Why do projects join the Apache Software Foundation?

Nan Yang*
Eindhoven University of Technology
n.yang1@tue.nl

Isabella Ferreira*
Polytechnique Montréal
isabella.ferreira@polymtl.ca

Alexander Serebrenik
Eindhoven University of Technology
a.serebrenik@tue.nl

Bram Adams
Queen's University
bram.adams@queensu.ca

ABSTRACT

While numerous open source projects operate on their own, others decide to join well-established foundations such as the Apache Software Foundation (ASF) and the Eclipse Foundation. Although many studies have investigated the motivations of individuals and companies contributing to open source, it remains unknown why projects decide to join software foundations. In this paper, we study the motivators behind the projects' decision to join the ASF, the geographical and organizational characteristics of these projects, and the differences between projects in terms of their motivations. To this aim, we analyzed 292 proposals submitted to ASF, and we found that there is an increasing number of company-based and Asia-based projects joining the ASF in recent years. Furthermore, we found that more than half of the projects are motivated by the desire to foster their community, strengthen the outcome of the project, increase interaction with other communities, and boost technical development. Our work shed some light on projects' expectations from the ASF. Having this understanding can help foundations to identify ways of supporting newly joined projects, while the prospective joiners can better decide on whether ASF is the right place for them by checking the alignment of their motivations and motivations of projects that have joined in the past.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**; • **Software and its engineering** → **Open source model**.

KEYWORDS

ASF, software foundation, motivations, open source, incubator

ACM Reference Format:
Nan Yang, Isabella Ferreira, Alexander Serebrenik, and Bram Adams. 2022. Why do projects join the Apache Software Foundation?. In *Software Engineering in Society (ICSE-SEIS'22), May 21–29, 2022, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3510458.3513006

*Both authors contributed equally to this work.

ICSE-SEIS'22, May 21–29, 2022, Pittsburgh, PA, USA

ACM ISBN 978-1-4503-9227-3/22/05.
https://doi.org/10.1145/3510458.3513006

LAY ABSTRACT

Open Source Software (OSS) is free to be used and modified by anyone in the world for any purpose. Nowadays, OSS is widely used in all kinds of products and hence plays an important role in our daily life. For example, as the most used OSS, the Linux kernel runs on 85% of all smartphones. To create a better environment for OSS developers to collaborate and for OSS projects to grow, many software foundations have been established, such as the Apache Software Foundation (ASF). We have seen that many OSS projects joined the ASF over the years. To join the ASF, OSS projects have to donate all their assets to the foundation, adhering to the foundation's rules and culture. In this paper, we study why these projects decide to join the ASF to understand their expectations from such foundations so that the expected technical and non-technical support can be provided to them. We identified their motivations by analyzing 292 proposals that these projects submitted when applying for joining the ASF. We observed that more than half of the projects try to join the ASF with motivations related to fostering a community, strengthening the project's outcome, increasing interactions with other OSS projects in the ASF, and boosting technical development.

1 INTRODUCTION

Open Source Software (OSS) plays a crucial role in our society. Because of its easy accessibility, it is attractive to many users, including government, individual users, schools, or business units. The importance and popularity of open source projects have motivated the creation of software foundations such as Apache Software Foundation (ASF), Eclipse Foundation, and Python Software Foundation [9], whose goals include providing a seal of quality, improving the integration of related OSS projects into an entire ecosystem, and providing support to newly open-sourced projects. To enable the latter, many software foundations, such as OpenStack Foundation and ASF Foundation, provide an incubation program that fosters the development of projects through governance, while minimizing the cost of potential failure.

Many studies have investigated the motivations of individuals [16, 34, 37, 38] and companies [5, 8] contributing to open source software. It has been identified that individuals contribute to OSS projects for career, kinship, fun, and learning [34, 38], while companies contribute more for economic and technological reasons [7]. However, it remains unclear about what motivates an OSS project, as a joint force of individuals or/and companies, to join software foundations instead of operating on their own. For example, joining a foundation requires donating the code to the foundation, adhering

to the foundation's rules and culture, and contributing towards the foundation's ecosystem of projects.

Studying the motivators behind the participation could help organizations wishing to open-source their project to better decide whether joining a foundation would be beneficial for them and to understand whether their motivations are commonly shared by other projects. Conversely, software foundations would understand what projects expect from them and how the foundations could better provide the environment and assistance for projects to grow. In addition, the insights into motivators can help researchers investigate how software foundations help projects achieve their goals and how foundations might impact OSS development in general.

To this aim, we conduct an explanatory case study on the phenomenon of projects wishing to join the ASF, the popular home of more than 200 OSS projects. Projects that wish to donate their code to the ASF must go through an incubation phase where incoming projects adopt the Apache style of governance, infrastructure, and operation. Projects that successfully graduate from the ASF incubator (ASFI) can then develop further as ASF projects. To enter the ASFI, projects must apply for admission with their proposals, which are then reviewed by the Project Management Committee (PMC) to decide about acceptance.

We are interested in understanding what kind of projects have entered the ASFI. Zhang *et al.* investigated the participation of companies in the OpenStack Foundation and found that developers from companies made far more contributions than volunteers [41]. Takhteyev *et al.* studied the geographical aspect of OSS projects on GitHub, and identified that developers are concentrated primarily in America and Europe [31]. A recent study by Wachs *et al.* shows that more and more developers from Asia and Latin America have joined the OSS development in the recent 10 years [35]. The geographic and organizational composition may be of interest to projects that consider joining the ASFI and can inform researchers about the diversity of such ecosystems. To reach our goal, we studied the projects and the drivers behind their decision to join the ASF. For that, we collected the submitted proposals of 292 projects accepted to the ASFI, and we aim at addressing the following questions:

RQ1: What are the primary organizational and geographical characteristics of the incubatees of the ASF? By conducting a qualitative analysis on the collected proposals, we identified the organizational and continent information of projects. We found that an increasing number of company-based projects have been accepted to the ASFI in recent years and that the number of Asia-based projects that entered the ASFI has shown a clear upward trend since 2016.

RQ2: What motivates projects to join the ASF? We aimed at identifying why projects join the ASF. We found 28 motivations for participation in the ASF, which are grouped into 10 high-level categories. We observed that more than half of the projects join the ASF with motivations related to fostering a community, strengthening the outcome of projects, increasing interactions with other communities, and boosting technical development.

RQ3: To what extent do projects have different motivations? We assessed whether the motivations vary between company and other projects, and the co-occurrence of projects' motivations. We found that, proportionally, non-company projects are more motivated to join the ASF due to re-organization, assistance, and external stimulation than company projects are. Projects that want to build their community typically also want to increase their user base.

As the main contribution of this study, the derived motivations of project participation in the ASF shed some light on what projects expect from the ASF, providing implications for foundations that would like to improve their support for OSS projects and for projects that consider joining the ASF.

2 THE APACHE INCUBATOR (ASFI)

The Apache incubator (ASFI), initiated in 2002, is the entry path into the ASF for projects that want to be supported and encouraged by the foundation's efforts. In general, the ASFI is responsible for filtering proposals about the creation of new projects, providing the necessary infrastructure to create new projects, supervising and mentoring the incubated projects, evaluating the maturity of incubated projects, and deciding whether projects should be promoted to an official Apache project or not [3].

