Open Source Community Joining & Management

i290m Open Collaboration and Peer-Production Class[1]

School of Information,
UC Berkeley,
Berkeley, United States

Abstract. Here is the abstract

Keywords: open collaboration, peer-production, community joining, education

1 Introduction

Open collaboration and peer production systems comprise a significant part of infrastructure and content on the Internet. Practicioners within these systems recognize the importance of *community*[1] in developing and sustaining these resources. Paradigmatically, an open source software project is created and maintained by an on-line community that engages each other in productive dialog using communication tools and contributes to a common pool[2], in a process that is sometimes called "private-collective innovation".[3] It has been suggested that the core features of such collaborative communities extend to other communities of practice, such as those surrounding wikis, open data, and citizen science, as well.

Our work is a contribution to the study of these open collaborative communities, with a focus on the process of joining them and on how the communities are managed, or manage themselves. We have been conspicuously reflexive in our approach to this study. The authors of this report are all students and instructors in a course at UC Berkeley's School of Information, entitled "Open Collaboration and Peer Production". Each student has spent the semester participating in an open collaborative community of their choosing and reporting back their observations. We have together developed a survey about our experiences joining these projects, and the projects' organization and demographics. By administering this survey to ourselves, we have surfaced data that samples across a wide range of communities and researcher experiences.

[1] Fogel, K. [**?**]

[2] some references to "common pool" resources, from Coye's CMC syllabus

[3] Gulley, N. and Lakhani, K. R. The determinants of individual performance and collective value in Private-Collective software innovation. Social Science Research Network Working Paper Series (2010)

Section 2 will elaborate on the context of how we have written this report, as an exploration of this novel method is one of our main research contributions. Section 3 will survey the background literature which we draw from in our survey design and analysis. Section 4 explains our research methods, including quantitative reporting and survey design. Section 5 presents the results of our empirical work. We discuss implications of our work in Section 6, and conclude.

2 Context

Discovering a Group to Join One of the central objectives of this report is to investigate factors that bring projects and potential contributors together. Prior to successfully joining and contributing to a project, newcomers must become aware of projects that are active and open to (or actively seeking) help, with available opportunities that resonate with the potential contributor's motivation to be part of an open source project. In the survey discussed here, we posed and analyzed questions relating to project outreach, contributor incentives, and the quality of the joining experience in order to gain novel perspective on the relationships between these factors for newcomers to a range of project types.

Citizen Science Intro As scientific data has become more open and accessible, either because of Federal regulations or changes in research culture, new collaboration opportunities were created among research teams. The citizen science movement is an example of crowd-sourced data collection or data processing that allows volunteers to access and collaborate in order to facilitate large-scale discovery. Volunteers, or citizen scientists, collaborate on a project or platform and have a variety of expertise levels. Although forms of citizen science have been seen throughout history, recent permutations of crowd-sourced science was illustrated in projects like SETI [4], where volunteers could donate their personal computer's processing power to help study life in the universe.

Most citizen science projects are hosted through academic organizations like universities, or non-profits focused on a specific research or issue area. Citizen scientists volunteer their time and expertise to projects of interest to them, and are usually not compensated for their research, unlike an academic scientific team would be. Often the crowd-sourced, citizen science projects may only represent a small fraction of an otherwise larger research area. Depending upon the level of community or volunteer engagement, a web-portal is often hosted to facilitate the collection of observations or data processing.

3 Background

Here is the Background

[4] http://www.seti.org/about-us

4 Methods

4.1 Qualitative Reporting

Here is the qualitative reporting method based based on the assignments

4.2 Survey Design

Survey Design Process To complement our qualitative understanding of the *joining process* we decided to design and run a survey. Each of us proposed a survey question and posted it on a Google Form [5]. We then categorized questions by assigning tags : communication, governance, contributions, social networks, joining (to be completed). One or more tags could be assigned to one question and has allowed to draw a bipartite network of relationships between questions (c.f. Figure, to be completed).

Once the first survey design was completed each of us took it. We debriefed the result in class and found several flaws that we described in a separate paragraph. However, the advantage of designing a survey in a collective way helped ensure that most questions would be relevant to most of us.

This is a starting point of course and in the future **this survey could be iteratively improved**

Survey Design The survey was designed *in collaboration* among the class members, and it reflected the course readings and our experiences in open source participation. The questions were iterated and categorized in the class and an effort was made to ensure their good formating and reduce potential replications; questions used in the survey are presented in Table 1. As the survey was intended for our class use only, we did not pilot it before collecting responses. Our final sample size was 24 responses, nn % of the class **TODO: FIXME**. Our idea is that these questions can be developed further in follow up questions based on our experiences, discussed below in detail.

