using Steinmacher's (2015) barrier model. In particular, this study examined delayed and non-existent responses and sentiment as reception barriers to new contributor recruitment and retainment. Three open source projects within the Linux Foundation ecosystem are used as representative data sources to pose the following research question.

> *How are contributors to organizational-communal open source projects received when interacting with a project for the first time?*

Background

Open Source Software Projects

Contributors have many different motivations for joining an open source project, including intrinsic and extrinsic factors (von Krogh et al. 2012) and provides a source of new ideas that may lead to innovation (Kraut et al. 2012). Established contributors play a central role in open source projects and shape the experience of new contributors. Established contributors may be regulators, community managers, and community members and outside collaborators who have the technological ability to decide what is allowed in their project. Some examples include the power to delete online posts or approve contributions (Barzilai-Nahon 2006). In this context, established contributors of open source projects may shepherd or block code changes based on many factors, including quality decisions (Asundi and Jayant 2007) or biases (Terrell et al. 2017).

New contributors and established contributors often seek out information about one another, and that information, along with first impressions, influences how receptive they are to collaboration (Marlow et al. 2013). For example, established contributors will often look at the profiles and github histories of new contributors to assess skills and reputation when they receive a pull request from a new contributor. A project's climate can determine whether a contributor becomes a part of a project or chooses to disengage (Zhou and Mockus 2012). Due to the transparent nature of open source projects, contributors are often aware of one another (Gutwin et al. 2004), and conflict resolutions are visible as elements of project climate (Elliott and Scacchi 2003; Filippova and Cho 2015, 2016). Further, emotionally negative communication has been shown to drive away contributors (Guberman et al. 2016; Shores et al. 2014).

Because the number of contributors working in open source projects may correlate with the success of a project (Crowston et al. 2006), the positive reception of contributions from new contributors is essential. To encourage participation for new contributors, open source projects use socialization mechanisms during the reception and joining process to increase the likelihood of successful interaction (Ducheneaut 2005). Some projects have processes in which new contributors are onboarded through sponsorship with established contributors (O'Mahony and Ferraro 2004), and many open source projects create joining scripts that provide guidance for new contributors. New contributors who follow those guidelines are more likely to be received positively by established contributors (von Krogh et al. 2003). New contributors who start by commenting on issues are more likely to become long-term contributors to a project (Zhou and Mockus 2012), and new contributors who have specialized knowledge or focus on specific types of tasks have more long-term success (von Krogh et al. 2003).

Social Barriers and Reception

Open source projects are complex social engagements, and the reception that new contributors get when they join a project determines future contributions (Fogel 2005). New contributors often face barriers related to reception such as delayed responses, no responses, and negative sentiment in responses when joining an open source project (Steinmacher et al. 2015). Steinmacher et al. (2015) explored the barriers to newcomers. These included reception issues, newcomers' characteristics, newcomers needing orientation, documentation problems, and cultural differences. Our research focuses on the reception issues, and within those issues, focuses on delayed answers and impolite answers. When a new contributor attempts to interact with a project but fails to get a response, research has shown that they are unlikely to return to the project (von Krogh et al. 2012; Singh et al. 2012; Steinmacher et al. 2015). Timeliness is also crucial, and newcomers rarely return if they fail to receive a response within 24 hours of an attempted interaction (Jensen et al. 2011). Impolite or negative responses within contributor communication have also been shown to make participation difficult (Steinmacher et al. 2015). Previous research has explored newcomer reception primarily in volunteer environments. Our research adds to this by exploring newcomer reception in organizational-communal open source projects environments.

Method

Three organizational-communal open source projects under the umbrella of the Linux Foundation, namely Kubernetes[1], Node.JS[2], and Zephyr[3], are used as representative data sources in this study to investigate new contributor reception. Of particular interest are first-time interaction and new contributor experiences. Kubernetes is an open source system for automating deployment, scaling, and management of containerized applications. Node.js is a JavaScript runtime built on Chrome's V8 JavaScript engine. Zephyr is an open source scalable real-time operating system for small Internet of Things devices. These projects were selected because they share similar, high levels of organizational interest. Additionally, project work for all three occur on the GitHub[4] platform, making data collection repeatable for each project. The three projects selected show different levels of activity, providing some variation in the contexts within which the research question is addressed, as demonstrated by project artifact levels at the time of the data extraction (see Table 1).