The main goal of the ASFI is to help incoming projects to adopt the Apache style of governance and operation (also called The Apache Way[1]) and to guide them through the ASF services so that they can become top-level ASF projects (TLPs). The life cycle of projects in the ASFI is shown in Figure 1. In particular, to achieve the aforementioned goals, the incoming project needs to go through a series of steps starting with finding incubation mentors from the ASF and preparing an incubation proposal.[2] Once the proposal is ready, the incubation mentors send it to an Apache mailing list for discussion. The Project Management Committee (PMC) then votes on the proposal and if it is accepted, the project enters into the *incubation* phase until *graduation* or *retirement* occur.

Graduation indicates that the project has successfully finished the incubation phase and can continue as an ASF project. Retirement occurs when the PMC decides to close down the project, removing it from the ASF ecosystem. Typically, projects are retired because of the lack of activity.[3] Retired projects no longer have their activities within the ASF and may further develop outside the ASF. Retirement may occur during the incubation phase (also called *incubating-retired*) or after the graduation (*graduated-retired*). In contrast, the projects that are graduated can further continue their development in the ASF (*graduated-continuing*). Hence, projects that were accepted to the ASFI at any particular moment in time can have the status of *incubating*, *graduated-continuing*, *graduated-retired*, or *incubating-retired*.

3 STUDY DESIGN

This section discusses the case study approach used by this paper to investigate the characteristics of the ASF's incubatees and what motivates projects to join the foundation.

3.1 Data collection

The collection of projects' proposals consists of three steps. First, we identify all the projects that have been accepted to the ASFI

[1] http://apache.org/theapacheway/
[2] https://cwiki.apache.org/confluence/display/INCUBATOR/New+Podling+Proposal
[3] https://incubator.apache.org/projects/

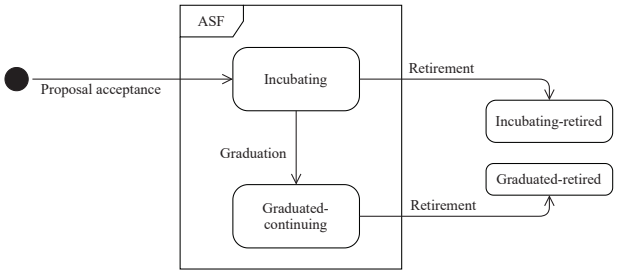

Figure 1: Life cycle of projects in the ASFI.

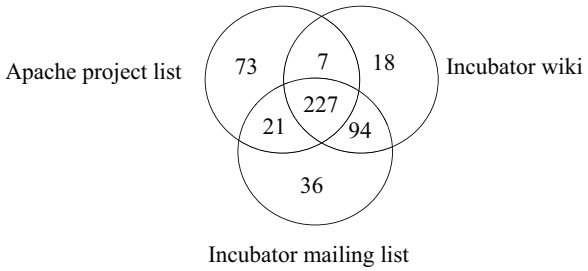

Figure 2: Venn diagram showing the relations between extracted projects from three different data sources.

and their current status (see Figure 1) from the list provided by the ASF [1]. For each project, we record its full name, status, and dates of starting the incubation, graduation (if applicable), and retirement (if applicable). To identify projects' status, we consult the list of accepted projects provided by the ASF [1] and the projects' website[4] to distinguish between *graduated-continuing* and *graduated-retired*.

Second, we collect all the available proposals from the Incubator wiki and the Incubator mailing list, both accepted and rejected ones. Incubator wiki[5] is an archive of incubation proposals. We collected the proposals from the Incubator wiki in December 2019. Since it is not mandatory for a project to upload its proposal to the Incubator wiki (even though it is recommended by the ASF[6]), the Incubator wiki might not contain all the proposals. Therefore, to increase the completeness of our dataset, we further collect the proposals from the Incubator mailing list, where proposals might be discussed by the ASF PCM members. To collect proposals from the mailing list, we download all mbox files from October 2002 to January 2020 using the Python mailbox API. We use "incubator" and "proposal" as keywords to search in the emails' subject.

Third, we merge the Apache project list with the two lists of project proposals collected from the Incubator wiki and the Incubator mailing list. In total, there are 476 projects. 346 of them are listed in the Incubator wiki, 328 on the Apache project list, and 378 mentioned in the Incubator mailing list. 227 projects are found in all these three sources and 7 projects are found on both the Apache project list and the Incubator wiki. Figure 2 shows the relations of these three lists.

For 234 (227 + 7) projects, proposals and status are available. 94 projects are found in both the Incubator wiki and the Incubator mailing list but not the Apache project list, therefore, the information about their status is missing. 73 projects are identified only on the Apache project list and their proposals are missing. Furthermore, we noticed that some projects have been renamed during or after incubation, resulting in inconsistency between their names on the Incubator wiki and on the Apache project list. Based on these observations, we augment the data using the following strategies:

- To identify the renamed projects, we first manually inspect the description of projects on the Apache project list: *e.g.,* one project proposal referred to *Optiq* but the project was later renamed to *Calcite*. Next, we search for the keywords "rename", "renamed", and "renaming" in the emails that we collected from the Incubator mailing list. By applying these two steps, we identified 74 projects that have been renamed.
- To obtain the missing proposals, we search for the keywords "Apache Project *[Project Name] Proposal*" on Google, which allowed us to find proposals that are stored elsewhere (*e.g.,* the project's website). In this way, we were able to identify proposals for additional 38 projects.
- To identify the missing projects' status, we search for the keyword "status" in the emails that we collected from the Incubator mailing list and manually examine the returned emails. However, we did not find any emails disclosing the status of these projects.

After resolving the renaming cases, we obtained 402 projects. Among them, 292 projects have both status and proposal available, they were therefore used for the analysis of our RQs. 51 projects have the status of *incubating-retired*, 46 projects of *incubating*, 31 projects of *graduated-retired* and 164 projects of *graduated-continuing*. The dataset contains the projects that entered the ASFI at the earliest on 2003-03-15 and at the latest on 2019-12-09.

3.2 Qualitative analysis on projects' proposals

To answer our research questions, four authors conducted an inductive coding [30, 32], which is an approach that allows the research findings to emerge from the interpretation of the raw data, on all proposals found in Section 3.1. We describe each step below. All the data of our qualitative analysis are hosted online[7] for replication or third party reuse.

Creating a preliminary list of codes. To establish a coding protocol and to create a preliminary list of codes, we randomly selected 20 proposals. Each coder was responsible for individually reading each proposal of the sample, identifying the motivation of why the project joined the ASF and attributing a code to each identified motivation. For each code, the person added the code and its definition to their personal codebook. Then, the four coders met online and discussed their codebooks by merging or splitting codes into more codes, therefore, generating a unified list of predefined codes.

Next, each coder relabeled their proposals according to the predefined list. We then created an agreement table, i.e., a table with proposals in the rows and codes from the predefined list of codes

[4] *e.g.,* project *ACE*, http://ace.apache.org/
[5] Incubator wiki: https://cwiki.apache.org/confluence/display/INCUBATOR/Proposals
[6] https://incubator.apache.org/guides/proposal.html
[7] https://doi.org/10.6084/m9.figshare.19105193.v1

in the columns. We added the initials of each coder for each proposal/code according to what each person individually coded. For the codes that were assigned to a proposal by less than three people, we discussed and solved the disagreements together in an online meeting. In this step, more codes could be created or merged, leading to a codebook of 17 codes.