Reflections on the survey design After reviewing the results in the class, the survey was found inaquent in many ways. First and foremost, it was biased towards software development, where as some participants were involved in other types of projects. Also exact questions were found difficult, such as the licencing question (#nn), speaking of licence families and missing major licences such as Apache licence, and number of female contributors (#nn) found hard to answer. I[6] attribute these problems to the rather cauothic process of contributions and not enough quality control in the end, many problems discussed also in the previous efforts of academic collaboration [**?**]. I suggest that in the future, instead of circa 20 participants editing the survey there should be a (benefictous) dictator chosen to coordinate the collaboration effort.

[5] `https://docs.google.com/forms/d/1KnkSkM3f_QBRQeYfUXpyXspSqavjcClKS2DoWAz1fyE/viewform`

[6] Matti

What is your birth year?
Have you previously contributed to open source? Yes/No
Prior to this project, did you have any experience in the past with open source projects? (Free text)
Which of following best describes your project? Corporation lead open source project Community originating open source project Academic open source project other (text form)
When was the project founded that you're contributing to? In other words, when were the initial contributions made to the project?
How did you first hear about the project you eventually joined? direct human contact(friend, classmate, etc) brought up in class article(blog, online periodical, etc) mailinglist print publication web search other: (text form)
Do you feel that branding matters to an open source project? Yes No I don't know
Did the popularity of the project have an impact on your choosing to participate? Yes No Undecided
Did you consider (but did not attempt to join) any other project before you joined your current one? If so, please identify the project(s), and the reason(s) why you did not choose them.
Did you attempt to join any other project before you joined your current one? If so, please identify the project(s), and the outcome of your join attempt.
Did you plan to join the project prior to this course? I was already a part of the project. I intended to become a collaborator. I found the project after joining the class. other: etc.

Table 1. Questions used in our survey

5 Results

Funding and Response Tone When separating out the questions on funding, who responded and response tone we can see several interesting trends. While the question asked several different ways that people could be paid for their contributions, it can be simplified to a binary attribute of paid / unpaid. In most cases, there were a mix of funding but in this analysis all that mattered was if there was any financial incentives. Additionally the responses to How did the response read formed a binary attribute as well with all participants selecting either Peer or Teacher.

Though there isn't much data to go by, in funded organizations more people reported the response tone to their initial contribution read as being answered as peers rather than as teachers. Further exploration shows that unlike unfunded projects, in paid organizations it was rarely the project founder who responded but rather a senior project member. This might explain the presence of a more peer-to-peer response tone verse a pedagogical response tone.

True Motivations for Open Source Contribution 42. What do you want to get in return for your contribution? My reason for asking about contributor motivation is to figure out if people really want anything in return for volunteering their time. Time is money, and it would be naive to think that

everyone is truly altruistic. Out of 24 responses, a *staggering* 50% want more experience in open source communities. This should be taken with a grain of salt because respondents are class peers. However, 25% are looking for career advancement or "nothing at all", while a few receive satisfaction, skills, societal benefit, and web development experience from contributing. It is apparent that career advancement is a major reason for open source collaboration. Strangely enough, an equal amount of people want nothing at all, and only 2 respondents wanted academic recognition. Our sample size is small and academically biased, so it will be interesting extrapolate from a larger and more diverse sample size.

Design and Previous Experiences In my experiences working on Hypothes.is for this class, a few questions had come to mind while I jumped into my first open source project. The first was trying to understand how difficult it would be for me to find a project that would suit my skills and what I'm looking for, since the only contribution I could really put forth was my design work. So some key quetsions that had arisen primarily involved how important design was to the project and what to expect from them. This tangentially led me to wonder how much experience others in the class had with open collaboration projects, since I felt that many had an easy time jumping into one or knowing what they wanted to do. I then felt like it was appropriate to ask the class these two pressing questions; "How important do you find design is to your project" and "what was your involvement with open source projects before this class".

I was pleasantly surprised at the results I got. Many agreed that design was important to their project, with "I Agree" holding 61% of the answers. I was more expecting a higher number of "I Somewhat Agree" which holds 35% of the current answers. While I'm glad to know that many projects find design to be an important aspect of it, in hindsight I wonder if I should have rephrased the question to more analyze the role of the design team on the project. I felt like it was a given to understand that design is important, but it's another thing to actually see that the project actively prioritizes design by establishing an open collaborateive design team.

The results for whether or not people had previous experience in open source projects did not surprise me, as many individuals seemed like they are generally active in this space. If anything I was slightly surprised by the number of "No" answers.

