

Open Source Software Sustainability: Combining Institutional Analysis and Socio-Technical Networks

LIKANG YIN, University of California, Davis, USA
MAHASWETA CHAKRABORTI, University of California, Davis, USA
YIBO YAN, University of California, Davis, USA
CHARLES SCHWEIK, University of Massachusetts Amherst, USA
SETH FREY, University of California, Davis, USA
VLADIMIR FILKOV, University of California, Davis, USA

ACM Reference Format:

Likang Yin, Mahasweta Chakraborti, Yibo Yan, Charles Schweik, Seth Frey, and Vladimir Filkov. 2022. Open Source Software Sustainability: Combining Institutional Analysis and Socio-Technical Networks. 1, 1 (May 2022), 23 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

ABSTRACT

Sustainable Open Source Software (OSS) forms much of the fabric of our digital society, especially successful and sustainable ones. But many OSS projects do not become sustainable, resulting in abandonment and even risks for the world's digital infrastructure. Prior work has looked at the reasons for this mainly from two very different perspectives. In software engineering, the focus has been on understanding success and sustainability from the socio-technical perspective: the OSS programmers' day-to-day activities and the artifacts they create. In institutional analysis, on the other hand, emphasis has been on institutional designs (e.g., policies, rules, and norms) that structure project governance. Even though each is necessary for a comprehensive understanding of OSS projects, the connection and interaction between the two approaches have been barely explored.

In this paper, we make the first effort toward understanding OSS project sustainability using a dual-view analysis, by combining institutional analysis with socio-technical systems analysis. In particular, we (i) use linguistic approaches to extract institutional rules and norms from OSS contributors' communications to represent the evolution of their governance systems, and (ii) construct socio-technical networks based on longitudinal collaboration records to represent each project's organizational structure. We combined the two methods and applied them to a dataset of developer digital traces from 253 nascent OSS projects within the Apache Software Foundation (ASF) incubator. We find that the socio-technical and institutional features relate to each other, and provide complimentary views into the progress of the ASF's OSS projects. Refining these

Authors' addresses: Likang Yin, lkyin@ucdavis.edu, University of California, Davis, CA, USA; Mahasweta Chakraborti, mchakraborti@ucdavis.edu, University of California, Davis, CA, USA; Yibo Yan, ybyan@ucdavis.edu, University of California, Davis, CA, USA; Charles Schweik, cschweik@umass.edu, University of Massachusetts Amherst, MA, USA; Seth Frey, sethfrey@ucdavis.edu, University of California, Davis, CA, USA; Vladimir Filkov, vfilkov@ucdavis.edu, University of California, Davis, CA, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

XXXX-XXXX/2022/5-ART \$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

combined analyses can help provide a more precise understanding of the synchronization between the evolution of institutional governance and organizational structure.

1 INTRODUCTION

Open Source Software (OSS) is a multi-billion dollar industry. A majority of modern businesses, including all major tech companies, rely on OSS without even knowing it. OSS contributions are an important manifestation of computer-supported collaborative work, for the high degree of technical literacy typical of OSS contributors. Even though this popularity attracts many software developers to open source, more than 80% of OSS projects are abandoned [39].

The failure of collaborative work in OSS has received attention from two perspectives. In software engineering, the focus has been on understanding success and sustainability from the socio-technical perspective: the OSS developers' day-to-day activities and the artifacts they create. In the management domain, on the other hand, emphasis has been on institutional designs (e.g., policies, rules, and norms) that structure governance and OSS project administration. In particular, systems that generate public goods address these and other endemic social challenges by creating governance institutions for attracting, maintaining, incentivizing, and coordinating contributions. Ostrom [33] defines institutions as "... prescriptions that humans use to organize all forms of repetitive and structured interactions ...". Institutions guide interactions between participants in an OSS project, and can be informal such as established norms of behavior, or more formalized as written or codified rules. These norms and formalized rules, along with the mechanisms for rule creation, maintenance, monitoring, and enforcement, are the means through which collective action in OSS development occur [39], and they can be tiered or nested, as in the context of OSS projects embedded within an overarching OSS nonprofit organization.

Both methods have separately been shown to be utilitarianly describing the state of a process, however, combining the two perspectives has been barely explored. In this paper, we undertake a convergent approach, considering from one side OSS projects' socio-technical structure and on the other aspects of their institutional design. Our goal is to use these two perspectives synergistically, to identify when they strengthen and complement each other, and to also refine our understanding of OSS sustainability through the two methodological approaches. Central to our approaches is the idea that trajectories of individual OSS projects can be understood in the convergent framework through the context provided by similar projects that already are being readily sustained or have been abandoned.