Conducting inductive coding on all proposals. After defining the coding protocol and the codebook with a preliminary list of codes, we coded all remaining proposals (excluding the 20 proposals used to create the preliminary list of codes). To avoid bias and misidentification of motivations, each proposal was coded by a pair of coders. We switch the pairs by doing all possible combinations of two coders out of four, resulting in six different pairs. As a result, five pairs coded 46 proposals, and one pair coded 43 proposals.

To support the manual coding, we used a web application called Labelling Machine [8]. This tool allowed us to select the text fragments that were related to motivations to join the ASF and attribute a code to each fragment. New codes could be created. All four coders had a common codebook, and every time a new code was created, the person added the code and its definition to the unified codebook. After everyone finished coding their proposals, we discussed the updated unified codebook together and merged similar codes. As a result, we identified 28 codes in this step. Additionally, 2,683 text fragments were coded by the four coders across the 292 proposals.

Solving coding disagreements. After everyone finished coding their proposals, we solved the disagreements in two parts. First, each coder checked offline and individually the fragments that only the other person of the pair of coders coded. The person could either accept the labeled fragment or leave it to be discussed online with the other coder part of the pair. Second, we had several online sessions to resolve the remaining conflicts. That is, the pair responsible for each proposal discussed each fragment that either both coders classified with different codes or one of the coders could not decide whether the coding was correct in the first step. In case both coders responsible for a proposal could not agree, a third person stepped in to solve the remaining conflicts. As a result, 2,137 (out of 2,683) fragments were considered relevant in this step. Among them, 474 fragments were obtained from the automatic resolution because (i) two fragments were identical and were coded with the same code and (ii) one fragment was part of another fragment and they were coded with the same code; 1,062 fragments were obtained from the offline resolution by one of the coders of the pair; 592 fragments were obtained from online discussion by the two coders; 9 fragments were obtained from the resolution by a third person.

Validating the inductive coding. To make sure that we have consistently coded similar sentences with the same code, we applied the Jaro-Winkler text similarity approach [17] between the coded fragments. If two fragments have a similarity score of at least 0.90 and different codes were assigned to such fragments, we then checked them manually. For that, each one of the four coders went through the fragments offline and individually and tagged if we should leave the code as it is, add a new code, or discuss the code with the other coders. Afterwards, the four coders had an online discussion to decide the final code of each fragment.

Furthermore, we check if we missed labeling fragments from the projects' proposals. For that, we checked for codes with low frequency, i.e., codes that were assigned to less than ten fragments. In total, we found 15 codes that had low frequency and we verified if there were other fragments in the proposals that we missed labeling with such codes. For that, we manually analyzed all fragments coded with each one of the 15 infrequent codes and we extracted a list of keywords that commonly appeared in such fragments. For example, one of the infrequent codes was "*Co-dependency with other Apache projects*". After manually analyzing the fragments coded with such code, we observed that the keywords "*aligned with*", "*alignment with*", "*co-dependence*", "*depend on*" were often present in the coded fragments. Then, we checked if such keywords appear in the proposals' body. If so, we extracted the fragment that has such a keyword and assigned the respective code to the respective fragment. Afterwards, the four coders manually analyzed whether the newly extracted fragments corresponded to motivations to join the ASF and if the code was correctly assigned to that fragment. As a result, we found 62 fragments related to motivations in this step. Hence, we have 2,199 (2,137 + 62) text fragments in total.

Conducting axial coding. With the final version of the codebook with 28 codes, we conducted axial coding [36], which is a technique used to synthesize and organize data into coherent hierarchically structured categories of data, by grouping the codes into themes. In total, the codes were grouped into 10 different themes.

Extracting projects' organization and continent. To answer our RQs, we also extract from the proposals if a project was donated by companies, universities/research labs, universities/companies, or private individuals. For that, two of the authors each went over half of the proposals to extract the name of the donor in case the donor was a company and/or a university/research lab. Then, each one of the authors extracting the information exchanged the extracted information (*i.e.,* author1 validated the information extracted by author2 and vice-versa) to guarantee that the extracted information is accurate. For projects donated by companies and/or universities/research labs, we search for the name of the institution on Google to identify the country of the companies' main office. After identifying the countries, we group the projects by continent. Then, we cross-validated the extracted information between the two coders. We excluded one project from our analysis in which it was not possible to find information about the donor. Additionally, two projects were donated by two companies on different continents. In that case, we consider both continents as "America & Asia" and "America & Europe".

3.3 Association rule mining

Each proposal may mention multiple motivations. Identifying the co-occurrence of the motivations can help us to better understand the incubatees' expectations from the ASF. To this aim, we applied association rule mining [4] to the motivations derived from proposals. An association rule is a rule of the form $X \Rightarrow Y$, where X and Y are two disjoint sets of items (itemsets). The rule indicates that if a transaction contains all items in X, then it is very likely that items

[8] https://github.com/emadpres/labeling-machine

in Y are also shown in the transaction. We consider each proposal as a transaction, and the motivations derived from it as items. We applied a statistical method that identifies significant associations of rules based on Fisher's exact test [19] and we adjusted the p-value using Benjamini-Yekutieli [6]. The rules that have an adjusted p-value smaller than 0.05 are seen as statistically significant.

4 RESULTS

In this section, we present the results for our research questions.

4.1 Projects' organizational and geographical characteristics (RQ1)

The total number of projects that have been accepted to the ASF grew linearly over time. Figure 3 shows the cumulative number of projects from different types of organizations that have been accepted to the ASFI. We found that 175 projects (59.9%) were donated by companies, and 117 projects (40.1%) were donated by non-companies. We observe a slowdown in the number of individual projects entering the ASFI from 2012 on, while the company-based projects continued growing in the same rapid, linear fashion. In particular, the number of company-based and individual-based projects grew at the same speed before 2012. Afterwards, only 24 individual-based projects have been accepted, compared to 106 company-based projects. In addition, university-research-based projects increase only slowly in number over time. There is only one university-company-based project which was accepted in 2016.

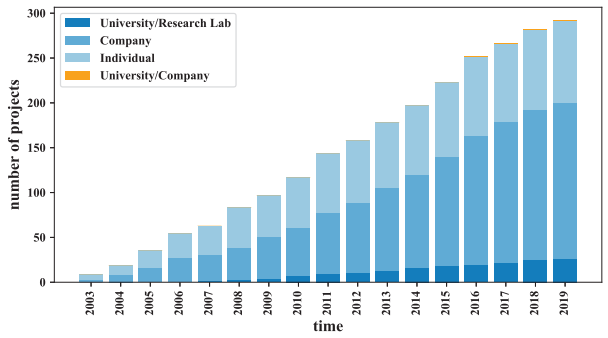

Figure 3: Number of projects from different types of organizations that have been accepted to the ASFI.

One possible reason for the continuously growing participation from industry compared to individual projects is that the ASF has gradually built a good reputation in the industry, attracting more companies to submit their proposals. Another reason could be that projects with industrial use cases are favored by the ASF, which increases the acceptance rate of company-based projects.