Project	Pull Request	Code Contributors	Comment Contributors	File Changes	Comments
Kubernetes	42,485	2,038	11,465	727,218	464,834
Node.js	15,531	2,405	8,274	717,526	132,516
Zephyr	7,078	439	515	87,858	22,474

Table 1: Project Activity

The methodology employed in this study is informed and enriched by an eight-year ongoing field study conducted by members of the research team, who have become active participants in numerous open source projects. This vantage point positions the research team and, by proxy, this study to provide a broad perspective on open source projects. Context is important and the methodology of the study applies learned insights from open source interaction to ensure project artifacts were examined within the frame of reference on GitHub in which they were created. The following sections detail the particular types of data used in this study, the mechanisms of collection used to gather them, and the kinds of analysis techniques applied across the data to explore the research question.

Trace Data

Trace data, or intermediate project artifacts emerging from interaction or contribution to a project, are the primary unit of analysis in this study. Most trace data relevant to reception emerges from *issue tracking threads* within the project. Common issue topics include bug reports, new feature requests, project planning discussions, and support activities. Responses to issue comments and the associated trace data provide a means to analyze timeliness and sentiment. To extract project trace data and tabulate aggregate statistics, the study uses GraphiQL (Jindal and Madden 2014) and the R programming language to access open source project repositories via the GitHub API. Three projects were queried, and data associated with the project's issue forums was extracted on February 4th, 2018. Specifically, the *issue URL*, *publish at time*, *author association* (issue author contributor type), *author username*, *first response name* (used to eliminate responses by the issue author), *first response time*, and *comment text* were extracted.

[1] https://kubernetes.io/

[2] https://nodejs.org/en/

[3] https://www.zephyrproject.org/

[4] https://github.com/

Activity Analysis

Response time correlates highly with reception (Jensen et al. 2011). In this study, *response time* is defined as the time difference between when an issue is originally created and the time of the first response by another user within the issue thread. The response time metric precludes scenarios where the first response to an issue was the issue author themselves. The first response time was subtracted from the published at time in order to generate a response time. As part of the activity analysis in this work, response time was computed for all issues across both new and established contributors to determine if there were significant timeliness differences between receptivity.

Sentiment Analysis

Natural language (NL) sentiment analysis techniques can be used to examine trace data. NL sentiment analysis techniques score textual content according to two factors: *polarity* and *subjectivity*. Polarity is a measure of emotional valence that ranges from negative to neutral to positive. Subjectivity is a continuum that indicates either fact-centered or objective commentary. This measure ranges from “objective” (very fact centered) to “subjective” where observations and statements are made with high amounts of personal perspective interlaced within them. Prior research has used sentiment analysis to examine contributions on GitHub (Guzman et al. 2014; Pletea et al. 2014; Sinha et al. 2016). In this study, TextBlob (Loria et al. 2014), an open source Python library for processing textual data, was used to analyze issue comment text. TextBlob provides a simple API for diving into common natural language processing, including sentiment analysis. Previous research has shown that TextBlob has a good accuracy for text extraction and NLTK (Natural Language Toolkit) for short texts such as social media comments and reviews (Arai and Tolle 2011; Gauba et al. 2017; Hasan et al. 2018; Micu et al. 2017; Rajesh and Gandy 2016; Sahni et al. 2017; Vijayarani and Janani 2016). GitHub Issue comments span the topics of bug reporting, design discussion, project organization, and feature requests. An inspection of the issue comments in GitHub revealed that the messages were generally short and written in informal language, making TextBlob a good candidate for analyzing whether or not there is a difference in response sentiment between user types. Given a text sample, in this case, the aggregate text spanning comments in the issue, TextBlob outputs an analysis report estimating the overall polarity and subjectivity of the text following the two ranges below:

Polarity
-1 (very negative) to 1 (extremely positive)

Subjectivity
0 (very objective) to 1 (very subjective)

Using TextBlob, the sentiment of the comments within issues created by new contributors and established contributors are compared in aggregate to determine if, and by how much, the overall subjectivity and polarity varied between the data sets.