The Difference between Experienced and Newcomers in Open Source To start my analysis, I divided the total number of responses into two categories based on how they answered the question : "Prior to this project, did you have any experience in the past with open source projects?" As a newcomer to open source communities myself, I wanted to analyze to **see if there were any discernable differences or interesting trends between newcomers and "veterans" of open source communities.** I started with the responses for how long did the individual lurk before making an attempt at reaching out to the community. We can see that in contrast to the "No" group, the "Yes" group

has 2/11 or 18% of people who didn't even lurk for a single day. On the "No" side, we can see that not a single individual spent 0 days lurking and all of them had to spend at least 1-3 days lurking prior to joining. This could indicate a stronger willingness to immediately dive in to a new Open Source community by the people who have had prior experience with them and a heightened shyness exhibited by the group with no experience with them before. The distribution for the YES group is very disjoint when compared to the No group. For 71%(5/7) of the Yes Group, it took them either more than a month or immediately to enter the OS community. This could indicate either *an extreme willingness or elongated hesitation that is not seen in the No Group* since there are no individuals that took 0 days and not as much of a frequency(20% vs 43%) of individuals that took more than 29+ days. Statistically, *the distribution of the "Yes" group is much more bimodal* with the peaks being at 0 days and 29+ days compared to the No group. Another interesting division between the two groups was their responses for the question "what do I plan on getting out of this contribution" For the Yes Group, the majority of them wanted not only experience in OS software but some sort of other reward like academic publication, career advancement, etc. However, this is decidedly false for the No Group where the majority wanted only experience in OS software or even nothing at all. This could be from the fact that because the Yes Group has contributed to Open Source communities before, they would need an extra incentive to join a new community while the No Group is simply excited to be part of this endeavor that they have never been on before. I think with further data, we could see if the requirement or desire for external incentives would increase linearly with the number of OS communities one would participate in.

Lastly, I thought it would be interesting to test if the two different groups would have differences in the roles that they play within their respective open source communities. I found that the number of roles a particular user from the Yes Group is significantly more than the number of roles a user from the No Group would take. This is to be expected as it is the No Group's first time. Other than that difference, the No Group had less developers and less documenters than the Yes Group but the same number of UX/Designers, Testers, and Analysts as the Yes Group.

From lurking to first contact Most of the participants answered that they **have been lurking** prior to contacting the project, although the days of lurking are widely spread among the projects. *21%* answered that the question is not applicable for them and their project. About half of the people used mail to get into contact with the community (either through a private mail or to the mailing list). Social media was only used by one participant [7]. This suggests that mail is still the main form of communication, at least for initial communication.

[7] Fogel pointed out that social networks might have an impact on OS, but the impact is still not well understood.

Personal Connections To understand whether open source collaborators actually try to know each other on a personal level, we asked them if they knew any of the collaborators to the project personally. Out of the 23 responses, 12 said that they knew other collaborators personally. Furthermore, we asked the people who knew other collaborators if they ever discussed their contributions with the others. 10 out of 12 people replied in the affirmative.

A good followup question would have been if this personal connection helped them in any way with their contribution, but it was too late to add it

Problems faced by new contributors to open source projects 25% of the people reported that lack of proper documentation is the biggest problem they faced while contributing to open source projects. This makes participation much more time consuming because many questions of mine need to be answered, and then there's a time delay. If there was standard, proper, detailed documentation, it would be a much easier process of contributing. One of the reasons could be that many open source projects have multiple platforms for members to communicate with each other - forums, the wiki, the mailing lists, and the issue tracker. However, all of these are separate systems with their own login info and digging into these huge mines and getting required information seems to be a major obstacle. If all the important information like "Where to start and How to contribute to the project" and "trouble shooting areas" from these systems had been put in single place, the newbie would have found it easier to take his first steps in the journey of open source contribution. Indeed, some of the members of this class are contributing in making documentation more comprehensive and organized.

Some members contributing to technical open source projects reported that they had troubles with installing and running the software on their machines. In spite of the availability of the documentation, they could not find a solution to their problems. The responses to their mails are very cryptic and are mostly stuck with solving the problem on their own, resulting in wastage of time. Similar to this issue is the high complexity of the code base which required lot of time to even understand the project. Because of this it has been hard to get up to speed and make a contribution to a current need or issue in the project.

Other major problems which members reported is the lack of sense of community in the open source projects they are trying to contribute. Some complained that they did not get any response or favorable warm response if at all, to their introduction mail in the mailing list. This led to a sense of alienation and fear of contribution in the newbies.