America-based companies have been incubating since 2003 and are still the dominant type of incubators nowadays. Figure 4 shows the company-based and university-based projects that have been accepted into the ASFI and their geographical information. While America-based companies are dominant, we noticed that its growth slightly slowed down since 2017, with fewer than

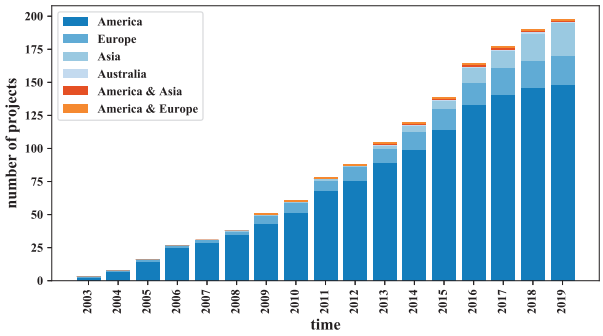

Figure 4: Number of projects from different continents that have been accepted to the ASFI.

5 new projects accepted each year. The number of Europe-based projects is 7 times lower, although the trend for these projects has been steadily growing In contrast, the first Asia-based project was accepted in 2008, and the second one was five years later in 2013. Only from 2016 on, the number of accepted Asia-based projects has increased significantly, growing from 6 projects in 2015 to 11 projects in 2016. We identified three projects, based on Australia (joined in 2013), America & Europe (joined in 2009), and America & Asia (joined in 2013) respectively. No other continents were identified in the proposals.

> **RQ1 summary:** Since 2012, company-based projects have become dominant. Geographically, America-based projects are the most dominant, being 7 times as common as other continents' projects. However, the number of Asia-based projects has shown a clear upward trend since 2016.

4.2 Projects' motivations to join the ASF (RQ2)

We identified 28 low-level categories, which were grouped into 10 high-level categories. Figure 5 shows the results of analyzing the motivations that projects reported in the 292 studied proposals. The gray boxes show the 28 low-level categories obtained from the inductive coding. The number shown in brackets indicates the number of projects that mentioned the corresponding motivation in their proposal. The 10 high-level categories are indicated by the white boxes. The number shown in these boxes indicates the number of projects that mentioned at least one of the low-level categories belonging to the corresponding high-level category. Since a project can state multiple motivations in their proposals, the number in the white box is not equal to the sum of the numbers in the gray boxes.

As shown in Figure 5, **more than half of the projects motivated their ASFI proposal with arguments related to the high-level categories *community, outcome, interaction,* and *technical development***. Among the low-level categories, *community building*, *community diversity*, *development process*, *user base* and *expected collaboration* are the top 5 motivations.

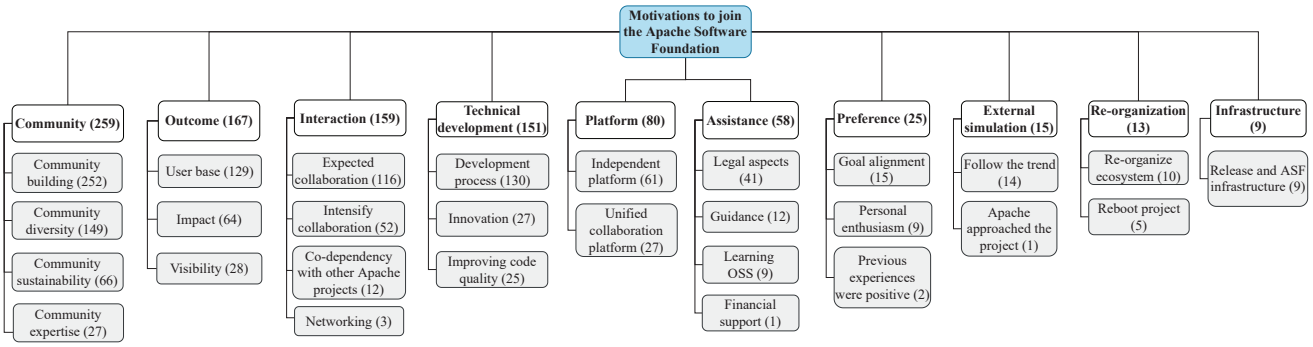

Figure 5: Motivations to join the ASF.

Community. Four low-level motivations related to *community* have been identified. Building a community via Apache incubation is the most frequent motivation ($N = 252$): "*We believe that the Apache brand will help us attract contributors to Stratosphere*" (project `Stratosphere`). 149 projects report that a healthy community is characterized not only by the number of contributors but also by their diversity: "*We hope that growing a diverse community around the project will become a guarantee against the project becoming orphaned*" (project `Crail`). 83 out of these 149 projects further specify the types of diversity that they would like to gain in the proposals. 70 of these projects mention that they want to "*diversify the developer base by involving developers from other companies, organizations and universities*" (project `NetBeans`), trying to reduce "*reliance on a single organization*" (project `Tobago`) and "*the concern of exclusive control [of a single organization]*" (project `Servicecomb`). The 13 other projects (out of 83) motivate that they would like to engage developers from a broader range of domains, which is considered to be useful, e.g., "*to ensure that Falcon is deployable for a broad range of scenarios*" (project `Falcon`). According to 10 projects (out of 83), diversity of expertise is also important: "*Grow the community to establish diversity of background and expertise*"(project `Pirk`). Apart from gaining diversity of technical skills, 7 projects (out of 83) mention that they would like to increase the diversity of geographies, involving developers from different countries or regions. For example, project `OpenOffice` states: "*We wish to diversify this more, especially encouraging greater participation from CJK,*[9] *India and South America*". Only one project mentions that they would like to involve developers of different genders.

We observed that 66 projects were motivated to build a self-sustaining community within Apache. Project `Airavata` states that community-driven development is more sustainable than other drivers such as funding: "*As a successful ASF project (after incubation), we will have created a community led, rather than funding led, environment for the development of our software*". 27 projects also indicate that by joining Apache they hope to gain expertise in certain domains (e.g., graph databases).

[9]China, Japan, Korea

Outcome. 129 projects mentioned in their proposal that the ASF could help them to increase their user base by increasing their software project's adoption: "*Currently, the primary goal is to widen the user and developer base.*" (project `StreamPipes`). Furthermore, 64 projects are motivated by the impact and contribution on the OSS community. For example, the DataSketches project mentioned in their proposal "*We believe that having DataSketches as an Apache project will provide an immediate, worthwhile, and substantial contribution to the open source community, will have a better opportunity to provide a meaningful contribution to both the science and engineering of sketching algorithms*". Finally, 28 projects are motivated by the visibility that the ASFI can bring to them. That means that the ASF can help people to be aware of their project, increase their networking: "*by joining The Apache Software Foundation, we will further boost the recognition and prestige of our product.*" (project `Tinkerpop`).

Interaction. This category groups four low-level motivations related to the interaction that the donated projects desire with other ASF projects. Collaboration is one of the ways to interact with other ASF projects. By incubating inside the ASF, 116 projects would like to establish collaborations with other ASF projects: "*We see the two projects benefiting from each other's experience of implementing similar classes of algorithms and look forward to a fruitful exchange of ideas between the two communities*" (project `Madlib`). 52 projects had already established a collaboration with the ASF projects prior to applying for the ASFI. By being part of Apache, they expected to strengthen the existing cooperation. Three projects are also motivated by the opportunities of networking, increasing their social ties with ASF developers. Finally, sometimes incubation is motivated not only by the ties between developers but also the dependencies between the codebase of projects ($N = 12$): "*During the course of Blur development, a couple of patches have been committed back to the Lucene project...due to the strong relationship...the incubator is a good match for Blur*" (project `Blur`).