We leverage a previously published dataset [50] of traces representing OSS developer's day-to-day activities as part of the Apache Software Foundation Incubator (ASFI) project. These developers are a part of projects that have decided to undergo the process of incubation, toward becoming part of the ASF, and benefiting from the services it provides to member projects. The dataset includes historical traces and a sustainability label (graduation or retirement) for each project. Graduation is an indication of successful incubation and the readiness of a nascent project to join ASF proper, otherwise the project is retired. In other words and importantly, in this paper, we use the ASFI project outcomes of graduation or retirement as a measure of sustainability of the project. We assume that graduated projects are sustained longer than retired ones, although that might not always be the case¹. But key hurdles that OSS projects have to demonstrate to graduate is that they can (1) produce new releases, and (2) show the ability to attract new developers. Both of these factors arguably are key toward the sustainability of OSS projects.

¹For example, it could be that some ASFI retired projects simply could not adapt to the policies and requirements set in the ASFI program but yet continue on, "in the wild" or perhaps aligned with a different OSS foundation.

We utilize this dataset to study the extent to which graduated and retired projects differ from each other, from the point of view of both the socio-technical structure and the institutional governance. On the socio-technical side, we construct the monthly longitudinal social and technical networks for each project, and calculate several measures describing features of the networks. On the institutional governance side, we implement a classifier trained on manual annotations of institutional statements in the publicly accessible email communications among ASF participants. Then we compare the findings of our socio-technical and institutional metrics for project-level and individual-level activities. Next, we perform exploratory data analyses, deep-dive case studies, and eventually, we look at how socio-technical measures associate with the prevalence of institutional statements, and evolutionary trajectories during OSS project incubation to sustainability. In summary, we find that:

- We can effectively extract governance content from email discussions in the form of institutional statements, and they fall into 12 distinguishable topics.
- Projects with different graduation (i.e., sustainability) outcomes differ in how much governance discussion occurs within their communities, and also in their socio-technical structure.
- Self-sustained projects (i.e., graduated) have a more socially active community, achieving it within their first 3 months of incubation, and they demonstrate more active contributions to documentation and more active communication of policy guidance via institutional statements.
- A project's socio-technical structure is temporally associated with the institutional communications that occur, depending on the role of the agent (mentor, committer, contributor) communicating institutional statements.

To provide most relevant context, recently, Yin et. al. [49] showed that socio-technical networks can be used to effectively predict whether a project will graduate or retire from the ASF incubator. That work did not include any institutional or governance analysis. Here, we focus on closing the gap by studying the relationship between the organizational structure (i.e., the socio-technical system) and institutional governance in peer-contributed OSS projects. Our study is the first attempt to provide a common framework for simultaneous, socio-technical structure and institutional, analysis of OSS projects, in order to describe and understand a process affected by both, that is, project gaining self-sustaining and self-governing community and eventually graduating from the ASF incubator. We are hopeful that refining this convergent approach, of structural and institutional analyses, will open new ways to consider and study emergent properties like project sustainability.

2 THEORETICAL FRAMEWORK

Here we introduce the theories behind the two different viewpoints, Institutional Analysis and Development (IAD) and Social-Technical Systems (STS), as well as Contingency Theory serving as the glue between institutional governance and the organizational structure of OSS projects.

2.1 Institutional Theory and Commons Governance

OSS projects are a form of digital commons, or more precisely, Commons-Based Peer Production (CBPP) [39]. Legal scholar Yochai Benkler [2] introduced the phrase CBPP to describe situations where people work collectively over the Internet, and where organizational structure is less hierarchical. While CBPP situations are found in a variety of settings (e.g., collaborative writing, open source hardware) Benkler argues that OSS is the 'quintessential instance' of CBPP.

There is a relatively long history of the study of governance in commons settings, arguably led by Nobel laureate Elinor Ostrom and her groundbreaking book *Governing the Commons* [32]. Ostrom's Institutional Analysis and Development (IAD) framework was developed to study the governance institutions that communities develop to self-manage natural resources. Much of this

research focuses on the governance and sustainability of natural resource settings, e.g., water [6], marine [19, 27], and forest [16] settings.

A key challenge in natural resource commons settings is that individuals who cannot easily be excluded from extracting resources from the pool of available natural resources often have little incentive to contribute toward the production or maintenance of that resource – what are commonly referred to as ‘free-riders’ [30, 37]. In forest, fishery or water settings, the free-rider problem in open access settings can lead to a problem termed by Hardin as the ‘Tragedy of the Commons’ [20]. Ostrom famously pushed back against Hardin’s analysis and over a course of a lifetime of work, highlighted that communities can avoid tragedy through hard work in developing self-governing institutions.