Findings: Organizational-communal Open Source Reception

Timely Interactions

A Kruskal Wallis test was performed to determine significance ($p < .05$) because response time data was right-skewed and non-normally distributed. After analyzing the trace data from each of the project GitHub repositories, results showed that response times for new contributors are significantly faster than responses for established contributors for two of the projects - Node.js and Zephyr (Table 2). For Kubernetes, the response time is slightly worse for new contributors than for established contributors, however the response times are close to one another (approximately 15 minutes difference). All three projects respond to new contributors in under 24 hours. Kubernetes established contributors respond to new contributors in a median time of approximately six hours, Node.js responds in approximately 35 minutes, and Zephyr responds in approximately eight hours. Figure 1 shows side-by-side response time boxplots for each author association type by project.

Project	Author Association	Issues	Median Response Time secs (hours)	CHI Square Statistic	p-value (p < .05)	Effect size (phi)
Kubernetes	New Contributor	5,210	22,931.00 (6.37)	23.675	1.14E-06	.19
	Established Contributors	10,698	21,917.00 (6.09)			
Node.js	New Contributor	4,488	2,120.00 (0.59)	126.02	2.20E-16	1.49
	Established Contributors	2,711	4,313.00 (1.20)			
Zephyr	New Contributor	202	29,930.00 (8.31)	47.07	6.85E-12	1.36
	Established Contributors	1,000	203,939.00 (56.65)			

Table 2: Response Time

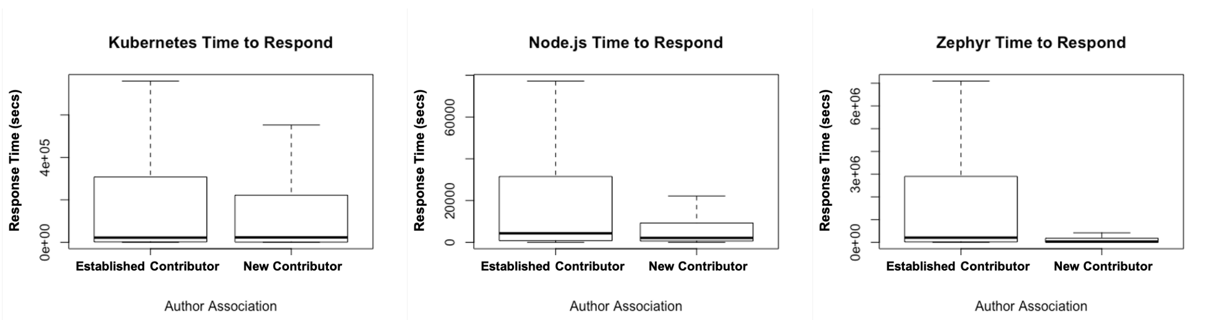

Figure 1: Boxplot of Response Time and Author Association

There are differences in project activity between the three projects that may explain some of the varying results. The largest amount of activity in Kubernetes and Zephyr is from established contributors. In Node.js, however, the largest amount of activity is from new contributors. This difference may help explain the incredibly fast response time that new contributors receive in Node.JS. Also, of note is the level of activity for the projects in regard to the number of contributors. At the time this data was extracted, Kubernetes and Node.JS had approximately 5 times the number of contributors of Zephyr which had the slower response times. The boxplots show that the range of response times for new contributors is consistently smaller than for established contributors for all three projects and there is less variation in the time it takes to respond to a new contributor as opposed to an established contributor. This may mean that new contributors post things that require a more urgent response, that they post things that people feel more comfortable answering quickly, or that established contributors feel more comfortable allowing certain posts of other established contributors to go unanswered longer. In general, new contributors get better response times to established contributors for all of the projects and the effect sizes for Node.js and Zephyr are pretty substantial.

Sentiment Analysis

The sentiment data had a normal distribution pattern, so an ANOVA test was used to determine statistical significance ($p < .05$) for polarity and subjectivity data across each author association type for each project. Sentiment analysis from each of the three projects broken down by contributor type shows an average polarity sentiment range from 0.05-0.10 (polarity ranging between -1 to 1). Kubernetes showed a polarity value of .07 for new contributors, compared to .05 for all established contributors (p = 0.001). Zephyr showed a polarity value of .09 for new contributors, compared to .05 for all established contributors (p = 0.001). Node.js showed the same polarity value of .10 for new contributors and for all established contributors; however, these results were not found to be statistically significant (see Table 3). Figure 2 shows side-by-side polarity boxplots for each author association type by project.