Technical development. The low-level motivation of 130 projects is to join the ASF to have an established development process, i.e., follow the Apache way governance model, have a meritocracy-centric model, and a self-governed open-source community. For example, the `CarbonData` project wrote in their proposal "*Our initial goals are to bring CarbonData into the ASF, [...], and foster a*

collaborative development model according to the "Apache Way"." Furthermore, 27 projects are motivated by the innovation that the ASF can bring to their projects: "*We envision accelerating innovation under ASF governance in order to meet the requirements of a wide variety of use cases for in-memory non-volatile and volatile data caching programming.*" (project Mnemonic). Finally, the goal of 25 projects is to improve their code quality by making their software more reliable, robust, fault tolerant, and easy to maintain: "*By also bringing these resources closer to their origin we hope to improve quality, freshness of the content and versatility (e.g. more language options).*" (project Training).

Platform. ASF is also considered by some proposals as an *independent* and *unified collaboration* platform. As an independent platform, it provides a way to formalize the neutrality of a project ($N = 61$): "*One of the reasons for our desire to move to the Apache Foundation is to formalise this volunteering/contribution effort so that it becomes obvious that it is not just the University of Manchester that is contributing to the core code base*" (project Taverna). This neutrality is essential not only to *attract independent developers* (project Milagro), but also to allow the community to make development decisions based on actual needs:"*Vendor influence and/or proprietary implementations may limit the community's ability to choose the hardware and software for use in the datacenter*" (project CloudStack). In addition to the neutrality of ASF as a platform, projects are also motivated by unification of the codebase ($N = 27$): "*This situation with separate codebases in separate source repositories was causing confusion and coordination problems...the project should [be] brought to the Incubator where it can grow in a more controlled and yet less constrained manner*" (project Chemistry).

Assistance. 58 projects join the ASF as it provides different types of *assistance*. 41 projects mentioned that the ASF provides assistance with legal aspects of software development, such as licensing and intellectual property management. For example, the Tinkerpop project wrote in their proposal "*TinkerPop is interested in The Apache Software Foundation for the legal support and protection it can offer our developers. We believe that by joining The Apache Software Foundation, our vendors, users, and contributors will feel more comfortable in terms of legal protection*". Furthermore, 12 projects mentioned that the ASF provides guidance on best software engineering practices since they have experience as a foundation and with experienced committers. For example, the Airavata project wrote in their proposal "*It is our desire, through the Apache Incubator, to take our software engineering efforts to a higher level by learning from the substantial experience of appropriate Apache Committers.*" Some projects also mentioned that the mentorship in the development process during incubation is very valuable to their projects: "*The project currently lacks sufficiently clear direction, leadership, and process; we believe the project will benefit greatly from Incubator mentorship.*" (project ESME).

Nine projects mentioned that the ASF assists projects in learning about open source development. For example, the ESME project mentioned in their proposal that "*moving to the Apache Software Foundation would enable the team to learn more about open source project governance and to correct bad habits such as a tendency to use our private Google group for discussion and a tendency not to record minutes during our team conference calls.*" Finally, one project joins the ASF to have financial support.

Preference. 15 projects joined the ASF because they believe that the ASF has the same goal as their projects: "*The ASF is the natural choice to host the Dubbo project as its goal of encouraging community-driven open source projects fits with our vision for Dubbo.* " (project Dubbo). Nine projects mentioned that they are Apache enthusiasts and they would like to join the ASF because of that: "*We have been enthusiastic users of Apache from the earliest hour (remember JServ anyone?), and feel honored at getting the opportunity to join the club.*" (project Wicket). Finally, five projects would like to join the ASF because their previous experiences were positive: "*We have all had experience working within the Apache community and benefiting from Apache-released software, and believe the Apache community remains the right home for this project.*" (project MRUnit).

External stimulation. 15 projects are motivated to join the ASF because other projects (normally from the same domain) are already part of it: "*The ASF is a natural host for HMS given that it is already the home of Hadoop, Pig, HBase, Cassandra, and other emerging cloud software projects.* " (project HMS). Furthermore, in one case, the ASF approached a project to submit the proposal: "*In fact the LDAPd Group was initially approached by Apache to submit an Incubator proposal.*" (project Directory).

Re-organization. This category groups motivations reported by projects that were part of the ASF before, but like to join incubation again (re-incubation). There are two scenarios where projects would like to re-incubate. First, it happens when a project would like to re-organize their ecosystems ($N = 10$) by merging: "*...combine several existing projects and teams currently working towards more or less the same or overlapping goals*" (project Rave), or splitting: "*better to be a separated project because...separation of concerns is good for project governance*" (project Hivemall). The second scenario of re-incubation is when the projects were terminated by the ASF (i.e., retirement), discontinuing their development within the ASF, and would like to reboot via the incubation, as indicated by five projects.

Infrastructure. Nine proposals report their motivations of donation related to the infrastructure provided by the ASF[10]. The infrastructure includes, but is not limited to, the various machines, mailing lists, version control systems, committer accounts, the distribution mirroring system, issue tracking systems, and technical support for the use of ASF infrastructure. Particularly, five projects elaborate the benefit that they expect to gain from the ASF release infrastructure: "*using the ASF's worldwide mirroring system, we will be able to deliver climate software broadly to the community as we release it, rather than sneaker netting the software around or establishing our own point release infrastructure*" (project Climate).

RQ2 summary: We identified 28 motivations grouped into 10 categories. More than half of the projects motivate their proposals using the categories community, outcome, interaction, and technical development.

[10] https://infra.apache.org/

4.3 Differences in motivations between types of projects (RQ3)

To answer this question, we first analyze if there is a difference in motivations in company projects vs. other projects (*i.e.,* universities/research labs and private individuals), then we assess if the co-occurrence of motivations to join the ASF can help us better understand the incubatees' expectations. For the latter case, we apply association rule mining (see Section 3.3).

Proportionally speaking, we observe that non-company projects are usually more motivated to join the ASF due to its *re-organization* (statistically significant), *assistance* and *external stimulation*, while the other motivations are more common to company projects. Figure 6 shows the distribution of high-level motivations for each type of organization. We present the percentage of projects relative to the total number of projects in each type of organization (y-axis) as well as the absolute number of projects (number on top of each bar). To determine whether the differences are significant, we computed for each high-level motivation the Fisher's Exact Test [19] with adjusted p-value using the Benjamini-Yekutieli method [6] to validate the null hypothesis "the motivation's observed differences in proportions are independent of a project being from a company or non-company". As a result, only the re-organization motivation is statistically significant ($p - value = 0.027$). Our results concerning the motivation *technical development* confirms previous research [8] that mentions that company projects contribute to OSS for the technical development gains, *i.e.,* innovation and improving software quality through bug reporting and testing.

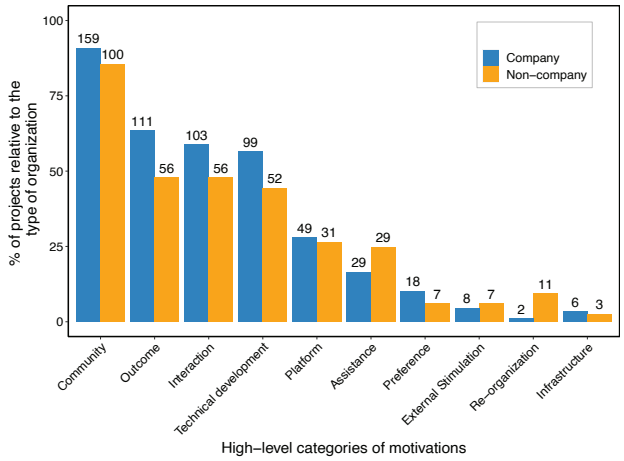

Figure 6: Percentage of projects relative to the type of organization and absolute number of projects from different types of organizations for different high-level categories of motivations to join the ASF.