OSS commons are fundamentally different from natural resources in that digital resources can be readily replicated and are not subject to degradation due to over-harvesting. Therefore, if over-appropriation is not a problem, is there a potential tragedy of the commons in an OSS context? Invariably the answer is yes, and it lies at the heart of the idea of OSS sustainability. The tragedy occurs when there are free-riders and insufficient human resources available to continue to further develop and maintain the software and, as a result, the software project fails to achieve the functionality and use that was perhaps envisioned when it began, and becomes abandoned [38]. Ostrom and Hess [22] aptly describe this tragedy as ‘collective inaction.’

Ostrom’s Nobel Prize winning body of work was studying how humans collectively act and craft self-governing institutional arrangements to effectively avoid the tragedy in natural resource settings. Central in this effort was the introduction and evolution of the Institutional Analysis and Development (IAD) framework [33]. Later, IAD was applied to the study of digital or knowledge commons [17, 22] and explicitly to the study of self-governance in OSS, where Schweik and English undertook the first study of technical, community, and institutional designs of a large number of OSS projects [39].

With that being said, prior work has found that self-governing OSS projects develop highly organized social and technical structures [5]. Those having foundation support, like the ASF, may additionally be in the process of organizing the developers’ structured interactions under a second tier of governance prescriptions as required by the ASF Incubator. We refer to an individual institutional prescription as an *Institutional Statement* (IS), which can include rules and norms, and which we define as a shared linguistic constraint or opportunity that prescribes, permits, or advises actions or outcomes for actors (both individual and corporate) [10, 41]. Institutions, understood operationally as collections of institutional statements, create situations for structured interaction for collective action. In other words, configurations of ISs affect the way collective action is organized. In the context of ASF and OSS projects, incubator ISs can affect OSS project social and technical structure.

With IS and other approaches to institutional analysis, it becomes possible to articulate the relationships between governance, organizational, and technical variables. For example, previous studies on OSS often report code modularity as a key technical design attribute [29, 31]. Hissam et al. [23] write: ‘A well-modularized system ... allows contributors to carve off chunks on which they can work.’ Open and transparent verbal discussion between OSS team members and other ASF officials (e.g., mentors) about OSS project or ASF institutional design, captured in the form of institutional statements, could then predict effort by project contributors to restructure their project’s technical infrastructure to be more modular and inviting to new contributors. Using the approaches of institutional analysis, we extract institutional content from open access email exchanges between OSS project contributors to understand the role of communication governance information in OSS project sustainability.

2.2 Socio-Technical System Theory

A Socio-Technical System (STS) comprises two entities [44]: the social system where members continuously create and share knowledge via various types of individual interactions, and the technical system where the members utilize the technical hardware to accomplish certain collective tasks. STS theory can be considered to combine the views from both engineers and social scientists – an intermediary entity of sorts – that transfers the institutional influence to individuals [36]. The theory of STS is often referenced when studying how a technical system is able to provide efficient and reliable individual interactions [21], and how the social subsystem becomes contingent in the interactions and further affects the performance of the technical subsystem [15]. Moreover, the socio-technical system theory plays an important role in analyzing collective behavior in OSS projects [3, 48]. OSS projects have also been studied from a network point of view [12, 24]. González-Barahona et al. [18] proposed using technical networks, where nodes are the modules in the CVS repository and edges indicate two modules share common committers, to study the organization of ASF projects. In socio-technical systems, organizations can intervene through long-term or short-term means. Smith et al. [42] propose two conceptual approaches, ‘outside’ and ‘inside’: ‘outside’ approaches represent the socio-technical and are managerial in approach. ‘Inside’ approaches are more reflexive about the role of management in co-constituting the socio-technical.

From that perspective, the Apache Software Foundation (ASF) community is a unique system that has both outside influence regulations from ASF board and members and inside governance managed or self-governed by individual Project Management Committees (PMC).

2.3 Contingency Theory, or There Are No Panaceas in Self-Governance

Contingency theory is the notion that there is no one best way to govern an organization. Instead, each decision in an organization must depend on its internal structure, contingent upon the external context (e.g., stakeholder [45], risk [9], schedule [47], etc.). Joslin et al. [25] find that project success is associated with the methodologies (e.g., processes, tools, methods, etc.) adopted by the project. Here, in particular, we treat the institutional statements as an abstraction of the methodologies in OSS development. As the organizational context changes over time, to maintain consistency, the project must adapt to its context accordingly. Otherwise, conflicts and inefficiency occur [1], i.e., not a single organizational structure is equally effective in all cases. Similar arguments have been made in the field of institutional analysis, arguing that there are no panaceas or standard blueprints for guiding the institutional design of a collective action problem [34].