Project	Author Association	Average Polarity	SD	SS	p-value (p < .05)	Effect Size (f)
Kubernetes	New Contributor	0.07	0.14	.49	0.001	.04
	Established Contributors	0.05	0.15			
Node.js	New Contributor	0.1	0.15	.06	0.093	.02
	Established Contributors	0.1	0.14			
Zephyr	New Contributor	0.09	0.17	.31	0.001	.10
	Established Contributors	0.05	0.15			

Table 3: Polarity

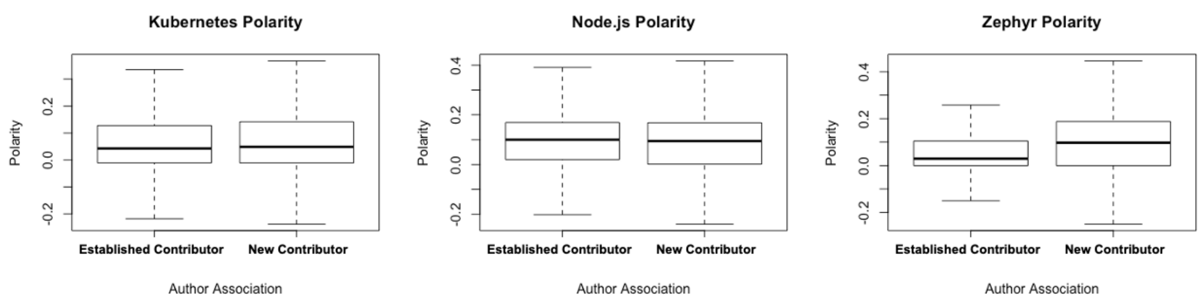

Figure 2: Boxplot of Polarity and Author Association

The boxplots show that in all three cases, the range of polarity for new contributors is slightly larger than for the established contributors and the variation in polarity is greater for new contributors than for established contributors. This is especially true of Zephyr. This indicates there is more variation in tone of response to new contributors than to established contributors in terms of polarity. A possible explanation for this could be that new contributors not knowing the norms of the group, may create more variety in their issues, which leads to more variety in responses. The difference in variation in responses to new contributors is similar across the three projects, but Zephyr gives more consistent responses to established contributors. This may be because Zephyr has fewer established contributors than either Kubernetes or Node.js, who have similar numbers of established contributors to each other.

Sentiment analysis from each of the three projects broken down by contributor type shows an average subjectivity range from 0.35-0.46 (subjectivity ranging between 0 to 1). Kubernetes showed a subjectivity value of .46 for new contributors, compared to .43 for all established contributors (p= 0.001). Zephyr showed a subjectivity value of .42 for new contributors, compared to .35 for all established contributors (p = 0.001). Node.js subjectivity was not statistically significant (see Table 4). Figure 3 shows side-by-side subjectivity boxplots for each author association type by project.

Project	Author Association	Average Subjectivity	SD	SS	p-value (p < .05)	Effect Size (f)
Kubernetes	New Contributor	0.46	0.19	3.37	0.001	.08
	Established Contributors	0.43	0.2			
Node.js	New Contributor	0.46	0.18	.01	0.567	.01
	Established Contributors	0.46	0.16			
Zephyr	New Contributor	0.42	0.22	.77	0.001	.11
	Established Contributors	0.35	0.24			

Table 4: Subjectivity

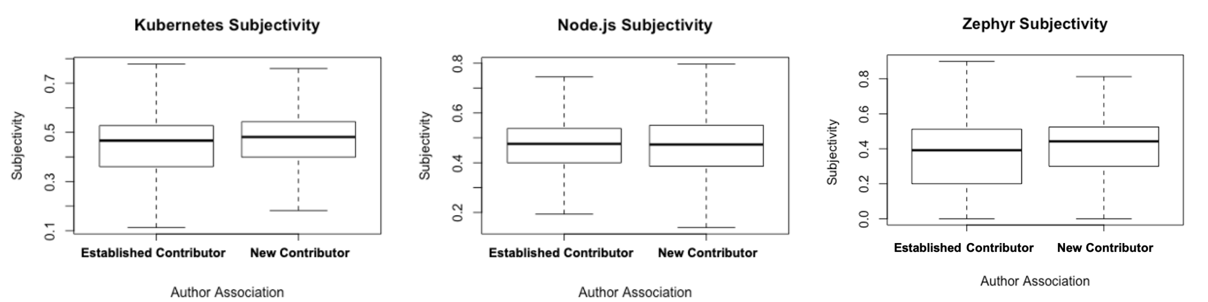

Figure 3: Boxplot of Subjectivity and Author Association

Kubernetes and Zephyr showed differences in subjectivity sentiment shown towards new contributors as compared to established contributors. Node.js showed the same subjectivity for new contributors and established contributors but was not significant. Average subjectivity sentiment for all projects ranged from 0.33-0.47 (subjectivity ranging between 0 to 1), indicating that the sentiment expressed was neither completely objective nor subjective. The boxplots show that in Kubernetes and Zephyr, the range of subjectivity for new contributors is slightly smaller than established contributors and there is less variation in subjectivity in responses. However, language expressed more often skewed towards objective wording.