Concerning the co-occurrence of motivations to join the ASF, **the results indicate that the projects that would like to build a community in the ASF also would like to gain diversity and enlarge their user base.** We found significant results for rules *community building* ⇒ *user base* (p-value = 0.0102) and *community building* ⇒ *community diversity* (p-value = 0.0102).

> **RQ3 summary:** More non-company projects are motivated to join the ASF due to re-organization, assistance, and external stimulation than company projects. Furthermore, projects that are motivated to join the ASF due to community building are also motivated to increase their user base.

5 DISCUSSION

Our results show that **there are more American projects than European and Asian projects joining the ASF.** Even if American projects are joining the ASFI since 2003 and they are dominant compared to other continents, there is a slight slowdown in the number of American projects joining the ASF since 2018. Interestingly, the number of Asian companies are significantly increasing since 2016. This result is expected since previous research shows that most OSS contributors are located in America [31] and that the number of contributors from Asia joining OSS projects is growing over time [35]. We also found that **more company-based projects have been accepted to the ASF than projects from individuals or universities.** That might happen because an increasing number of companies are contributing to open source due to the open innovation approach and cost reduction through opensourcing [5].

Interestingly, **projects are mostly motivated to join the ASF to *build, diversify,* and *sustain* their community**. The tendency is that open source projects with more contributors have higher survival rates [33] and projects with fewer but more centralized committers are more self-sustainable in the ASFI [39]. One of the aspects that projects should consider when building a community is a mentoring program to onboard new developers, since previous research has shown that OSS projects with onboarding programs tend to have high retention rate, productivity, quality, and diversity [15].

We also made some contradictory findings. Although **28 projects mentioned that they would like to join the ASF due to the visibility that the Apache brand can give to them**, the ASFI project proposal template explicitly states that "*an excessive fascination with the Apache brand*" is not a good motivation to join [2]. Similarly, **projects often *expect to collaborate with other projects* by joining the ASF**, yet joining the ASF by itself cannot be expected to increase collaboration. Previous research has found that collaboration is more likely to happen through prior social connections [11], which can impact the projects' survivability [29].

Finally, **although projects might have different motivations for joining the ASF, there might exist external factors that impact the projects' success in the ASF**, making it difficult to quantitatively link a project's success with their initial motivation. Yin *et al.* [39] have forecasted the survivability of projects in the ASFI using a Long Short-Term Memory (LSTMs) model and features involving socio-technical metrics. Then, the authors qualitatively analyzed three projects that had either up or down turning points in their sustainability trajectories.

To assess if projects were able to achieve their stated motivations in practice or if any other factors impacted their success in the ASF, we now contrast the motivations found in our study with the

qualitative analysis of Yin *et al.* for the projects `Commonsrdf`, `Etch`, and `Ariatosca`. The projects `Commonsrdf` and `Etch` are graduated projects that almost failed according to Yin *et al.*'s analysis, and `Ariatosca` is a retired project that almost succeeded.

A success case aligned with the initial motivations. The project `Commonsrdf` joined the ASF with the goal to have a **unified collaboration platform**: "*...we think that in this phase of the project, which focuses on the API design and actively involves the developers of existing toolkits, it is better to have a more focused community and infrastructure.* ", to have an **established development process**, and to **improve the code quality**: "*Part of the motivation for doing the incubator process would therefore be to bring together the existing Commons RDF community in the Apache Way, mature the API, and then gradually prepare the Commons RDF community for working within the larger Apache Commons community*".

Even with such ambitious motivations in mind, the project had a lack of commit and email activity in the first half of the incubation as well as a decreasing trend in the number of active developers [39]. Additionally, in the month with the lowest graduation forecast predicted by Yin *et al.*'s model, the PMC routinely requested the project's monthly report.[11] Lastly, discussions in the mailing list focused on the future of the project, with contributors mentioning that the project had no longer a viable community. However, at some point, a developer mentioned "*there was a small group of 5 core committers to begin with. As of right now, the number is 22. We've actually done pretty well*", after another developer suggested technical directions saying "*...if you actually tried to use this [algorithm] would (i) hurt speed, and (ii) hurt the perception of speed... I'd be inclined to go another step further and add a generic parameter*" [39]. After this, the community became more engaged and increased the development activities. This is an example of a successful case, in which the project graduated after having received technical directions on how to continue the development.

External factors may impact the projects' success. The `Etch` project was motivated to join the ASF by the **impact** that the project could have on the OSS community: "*We expect that as Etch becomes available, it will be seen as a very compelling technology and others will begin to depend upon it*". However, Yin *et al.* observed that there was a reduction in the number of senior developers and commit activity during the incubation process. After many months of inactivity, the project retired and the project mentor mentioned in the mailing list that lack of diversity was the project's main problem since all developers were from the same company [39]. Even if eventually the projects' contributors made a list of objective to make the community thrive again, this example suggests that projects might not be aware of what they need to succeed in before joining the ASFI. Even though `Etch` wanted to increase its impact, community diversity should also have been one of its motivations to join the ASF, and a concrete action point to tackle, even early on during incubation.

Finally, the `AriaTosca` project joined the ASF in order to have an **independent platform**, increase their **user base**, **build their community**, **unify the collaboration platform**, and have an **established development process**. The project retired before achieving its goals because all senior contributors had left the project due to being too busy in their jobs [39]. That was a major problem since even if newcomers wanted to contribute to the project, their changes were never merged because the senior contributors were the ones responsible for reviewing and merging new changes. In this case, the project could not succeed and achieve its goals due to the inability to build a community.

Based on our findings, we suggest the following implications for both researchers and practitioners.

Implications for researchers. This work paves the road for future research on investigating to what extent the motivations identified in this study can be achieved in practice. In other words, to what extent and how does a software foundation like the ASF help projects achieve their expectations? What kind of projects successfully achieved their expectations? How do projects' motivations relate to their success and failure in foundations? What causes the retirement of projects? Future research could investigate which external factors impact the projects' success in the ASF.

Implications for software foundations. In this work, we found what projects that join the ASF expect from the foundation. As a consequence, foundations can create means to help projects achieve their goals, as well as a way to measure if their goals are being achieved in practice. Furthermore, future work could help determine for a given incubation proposal which motivations (e.g., community building) make sense for the candidate project, and might stand a good chance of being achieved. In other words, how realistic are the expectations, given a project's current development process, community, etc.

6 THREATS TO VALIDITY

As with any other empirical study, our study suffers from threats to validity.

Construct validity. To identify the projects' motivations, we coded their incubation proposals. There is a risk that projects do not necessarily express their real motivation in the proposals, hiding their actual intent. Since our study is an explanatory study, we were not able to analyze this threat in depth.