To address the conflicts caused by incompatibilities with the project’s context, previous work suggests thinking holistically. Lehtonen et al. [26] consider the project environment as all measurable spatio-temporal factors when a project is initiated, processed, adjusted, and finally terminated. They suggest that the same factor can have an opposite influence on the projects under a different context. Joslin et al. [25] consider project governance to be part of the project context, concluding that project governance can impact the use and effectiveness of project methodologies.

As per contingency theory, during ASFI projects’ incubation, developers and mentors have to make in-time decisions on their organizational structure, contingent on what is happening on the institutional rules and governance, and vice versa.

3 RESEARCH QUESTIONS

Reflecting on the previous discussion, the primary goal of this paper is to demonstrate that the evolution of a project from a nascent state to a sustainable state can be studied effectively by combining the two different methodologies of socio-technical network analysis and institutional analysis.

We reported in prior sections that a variety of scholars have utilized a socio-technical systems approach to analyze collective behavior in OSS projects. We also described how institutional analysis is useful in understanding collective action in OSS settings. To enable to dual-view on sustainability, we first describe and evaluate our automated approach to identifying institutional statements in project emails.

RQ₁: Are there institutional statements contained in ASF Incubator project email discussions? Can we effectively identify them?

With the next two research questions, we assess the utility of our convergent approach to the Institutional Analysis (IAD) and STS frameworks. In the case of the ASF incubation program, there are two eventual outcomes: either a project graduates from the ASF incubator and becomes a full-fledged ASF associated project, or it “retires” without achieving that goal. In this context, we operationalize a sustainable state as one where an OSS project graduates from the ASF incubator program, rather than retires. We ask:

RQ₂: Is OSS project evolution toward sustainability readily observable through the dual lenses of institutional and socio-technical analysis? And how do such temporal patterns differ?

Per institutional analysis theory, strategies, norms, and rules can affect the social and technical organizations of projects. Governance and organization, per social theories, must work hand-in-hand to make viable socio-technical systems. Ill-designed institutional arrangements would introduce inefficiencies into the system, and such inefficiencies may amplify deviant behaviors and irregular structure in the systems. Such influential links from institutional design to the organizational structure can be, in fact, bi-directional. In effect, in a sustainable system, an ill-formed organizational structure may instigate new rules to adjust and improve such structure, further improving efficiencies in the systems.

Thus, we hypothesize that the feedback, if any, between project governance and project organization should be observable, specifically in that intensified governance discussion should precede and/or follow changes to the project organizational structure. As a reminder, we consider institutional statements as indicators of intensified discussions of OSS project self-governance or new incubator requirements on that self-governance. We also consider socio-technical network parameters as indicators of organizational structure. Thus, we ask:

RQ₃: Are periods of increased Institutional Statements frequency followed by changes in the project organizational structure, and vice-versa?

In the following section, we introduce the methodologies approaching the above three research questions.

4 DATA AND METHODS

To study the difference between projects that graduate ASFI (i.e., become sustainable) and those that do not, in this paper we use a collection of large-scale data sets comprising Institutional Statements and Socio-Technical variables extracted from all graduated and retired projects from the Apache Software Foundation Incubator, ASFI. In ASFI, graduation is an indication that a nascent project is sufficiently sustainable to join ASF proper², otherwise the project is retired. Our combing through the Apache lists, inspecting the data, and speaking to project and community members have shown that almost all failures to graduate are sustainability failures. In rare occasions, some projects have retired for reasons other than sustainability, e.g., some are not a good fit for the Apache model³,

²ASF’s guide to project graduation: <https://incubator.apache.org/guides/graduation.html>

³ASF’s reason behind projects’ retirement: <https://incubator.apache.org/projects/#retired>

despite evidence that projects are generally sufficiently aware of the ASF model before entering incubation according to their project proposal⁴.

For the socio-technical networks, we collected historical trace data of commits, emails, and incubation outcomes for 253 ASF projects, which have available archives of both commits and emails from 03/29/2003 to 02/01/2021⁵. Among those, 204 projects have already graduated, and 49 have retired. ASF incubator projects that are still in incubation are not studied in this paper.