Regular Community Meetings *Are regular community meetings related in some way to project governance and the ease or difficulty of a joining script?*

Looking into the data from our survey and experience as a class. It seems that regularly held meetings, do help with good project governance. In our experience from class, it seems that regular meetings play a significant role in driving project progress. This may be a result of the general environment, that

the course demographic is specifically busy graduate students. In the survey results, it will be interesting to see if respondents who claimed their community holds regular meetings, found the project governance and the joining script to be easier, simpler, or more user friendly. From my own experience with the IPython community, regularly scheduled meetings, whether in person or remote, seemed to be an important tool for decision making for medium and long term project goals. For example, the community holds a meeting along with each release, presently every six months, for core developers. The main purpose of the meeting is to create a roadmap for the next release, a seemly important step for making progress in the project.

FAQs and Onboarding Documentation Around 50% of our projects have no actual FAQ or onboarding document, so an opportunity to help newcomers likely exists. Since only 8% have an initial contact designated, there may be varying org structures, or a deliberate choice to have the entry point be distributed and in flux (according to needs). I do wonder if there could be some voluntary assignments. However, since members probably have forum or other contact points, presumably, newbies can take initiative and contact them.

I was exploring the qualitative potential of onboarding practices- text analysis may help (as well as some follow-up questions. Perhaps just raising the issue itself would capture attention of some project people.). Perhaps there is a connection between FAQ/onboarding info available and the org structure/governance of the actual projects.(Of course, the product and/or nature of the project is also relevant). My tentative summary:

Hypothesis: Development-centric (may be sufficient for their needs) Peerlibrary: Ways to contribute are listed, described (Nice!). Hmm. Since Mitar wrote it, it may not be a coincidence that this is one of the best joining guides. Mozilla PDF: Mentions all ideas are open (also, mainly bug, feature, and development-oriented) Courtlistener: Great "ways to help" page Facebook React: Developer and bug-centric, apologizes that easy flow for contribution is still being smoothed out (as well as transparency!) Chromium: Developer-centric Civic-CRM: no FAQ, but does express and embrace a "just dive in" culture, which may be sufficient Geonode: The about us page is a bit FAQ-like, with some info on getting started, contact, etc. OTHER PROJECTS: No FAQ/onboarding document (yet?)

Exploring Alternative Projects to Join A significant portion (40%) of the class had a clear idea at the beginning of the semester which project they wanted to join, and did not consider alternative projects. The majority of the class, however, spent some time investigating other open-source projects before settling on their selected project. The alternatives were rejected for a variety of reasons. Some projects were seen as having enough contributors; others were seen as not open to outsiders; **(add more reasons)**. Three students actively sought to join a project but were frustrated in their attempts, due to a lack of response

from the community, or more generally a outside appearance of *closeness* of the community.[8]

Funding Models Within a sample size of 24, a high degree of variance in the types of funding models of open source projects has been observed from analysis of the survey results. 22% of the responses indicate that funding was received from a for-profit source such as through venture capital investment (7%) or a direct funding from a for-profit corporation (15%). On the contrary, a large subset of the surveyed indicated that their funding came from non-profit sources. 24% of the respondents indicated their project was backed by academic or research grants, 22% responded that the funding came from donation sources and 17% responded that the project was financially backed by the members of the core team. The remaining 15% of the responses indicated that the projects were funded through a crowdfunding campaign such as Kickstarter or Indiegogo.

6 Discussion

Here is the Discussion

The Importance of Personal Touch The questions **How did you first hear about the project you eventually joined?** and **Which technology/tool did you use for initial communication?** seemed to correlate to a specific place in my mind. Both of the most popular responses to these questions seem to indicate that people were biased towards direct human contact in both choosing and making first contact with their project. It was suggested in class that this is because we are in a unique setting in that we have direct access to many open collaboration projects and the people working on them. I am curious if this analysis is correct, or if instead this would be true of most open collaboration projects. Are most contributors to open collaboration projects people who have some direct connection with the project, or is our experience with these results simply unique because of our specific situation?

Commentary on the Process of Writing This Report "Commentary on the process of writing this report: I'm lost. I can't see what others are writing and I have no perspective on how this paragraph will fit into the larger whole. Will it just fit? Will it be meaningful? Will it exist in a flow of thought? All of this makes me wonder how far we need to come with collaborative writing and coding tools. There needs to be a better vantage point of what others are doing and what needs to be done. GitHub provides some of this, but it's not in real time. Google docs is in real time, but it's uncontrolled chaos if 30 people are writing. What would a system in the middle look like?"

[8] The three projects are Bitcoin, Chromium, and Mozilla Open Badges.

7 Conclusion

Here is the Conclusion