Internal validity. First, our results could be influenced by the completeness of the data since the information and proposal of these projects might be distributed across different data sources. To increase the completeness of data (*i.e.,* name, proposal, and status of projects), we collected data from different sources. We obtained the project list by crawling the Apache website, Incubator wiki, and mailing list. We then resolved 74 renaming cases by manually inspecting project descriptions and checking emails in the mailing list of projects. In addition, for the project that does not have proposals available in the data sources that we mentioned above, we search them on Google to obtain the proposals stored in other archives (*e.g.,* the project's website), which allows us to obtain proposals for 38 projects. Second, the coding we applied to the proposals is an interpretive procedure. To reduce bias and misinterpretation, each proposal was coded by a pair of coders. In case of disagreement between the two coders, we involved a third coder to resolve the conflicts. Third, when identifying the status of the projects, we search keyword "status" in the mailing list of projects. It could be possible that developers use other words when discussing the

[11] https://cwiki.apache.org/confluence/display/INCUBATOR/Reports

project status. Therefore, there is a risk that we might miss some projects which have stated their status on the mailing list.

External validity. We limited our study to the projects of the ASF. It could be possible that OSS projects have a different set of motivations when joining other foundations. We expect that the results obtained from the ASF can be a start point to investigate the effect of the ASF in terms of helping projects achieve their goals.

Conclusion validity. The ASFI provides guidelines on how to create the proposals[12]. These guidelines include examples of successful incubation proposals potentially causing the motivation arguments of these proposals to be imitated by newcomers, increasing popularity of the common motivations akin to the so-called "Matthew effect" [24]. It is hard to estimate to what extent this could have affected our results; while the current proposal preparation guidelines go back to 2019, similar documents might have existed prior to this date.

7 RELATED WORK

In this section, we discuss related work about business and OSS incubation and motivations for joining OSS projects.

7.1 Business and OSS Incubation

The concept of incubation has widely been practiced outside OSS. Sherman *et al.* [28] found that incubated firms have significantly lower rates of failure compared to non-incubated firms. Colombo *et al.* [12] studied 43 new Italian technology firms and observed the effectiveness of attracting high-quality human capital. On the other hand, it has been shown in the same study that the input and output measures of innovative activities are only marginally different between incubated and non-incubated firms. Mas *et al.* [23] found that incubators alone cannot affect the survival of firms.

A large body of studies has been performed to understand the phenomenon of business incubation. Kemp *et al.* [21] studied the advantages and disadvantages perceived by incubatees of an incubator environment and the motivations of locating their firm within an incubator. The primary motivation identified in this study is to reduce the price of office space, rather than the business development assistance provided by the incubator. However, after operating their business within the incubator, the primary advantage was the assistance in the development and growth of their businesses. By conducting a systematic review, Hackett *et al.* [18] concluded that the questions related to the success of incubation and the survivability of incubatees are not sufficiently studied with empirical methods because access to information regarding politically sensitive incubation failures remains challenging.

In the OSS world, the accessibility of project repositories and mailing lists could enable the use of empirical methods to study the success of incubation and the survivability of incubatees. Duenas *et al.* [14] studied the number of committers of incubating projects and their incubation time at the ASF and suspect that the most critical variables that might impact incubation time and increase project risks are the period of project inactivity, the engagement of new committers, and the delivery of an early release with enough technical impact. To facilitate further investigation of sustainability of incubatees, Yin *et al.* [40] provided a longitudinal dataset of the ASF developer coding and communication activities by mining the mailing list of ASF projects. By further collecting commit data from repositories, Yin *et al.* [39] built models of project sustainability. The authors found that the projects with fewer but more centralized committers and those with more but distributed communicators are more likely to sustain longer in the ASFI. However, these studies do not focus on what motivates projects to join the ASF and how the ASF helps projects achieve their goals.

Our study on the phenomenon of incubation in OSS explores the motivation for joining the ASFI by analyzing the ASF publicly available data (*i.e.,* proposals). The derived motivations reflect the goals and expectations of the donated projects, which serves as the starting point towards understanding of the effect the foundation on the development and growth of the donated projects.

7.2 Motivations in OSS development

Many studies have investigated individual developers' motivations of contributing to OSS projects [16, 34, 37, 38]. The derived motivators have been grouped into extrinsic and intrinsic aspects. Extrinsic motivators are external factors given by the organization to an individual, such as career advancement and learning. Developers can also internalize extrinsic motivators in a way that their behavior is perceived as a result of external force [13]. For example, Ye *et al.* [38] theorize that learning is the main drive for individual. OSS development provides a chance of learning to less skilled developers. During this learning process, the contributors develop a shared way of addressing problems. Eventually, less experienced participants can develop useful skills to compete in the labor market. Indeed, according to Hann *et al.* [20], high-ranking OSS contributors were paid high salaries by their employers.

In contrast to the extrinsic motivators, intrinsic motivators are related to the need for satisfying the individual [37]. Several studies [22, 34] have argued that individuals are also motivated by the altruistic nature of helping others or the intention of having fun. Gerosa *et al.* [16] observed a shift of contribution motivations; experienced contributors often shift toward altruism, novices often shift toward career, fun, kinship, and learning. Roberts *et al.* [26] studied how differences in OSS contributors' motivations relate to differences in their participation, and how past performance affects subsequent motivations in OSS communities. Interestingly, the study found that developers' performance and motivations are interrelated; developers with higher status motivations appear to be the more substantive contributors, and developers with higher past-performance rankings tend to have a higher status motivation.

Several studies have focused on the motivators for companies to join OSS development [5, 7, 10, 25, 27]. Andersen *et al.* [5] identified that selling complementary services, building greater innovative capability, and cost reduction through open-sourcing to an external community are three main drivers for companies to contribute to open source. Same as the motivator for individuals, the acquisition of knowledge is also an important driver of companies' contribution. Munir *et al.* [25] observed the value a company gained by knowledge interchange with OSS projects in an open innovation process. Bonaccorsi *et al.* [7] compared differences between the set of motivations of individuals and those of companies and found

[12] https://cwiki.apache.org/confluence/display/INCUBATOR/New+Podling+Proposal

that while joining OSS development, companies emphasize more on the economic and technological reasons rather than many social motivations that are valued by individual programmers.

8 CONCLUSION

In this study, we investigated the phenomenon of OSS projects joining the Apache Software Foundation (ASF). We collected the incubation proposal of 292 projects. Specifically, we studied the geographical and organizational characteristics of these projects, the motivations for joining the foundation, and the difference between projects in terms of their motivations. We observed an increasing number of company-based and Asia-based projects joining the incubator in recent years. As the main contribution, we identified 28 motivations for joining the ASF. These motivations are grouped into 10 high-level categories. Among them, fostering a community, strengthening the outcome of projects, increasing interactions with other communities, and boosting technical development are the most frequent ones. In addition, we also observed that non-company projects are more motivated to join the ASF because of re-organization, assistance, and external stimulation than company-based projects. The results of this study provide insights into what the analyzed projects expect from the ASF, which can be useful for software foundations that would like to improve their support for OSS projects and for projects which consider joining them.

REFERENCES

[1] [n. d.]. All Incubator Projects By Status. https://incubator.apache.org/projects/ Accessed on 2021.10.01.

[2] [n. d.]. New podling proposal. https://cwiki-test.apache.org/confluence/display/INCUBATOR/New+Podling+Proposal?desktop=true&macroName=markdown Accessed on 2021.10.01.

[3] [n. d.]. What is the apache software foundation? http://www.apache.org/foundation/how-it-works.html Accessed on 2021.10.01.