We collected the ASF incubator project data from the ASF mailing list archives⁶, which are open access and can be retrieved through the archive web pagelists, http://mail-archives.apache.org/mod_mbox/. They contain all emails and commits from the project's ASF incubator entry date, and are current. The project URLs follow the pattern: `proj_name - list_name/(YYYYMM).mbox`. For example, the full URL for the *dev* mailing list of the *Apache Accumulo* project, in Dec 2014, is http://mail-archives.apache.org/mod_mbox/accumulo-dev/201412.mbox. Each such *.mbox* file contains a month of mailing list messages from the project, for the date specified in the URL. Here *dev* stands for 'emails among developers'. Notably, there are some sites that are not following the pattern, e.g., 'ASF-wide lists' are not project-owned mailing lists, and list 'incubator.apache.org' contains data of more than one project.

To extract Institutional Statements, we combined our email data set with a prior data set on ASF policy documents. In a given organization, institutional statements are characterized by a finite set of semantic roles (e.g. ASF Board, Mentors, contributors, etc. in ASF), and their interactions (e.g. management committees requesting reports from projects, developers voting to induct committers in ASF), in specific contexts. To account for their representation in our training corpus, we included institutional statements from not only ASF project-level email exchanges among participants, but also ASF policy documents. The supplementary set of Institutional Statements included 328 policies, which were compiled from ASF policy documents (e.g., Apache Cookbook, PPMC Guide, Incubator Policy, etc), in an economic analysis of the ASF Incubator's policies [40].

4.1 Pre-processing

We collected all 1,330,003 emails across the ASF Incubator projects, from 03/29/2003 to 02/01/2021 (under mailing lists of 'commit', 'dev', 'user', etc.). We find that 128,257 (about 9.6%) emails are automatically generated and broadcast by continuous integration tools (i.e., bots). Because the amount of such emails is substantial, but they carry less meaningful social or institutional information, and list members almost never reply to them, we use regular expression rules to identify and eliminate them from the corpus, leaving us 1,201,746 emails.

And, for the technical contribution side, many projects, especially those over ten years old that used SVN, utilized a bot for extensive mailings, thus forming outliers in the dataset. Thus, we eliminate commit messages from automated bots (e.g., 'buildbot'), 253,758 out of 3,654,196 (about 14.4%) commit messages, and email messages from issues/bug tracking bots (e.g., 'GitBox'). Moreover, we find some developers contributed commits by directly changing/uploading massive non-source code files (e.g., data, configuration, and image files). Since committing non-coding files can form outliers in the data set, we choose to apply the *GitHub Linguist*⁷ to identify 731 collective programming language and markup file extensions, and exclude any other non-coding commits (e.g., creating/deleting folder, upload images, etc.).

⁴ASF incubator projects' proposal <https://cwiki.apache.org/confluence/display/INCUBATOR/Proposals>

⁵Our code and data is available at Zenodo: <https://doi.org/10.5281/zenodo.5908030>

⁶During the submission of this study, ASF has moved their email archives to Pony Mail.

⁷GitHub Linguist <https://github.com/github/linguist>

Table 1. Selected Examples of Institutional Statements Found in ASFI Project Email Discussions.

Project	Date	Institutional Statements
Airflow	21 Dec 2016	... running in our Lab there is virtually no restriction what we could do, however I will hand select people who have access to this environment. I will also hold ultimate power to remove access from anyone ...
ODF	07 Dec 2011	Please vote on releasing this package as < Package >. The vote is open for the next 72 hours and passes if a majority of at least three +1 ODF Toolkit PMC votes are cast ...
Airflow	24 Feb 2017	... Next steps: 1) will start the voting process at the IPMC mailinglist. ... So, we might end up with changes to stable. ... 2) Only after the positive voting on the IPMC and finalisation I will rebrand the RC to Release.

4.2 Constructing Socio-technical Networks

Network science approaches have been prominent in studying complex systems, e.g., OSS projects [4, 43]. Since networks can contain rich information for both the elements (i.e., nodes) and their interactions (i.e., edges), in this study, we use socio-technical networks to anchor the abstraction of socio-technical systems. We define the projects' socio-technical structure using social (email-based) and technical (code-based) networks, extracted from their emails to the mailing lists and commits to source files. Similar to the approach by Bird et al. [3], we form a social network (weighted directed graph) for each project, at each month in their incubation, from the communications between developers: a directed edge from developer A to B forms if B has replied to A 's post in a thread or if A has emailed B directly. The weight of the edge represents the communication frequency between a pair of developers. The technical bipartite networks (weighted bipartite graph) are formed in a similar way. For each project, at each month, we include an un-directed edge between a developer A and a source file F if developer A has committed to the source file F that month (excluding the SVN branch names). The weight of the edge represents the committing frequency between the developer and the source file. In summary, social networks are weighted directed graphs. We form edges between two developers nodes, if one developer replied to or referenced the other's email. Technical networks are undirected bipartite graphs, with developers forming one set of nodes, coding files forming the other, and a link being drawn when a developer contributed to a coding file. We use the `NETWORKX` package from *Python* for the network-related implementation.