Results of the sentiment analysis showed that for all three of the projects, sentiment skewed more towards neutral rather than positive or negative. Comments in GitHub issue forums are often technical, so it isn't surprising that the sentiment of the comments for all three projects is close to neutral polarity and leaning towards objective language. While the effect size and the differences in sentiment may seem small, there was a measurable difference in sentiment between the new contributors and existing contributors for issues in the Kubernetes and Zephyr projects. For the sentiment analysis what might be interesting is what we didn't find. There do not appear to be systematic issues of rudeness or incivility in any of the projects we explored.

Discussion

Prior research has shown that when contributors get a slow or highly negative response when joining a project, they may feel their contribution is not welcome (Steinmacher et al. 2015). This study examined reception issues in three large open source projects in the organizational-communal open source space. For the three cases we looked at, we noticed a measurable difference between response time and sentiment between the different types of author associations. In general, new contributors receive faster response times and different sentiment than established contributors however sentiment does not seem to be a factor in the projects we observed.

The importance of timely responses, particularly within 24 hours, was outlined by Jensen et al. (2011), which found that new contributors who fail to receive a response within 24 hours are likely to abandon the project. By that standard, all of the projects analyzed in this study show a concerted effort to respond to new contributors in a timely way. This result, while interesting, is not surprising given prior research, which has shown that more receptive open source projects are more likely to keep new contributors (Choi et al. 2010). From our field study observations, we posit that the differences in response times may indicate that different contributor types use the issues forum for different purposes. For example, in Zephyr slower response times for established contributors may indicate that the issues forum is being used for long-term planning, which may not require a timely response. Additionally, these contributors may use alternate forms of communication such as email and chat that new contributors do not always have access to.

The study also used sentiment analysis to establish if there was a difference in sentiment between new and established contributors. The study does not make positivity judgments about the projects themselves. Our results show that there is a difference in sentiment for two of the project cases. Whether this difference is good or bad is left to the projects themselves and future research to determine. Different projects show different variations in sentiment to new vs established contributors, but all of them show a difference, indicating that whether they are aware of it or not established contributors respond to new contributors using different sentiment. Future work could examine the reason for the observed differences. To fully understand sentiment in open source communities, a tool trained on content specific data should be trained and evaluated by a context expert.

For the projects we observed, differences in response time and sentiment show that established contributors are aware of which users are newcomers. The difference in how contributor types are received has implications for how we design open committees for sustainability (O'Mahony and Ferraro 2004). Our study shows that reception patterns may not be consistent between all types of open source projects. Even between some of the open source projects in our study, we saw variations in newcomer reception. From our field study we have observed that different projects have different strategies in receiving newcomers, which are likely based on the culture of each project and the viewpoints of the established users. All three of the projects analyzed in this study have documentation that addresses welcoming new contributors. Zephyr and Node provide a Contributor Covenant Code of Conduct, instructing existing contributors to use "welcoming and inclusive language". Kubernetes includes a contributor guide with a section listed as "Your First Contribution", which provides potential new contributors information on what issues would be "beginner-friendly" for new contributor support. The inclusion of these public guidelines may lower perceived barriers for new contributors in terms of reception.

This research examined three organizational-communal open source projects, which is a small sample compared to the many open source projects in existence. This small sample means the study doesn't necessarily describe all open source projects and, indeed, this study found inconsistencies in how open source projects receive newcomers when compared to previous results. Future work in this area could further study how open source projects differ from each other regarding newcomer reception, and why these differences exist. Another topic for future study could be how these differing approaches impact newcomers, and how these approaches influence newcomers to choose one open source project over another.

Conclusion

New contributors are important to maintain activity and grow open source projects. This study examined the reception of new contributors to organizational-communal open source projects and found that newcomers are received with quick responses, sometimes faster than responses to existing contributors, and with different levels of sentiment. These initial observations are likely the result of a concerted effort by community managers and established contributors. Further research is needed to explore how organization-communal open source projects are utilizing governance mechanisms to improve reception and how these initiatives affect long term sustainability.