[4] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. 1993. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*. 207–216.

[5] Morten Andersen-Gott, Gheorghita Ghinea, and Bendik Bygstad. 2012. Why do commercial companies contribute to open source software? *International journal of information management* 32, 2 (2012), 106–117.

[6] Yoav Benjamini and Daniel Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics* (2001), 1165–1188.

[7] Andrea Bonaccorsi and Cristina Rossi. 2006. Comparing motivations of individual programmers and firms to take part in the open source movement: From community to business. *Knowledge, Technology & Policy* 18, 4 (2006), 40–64.

[8] Simon Butler, Jonas Gamalielsson, Bjorn Lundell, Christoffer Brax, Johan Sjoberg, Anders Mattsson, Tomas Gustavsson, Jonas Feist, and Erik Lonroth. 2019. On company contributions to community open source software projects. *IEEE Transactions on Software Engineering* (2019).

[9] Javier Luis Cánovas Izquierdo and Jordi Cabot. 2020. A Survey of Software Foundations in Open Source. *arXiv e-prints* (2020), arXiv–2005.

[10] Peter G Capek, Steven P Frank, Steve Gerdt, and David Shields. 2005. A history of IBM's open-source involvement and strategy. *IBM systems journal* 44, 2 (2005), 249–257.

[11] Casey Casalnuovo, Bogdan Vasilescu, Premkumar Devanbu, and Vladimir Filkov. 2015. Developer onboarding in GitHub: the role of prior social links and language experience. In *Proceedings of the 2015 10th joint meeting on foundations of software engineering*. 817–828.

[12] Massimo G Colombo and Marco Delmastro. 2002. How effective are technology incubators?: Evidence from Italy. *Research policy* 31, 7 (2002), 1103–1122.

[13] Edward L Deci and Richard M Ryan. 1987. The support of autonomy and the control of behavior. *Journal of personality and social psychology* 53, 6 (1987), 1024.

[14] Juan C Duenas, Félix Cuadrado, Manuel Santillán, José L Ruiz, et al. 2007. Apache and Eclipse: Comparing open source project incubators. *IEEE software* 24, 6 (2007), 90–98.

[15] Armstrong Foundjem, Ellis E Eghan, and Bram Adams. 2021. Onboarding vs. Diversity, Productivity, and Quality-Empirical Study of the OpenStack Ecosystem. In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*. IEEE, 1033–1045.

[16] Marco Gerosa, Igor Wiese, Bianca Trinkenreich, Georg Link, Gregorio Robles, Christoph Treude, Igor Steinmacher, and Anita Sarma. 2021. The shifting sands of motivation: Revisiting what drives contributors in open source. In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*. IEEE, 1046–1058.

[17] Wael H Gomaa, Aly A Fahmy, et al. 2013. A survey of text similarity approaches. *International Journal of Computer Applications* 68, 13 (2013), 13–18.

[18] Sean M Hackett and David M Dilts. 2004. A systematic review of business incubation research. *The Journal of Technology Transfer* 29, 1 (2004), 55–82.

[19] Michael Hahsler and Kurt Hornik. 2007. New probabilistic interest measures for association rules. *Intelligent Data Analysis* 11, 5 (2007), 437–455.

[20] Il-Horn Hann, Jeff Roberts, Sandra Slaughter, and Roy Fielding. 2002. Why do developers contribute to open source projects? First evidence of economic incentives. In *2nd workshop on open source software engineering, Orlando, FL*.

[21] Phillip Kemp. 2013. The influence of business incubation in developing new enterprises in Australia. (2013).

[22] Brenda L Mak and Hy Sockel. 2001. A confirmatory factor analysis of IS employee motivation and retention. *Information & management* 38, 5 (2001), 265–276.

[23] Francisco Mas-Verdú, Domingo Ribeiro-Soriano, and Norat Roig-Tierno. 2015. Firm survival: The role of incubators and business characteristics. *Journal of Business Research* 68, 4 (2015), 793–796.

[24] Robert K. Merton. 1968. The Matthew Effect in Science. *Science* 159, 3810 (1968), 56–63.

[25] Hussan Munir, Johan Linåker, Krzysztof Wnuk, Per Runeson, and Björn Regnell. 2018. Open innovation using open source tools: A case study at Sony Mobile. *Empirical Software Engineering* 23, 1 (2018), 186–223.

[26] Jeffrey A Roberts, Il-Horn Hann, and Sandra A Slaughter. 2006. Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the Apache projects. *Management science* 52, 7 (2006), 984–999.

[27] Cristina Rossi and Andrea Bonaccorsi. 2006. Intrinsic motivations and profit-oriented firms in open source software: Do firms practise what they preach? In *The economics of open source software development*. Elsevier, 83–109.

[28] Hugh D Sherman. 1999. Assessing the intervention effectiveness of business incubation programs on new business start-ups. *Journal of developmental entrepreneurship* 4, 2 (1999), 117.

[29] Param Vir Singh, Yong Tan, and Vijay Mookerjee. 2011. Network effects: The influence of structural capital on open source project success. *Mis Quarterly* (2011), 813–829.

[30] Anselm Strauss and Juliet Corbin. 1990. Open coding. *Basics of qualitative research: Grounded theory procedures and techniques* 2, 1990 (1990), 101–121.

[31] Yuri Takhteyev and Andrew Hilts. 2010. Investigating the geography of open source software through GitHub. *Manuscript submitted for publication* (2010).

[32] David R Thomas. 2003. A general inductive approach for qualitative data analysis. (2003).

[33] Marat Valiev, Bogdan Vasilescu, and James Herbsleb. 2018. Ecosystem-level determinants of sustained activity in open-source projects: A case study of the PyPI ecosystem. In *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 644–655.

[34] Georg Von Krogh, Stefan Haefliger, Sebastian Spaeth, and Martin W Wallin. 2012. Carrots and rainbows: Motivation and social practice in open source software development. *MIS quarterly* (2012), 649–676.

[35] Johannes Wachs, Mariusz Nitecki, William Schueller, and Axel Polleres. 2021. The Geography of Open Source Software: Evidence from GitHub. *arXiv preprint arXiv:2107.03200* (2021).

[36] Michael Williams and Tami Moser. 2019. The art of coding and thematic exploration in qualitative research. *International Management Review* 15, 1 (2019), 45–55.

[37] Chorng-Guang Wu, James H Gerlach, and Clifford E Young. 2007. An empirical analysis of open source software developers' motivations and continuance intentions. *Information & Management* 44, 3 (2007), 253–262.

[38] Yunwen Ye and Kouichi Kishida. 2003. Toward an understanding of the motivation of open source software developers. In *25th International Conference on Software Engineering, 2003. Proceedings.* IEEE, 419–429.

[39] Likang Yin, Zhunagzhi Chen, Qi Xuan, and Vladimir Filkov. 2021. Sustainability Forecasting for Apache Incubator Projects. *arXiv preprint arXiv:2105.14252* (2021).

[40] Likang Yin, Zhiyuan Zhang, Qi Xuan, and Vladimir Filkov. 2021. Apache Software Foundation Incubator Project Sustainability Dataset. In *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)*. IEEE, 595–599.

[41] Yuxia Zhang, Minghui Zhou, Audris Mockus, and Zhi Jin. 2019. Companies' Participation in OSS Development-An Empirical Study of OpenStack. *IEEE Transactions on Software Engineering* (2019).