4.3 Extracting Institutional Statements

We combined the email exchange data set with the ASF policy document data to fine-tune a BERT-based [8] classifier, for automatic detection of ISs (see Sect. 2.1 for definition of IS).

To start, we hand-annotated a small subset of our data for ISs as follows⁸. After selecting a random subset of 313 email threads from incubator project lists, two hand-coders labeled the sentences in them as "IS" or "Not IS", on the basis of whether they fit the definition of Institutional Statements. They resolved disagreements through discussion and recorded these conclusions, achieving a peak out-of-sample agreement between 0.75 to 0.80. A sentence was coded as an IS only if it was a complete sentence; fragments such as parenthetical mentions of rules or resources were not annotated as positive. This resulted in 6,805 labeled sentences (i.e., "IS" or "Not IS"); 273 were labeled as IS.

We treated all 328 policies from the ASF documents as institutional statements, since policy documents provide arguably more formal institutional sample text compared to the norm in the email discussions. Thus, we had 601 Institutional Statements in total across these two coded datasets.

⁸Our coding manual for training the annotators can be found here: <https://doi.org/10.5281/zenodo.5908030>.

Institutional statements refer to prescriptions and shared constraints in the form of norms, rules, and strategies that are meant to mobilize and organize actors towards collective actions. The examples of institutional statements provided in Table 1 provide some instances of developer exchanges that encompass norms and strategies with institutional implications. The first example from the Airflow project, dated 12/21/2016, involves a situation where certain developers find the computational infrastructure provided by ASF insufficient for testing and development requirements, and discuss setting up alternate arrangements to meet the bottleneck. Faced with resource limitations, one developer offers an externally hosted cloud environment through his private resources. The selected excerpt is a quote from the individual establishing the terms for using the alternate resources he may offer to the project members, including access permission and usage restrictions. ASF projects conduct voting from time to time to gather community consensus in matters of significance. The following example from ASF project ODF, dated 12/07/2011 describes the stepwise process expected to be followed by members project-wide to conduct a vote that decides on the approval of the release of the current candidate under development. The final example from Airflow, 02/24/2017 also pertains to a similar process, where a developer discusses the voting process and the implications, especially in terms of subsequent steps that need to be fulfilled to ensure product release.

BERT-based Sequential Classifier. In natural speech, such as emails, ISs can appear as whole sentences, parts of sentences, or span multiple sentences. They are also relatively sparse, with their institutional quality dependent on their inherent interpretation as well as context. Framing IS extraction as a sequential sentence classification task in the context of self-contained email segments, instead of labeling individual sentences helps take into account contextual cues.

We used the sequential sentence classifier developed by Cohan et al. [8], which leverages Bidirectional Encoder Representations from Transformers (BERT) sequence classifier [11] to classify sentences in documents. BERT can be employed to generate representation for a sentence in a document's context, through joint encoding over its neighbouring sentences and then leveraging the corresponding sentence separator '<SEP>' token's tuned embedding for downstream applications, such as sentence labeling, extractive summarizing, etc. Thus, our classifier comprises of BERT for attention-based joint encoding across sentences followed by a feedforward classifier to predict sentence labels based on these separator '<SEP>' vectors.

To test the performance of the classifier on email IS extraction, we held-out 40 email threads (12.5%, randomly split) out of our 313 hand annotated email threads. The training was performed on the combined set of the remaining 273 coded email threads and the ASF policy documents. The coded training and, respectively, testing email data contained 231 and, respectively, 42 institutional statements. For both training and testing, email threads were processed to generate classifier inputs as follows. To include neighboring context while meeting length limits of the BERT-based text classifier, for each email document, sentences were first chunked into segments using a sliding window of up to 256 BERT sub-word (wordpiece) tokens. This resulted in segments containing 6 contiguous sentences each, on average, comprising as many full sentences as could be accommodated in the specified subword limit. The rolling window had a step of 1 full sentence. We generated 3322 and 384 email segments for training and testing, respectively. For the policy documents, each policy with its sentences was treated as a segment, leading to 328 additional segments in the training data. There are several reasons to support inclusion of ASF policies to augment positive training examples. (1) In terms of semantic information, they are about institutional themes and actions. This was expected to help the language model learn what sets apart Institutional themes from regular development activity and artifacts. (2) ASF policies are critical in common pool resource

management and institutional operations as they describe roles, responsibilities and regulate actions, and are often invoked in email discussions⁹. (3) The institutional statements of the formal policies are the source texts that in-email references to IS are drawing from when they discuss ASF's rules in email. From this perspective, they are a vital source text for detecting these statements as they occur in email settings. Hence, while apparently sourced from formal bylaws beyond emails, ASF policies are indeed institutional statements relevant and recurring in developer conversations and are hence included in the training data.