Acknowledgements

This project received funding from the Alfred P. Sloan Foundation Digital Technology grant on Open Source Health and Sustainability, Num: 8434 (https://sloan.org/grant-detail/8434).

References

Arai, K., and Tolle, H. 2011. "Text Extraction from TV Commercial Using Blob Extraction Method," *International Journal of Research and Reviews in Computer Science* (2:3), p. 895.

Asundi, J., and Jayant, R. 2007. "Patch Review Processes in Open Source Software Development Communities: A Comparative Case Study," in *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, , January, pp. 166c–166c.

Barzilai-Nahon, K. 2006. "Gatekeepers, Virtual Communities and the Gated: Multidimensional Tensions in Cyberspace," *International Journal of Communications Law & Policy* (11), pp. 1–28.

Choi, B., Alexander, K., Kraut, R. E., and Levine, J. M. 2010. "Socialization Tactics in Wikipedia and Their Effects," in *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, CSCW '10, New York, NY, USA: ACM, pp. 107–116.

Cleary, B., Gómez, C., Storey, M.-A., Singer, L., and Treude, C. 2013. "Analyzing the Friendliness of Exchanges in an Online Software Developer Community," in *Cooperative and Human Aspects of Software Engineering (CHASE), 2013 6th International Workshop On*, IEEE, pp. 159–160.

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," in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, New York, NY, USA: ACM, pp. 1277–1286.

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

Elliott, M. S., and Scacchi, W. 2003. "Free Software Developers as an Occupational Community: Resolving Conflicts and Fostering Collaboration," in *Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work*, ACM, pp. 21–30.

Filippova, A., and Cho, H. 2015. "Mudslinging and Manners: Unpacking Conflict in Free and Open Source Software," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ACM, pp. 1393–1403.

Filippova, A., and Cho, H. 2016. "The Effects and Antecedents of Conflict in Free and Open Source Software Development," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, New York, NY, USA: ACM, pp. 705–716.

Fogel, K. 2005. *Producing Open Source Software: How to Run a Successful Free Software Project*, O'Reilly Media, Inc.

Gauba, H., Kumar, P., Roy, P. P., Singh, P., Dogra, D. P., and Raman, B. 2017. "Prediction of Advertisement Preference by Fusing EEG Response and Sentiment Analysis," *Neural Networks* (92), pp. 77–88.

Germonprez, M., Link, G. J. P., Lumbard, K., and Goggins, S. 2018. "Eight Observations and 24 Research Questions about Open Source Projects: Illuminating New Realities," *Proc. ACM Hum.-Comput. Interact.* (2:CSCW), 57:1–57:22.

Guberman, J., Schmitz, C., and Hemphill, L. 2016. "Quantifying Toxicity and Verbal Violence on Twitter," in *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, CSCW '16 Companion, New York, NY, USA: ACM, pp. 277–280.

Gutwin, C., Penner, R., and Schneider, K. 2004. "Group Awareness in Distributed Software Development," in *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, CSCW '04, New York, NY, USA: ACM, pp. 72–81.

Guzman, E., Azócar, D., and Li, Y. 2014. "Sentiment Analysis of Commit Comments in GitHub: An Empirical Study," in *Proceedings of the 11th Working Conference on Mining Software Repositories*, ACM, pp. 352–355.

Hasan, A., Moin, S., Karim, A., and Shamshirband, S. 2018. "Machine Learning-Based Sentiment Analysis for Twitter Accounts," *Mathematical and Computational Applications* (23:1), p. 11.

Howison, J., and Crowston, K. 2014. "Collaboration through Open Superposition: A Theory of the Open Source Way.," *MIS Quarterly* (38:1), pp. 29–50.

Jensen, C., King, S., and Kuechler, V. 2011. "Joining Free/Open Source Software Communities: An Analysis of Newbies' First Interactions on Project Mailing Lists," in *2011 44th Hawaii International Conference on System Sciences*, IEEE, pp. 1–10.

Jindal, A., and Madden, S. 2014. "GraphiQL: A Graph Intuitive Query Language for Relational Databases," in *2014 IEEE International Conference on Big Data (Big Data)*, IEEE, pp. 441–450.