We fine-tuned our classifier end-to-end against the corresponding labels for sentences in the segment. Training was conducted with a batch size of 16 and learning rate of $2 \cdot 10^{-5}$, for 6 epochs. All other hyperparameters were left as defaults. To account for the class imbalance, we randomly oversampled training data segments which had at least one IS sentence to match the number of segments which had no IS sentences (1:1). In both the training and predicting phase, we did not incorporate any temporal information, other than the sequentiality captured by the segments. That is, when extracting the institutional statements, the model does not require the exact time of the discussion.

During testing or prediction, due to variable length of context preceding or following each sentence in any particular segment, we treat a sentence in an email as a "positive" classification, if it has been detected as an IS in at least one segment. The performance of the model has been reported in terms of the F1-score, precision and recall with respect to the positive ('IS') label detected for sentences in the test email set in Sect. 5.1.

4.4 Topics Identification in Institutional Statements

The purpose of text modeling is to describe the text given a specific corpus, and provide numerically measurable relationships among texts, e.g., topics identification, measuring similarity, etc. We use a Latent Dirichlet Allocation (LDA) model to get semantically meaningful topics to better understand the extracted institutional statements. LDA is an unsupervised clustering approach [51], which when given a set of documents, iteratively discovers relevant topics present in them, based on the word distributions and relative prevalence in each document. We used LDA to identify prominent topic clusters occurring among all institutional statements extracted from our email archives through our trained classifier (see Sec 4.3). No prior training from our coded email set against pre-identified topic labels was used to train the LDA model. We use the coherence score provided by the *gensim* package [46] to optimize the performance of the LDA model with respect to the number of topics; a higher coherence score represents a better clustering performance. We select the LDA model with the highest coherence score from which to draw the clusters. However, since the LDA model does not automatically generate us a label for each cluster, we need to assign a label intuitively based on our domain knowledge of the ASF incubation process. Naming each topic cluster certainly carries some risks on interpretation, however, we believe that providing all top keywords for each cluster reduces such risk.

4.5 Variables of Interest

We draw institutional and socio-technical project features and variables on the basis of each framework's predictions for our research questions. Our socio-technical variables are pulled from a recent study on forecasting the sustainability of OSS projects [49], showing high predictive power of socio-technical variables. All metrics are aggregated over monthly intervals, for each project, from the start to the end of its incubation.

⁹<https://lists.apache.org/thread/zykybdvnk9cwx03pnrfl2br9nkcb7q3f>

Table 2. Summary statistics for the monthly socio-technical variables and the counts of institutional statements from project mentors, committers, and contributors after removal of the top 2% of outliers. The numbers in parentheses denote the values after the removal of inactive months (i.e., absent of emails/commits). Prefix $s_$ denotes features in the social network while $t_$ represents the technical network.

Statistic	Mean	St. Dev.	25%	75%
s_num_nodes	13.04 (16.96)	14.56 (15.04)	4 (7)	17 (22)
$s_graph_density$	0.30 (0.30)	0.27 (0.22)	0.12 (0.14)	0.40 (0.40)
$s_avg_clustering_coef$	0.22 (0.29)	0.23 (0.21)	0 (0.11)	0.39 (0.43)
$s_weighted_mean_degree$	11.83 (15.56)	12.03 (12.81)	4 (7.43)	16 (19.71)
$t_graph_density$	0.37 (0.68)	0.41 (0.32)	0 (0.36)	1 (1)
$t_num_dev_nodes$	1.18 (2.21)	1.59 (1.60)	0 (1)	2 (3)
$t_num_file_nodes$	60.99 (114.83)	153.94 (197.25)	0 (6)	38 (126)
$t_num_file_per_dev$	28.79 (53.57)	80.46 (104.23)	0 (4)	20 (54.5)
num_IS_mentor	15.46 (15.99)	24.46 (25.01)	0 (1)	20 (20)
$num_IS_committer$	9.34 (12.89)	19.36 (22.36)	0 (0)	10 (16)
$num_IS_contributor$	13.18 (16.36)	21.72 (24.42)	0 (2)	18 (21)