Kraut, R. E., Resnick, P., Kiesler, S., Ren, Y., Chen, Y., Burke, M., Kittur, N., Riedl, J., and Konstan, J. 2012. *Building Successful Online Communities: Evidence-Based Social Design*.

von Krogh, G., Haefliger, S., Spaeth, S., and Wallin, M. W. 2012. "Carrots and Rainbows: Motivation and Social Practice in Open Source Software Development," *MIS Quarterly*, pp. 649–676.

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), Open Source Software Development, pp. 1217–1241.

Link, G. J. P., and Germonprez, M. 2018. "Assessing Open Source Project Health," in *Twenty-Fourth Americas Conference on Information Systems Proceedings*, New Orleans, LA, August.

Loria, S., Keen, P., Honnibal, M., Yankovsky, R., Karesh, D., and Dempsey, E. 2014. "Textblob: Simplified Text Processing," *Secondary TextBlob: Simplified Text Processing* (3).

Marlow, J., Dabbish, L., and Herbsleb, J. 2013. "Impression Formation in Online Peer Production: Activity Traces and Personal Profiles in Github," in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, ACM, pp. 117–128.

Micu, A., Micu, A. E., Geru, M., and Lixandroiu, R. C. 2017. "Analyzing User Sentiment in Social Media: Implications for Online Marketing Strategy," *Psychology & Marketing* (34:12), pp. 1094–1100.

O'Mahony, S. C., and Ferraro, F. 2004. "Managing the Boundary of an 'Open' Project," SSRN Scholarly Paper No. ID 474782, SSRN Scholarly Paper, Rochester, NY: Social Science Research Network, March 30.

Pater, J. A., Nadji, Y., Mynatt, E. D., and Bruckman, A. S. 2014. "Just Awful Enough: The Functional Dysfunction of the Something Awful Forums," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2407–2410.

Pletea, D., Vasilescu, B., and Serebrenik, A. 2014. "Security and Emotion: Sentiment Analysis of Security Discussions on GitHub," in *Proceedings of the 11th Working Conference on Mining Software Repositories*, pp. 348–351.

Qureshi, I., and Fang, Y. 2011. "Socialization in Open Source Software Projects: A Growth Mixture Modeling Approach," *Organizational Research Methods* (14:1), pp. 208–238.

Rajesh, N., and Gandy, L. 2016. "CashTagNN: Using Sentiment of Tweets with CashTags to Predict Stock Market Prices," in *2016 11th International Conference on Intelligent Systems: Theories and Applications (SITA)*, IEEE, pp. 1–4.

Sahni, T., Chandak, C., Chedeti, N. R., and Singh, M. 2017. "Efficient Twitter Sentiment Classification Using Subjective Distant Supervision," in *2017 9th International Conference on Communication Systems and Networks (COMSNETS)*, IEEE, pp. 548–553.

Shores, K. B., He, Y., Swanenburg, K. L., Kraut, R., and Riedl, J. 2014. "The Identification of Deviance and Its Impact on Retention in a Multiplayer Game," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, New York, NY, USA: ACM, pp. 1356–1365.

Singh, V., Kathuria, S., and Johri, A. 2012. *Newcomer Integration and Learning in OSS Technical Support Communities*, ACM Press, p. 215.

Sinha, V., Lazar, A., and Sharif, B. 2016. "Analyzing Developer Sentiment in Commit Logs," in *Proceedings of the 13th International Conference on Mining Software Repositories*, pp. 520–523.

Steinmacher, I., Conte, T., Gerosa, M. A., and Redmiles, D. 2015. "Social Barriers Faced by Newcomers Placing Their First Contribution in Open Source Software Projects," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, New York, NY, USA: ACM, pp. 1379–1392.

Terrell, J., Kofink, A., Middleton, J., Rainear, C., Murphy-Hill, E., Parnin, C., and Stallings, J. 2017. "Gender Differences and Bias in Open Source: Pull Request Acceptance of Women versus Men," *PeerJ Computer Science* (3), p. e111.

Vijayarani, S., and Janani, R. 2016. "Text Mining: Open Source Tokenization Tools-an Analysis," *Advanced Computational Intelligence: An International Journal (ACII)* (3:1), pp. 37–47.

Zhou, M., and Mockus, A. 2012. "What Make Long Term Contributors: Willingness and Opportunity in OSS Community," in *Proceedings of the 34th International Conference on Software Engineering*, IEEE Press, pp. 518–528.