Longitudinal Socio-Technical Metrics: For each project network, for each month, we constructed the social and technical networks, and from them calculate various organizational structure measures. In our tables and results, the prefix $t_$ in a variable's name indicates it is of the technical (code) network, while the prefix $s_$ in a variable's name indicates it is of the social (email) network. For the monthly social networks, we calculate the weighted mean degree $s_weighted_mean_degree$ (sum of all nodes' weighted degree divided by the number of nodes), average clustering coefficient $s_avg_clustering_coef$ (the average ratio of closed triangles over open triangles), graph density $s_graph_density$. In the technical bipartite networks, for each month, we calculate the number of unique developer nodes $t_num_dev_nodes$, the number of unique file nodes $t_num_file_nodes$, the number of files per developer $t_num_file_per_dev$, and the graph density $t_graph_density$.

Institutional Statements Frequency Metrics: For each project, in each month, we added up the ISs in all emails in that month sent by each of the following three separate and identifiable groups of people: ASF mentors (num_IS_mentor), registered ASF committers ($num_IS_committer$), and contributors ($num_IS_contributor$). We summarize their statistics in Table 2. As noted earlier, there is a final group of emails not accounted here, sent by bots. Similar to calendar entries, they may be useful, but are not the object of our study here.

4.6 Granger Causality

Time series data allows for the identification of relationships between temporal variables that go beyond association. One approach, *Granger causality*, is a statistical test for identifying quasi-causality between pairs of temporal variables [13]. Given two such variables, X_t and Y_t , the Granger causality test calculates the p-value of Y_t being generated by a statistical model including only Y 's prior values, Y_{t-1}, Y_{t-2} , etc., versus it being generated by a model that in addition to Y 's prior values, also includes X 's prior values X_{t-1}, X_{t-2} . Thus, Granger causality simply compares a base model involving only Y to a more complex model involving Y and X , and calculates if the latter is a better fit to the data. In the context of Granger causality, prior values are called *lagged values*, with X_{t-1} having a lag of 1, X_{t-2} having a lag of 2, etc. If the Granger causality test returns a small enough

p-value (e.g., < 0.01) it is interpreted as rejection of the null hypothesis, and thus establishing that X Granger causes Y .

The Granger causality test makes an assumption that the time-series on which it is applied are stationary, meaning they do not have trend or seasonal effects. It is necessary to test for stationarity before running the Granger causality. We use the augmented Dickey-Fuller test [7], as implemented in `ADF.TEST` from the *R* package `TSERIES` [28], to test stationarity. Both institutional and socio-technical variables were found to be stationary. We note that a distinction is typically made between scientific causality based on controlled experiments, and Granger causality, with the latter only satisfying one (precursor property) of multiple different properties of causality. Because of that, when Granger causality is used, the word ‘causality’ is always preceded by ‘Granger’. We also note that this test does not identify the sign, if any (i.e., positive or negative) of the Granger causality. It simply says if one exists. We use the `PGRANGERTEST` function to test Granger causality in *R*.

5 RESULTS

In this section, we answer the proposed research questions by adopting a dual-view, from the institutional analysis and socio-technical network perspectives. We first establish the utility of our IS identification methodology.

5.1 RQ₁: Are there institutional statements contained in ASF Incubator project discussions? If any, can we effectively identify the content of ISs?

Detecting Institutional Statements. First, we focus on the ability of our BERT-based classifier to identify institutional statements in the emails. When tested on the 857 held out sentences from the 40 email threads in our test set, see Sect 4.3, our classifier achieved a precision score of 0.667, recall score of 0.681, and F1 score of 0.674 on classifying Institutional Statements, demonstrating it is able to extract ISs from developer email exchanges in spite of there being only 5.1% ISs.

For model validation against overfitting, we sought to perform stratified cross-validation (CV) on our training data. We note that our data was not ideal for a CV study: we had (1) limited data size (2) uneven distribution of ISs across the email threads and (3) class imbalance between IS and non IS sentences. E.g., due to the limited data size, emails with high IS density could find their way in the train but not the test split, and dramatically increase the variance in cross-validation results. To ameliorate that, for more uniform stratification we chunked up each of the 273 threads in our training data into 442 sub emails of 20 contiguous sentences each (the email threads had a mean length of 22 sentences). We fine-tuned our classifier end-to-end against the corresponding labels for sentences in the sub emails. Training was conducted with a batch size of 16 and a learning rate of $2e-5$, for 6 epochs. All other hyperparameters were left as defaults. The subsequent input segment generation and training of the pipeline were otherwise kept unchanged. We obtained a mean F1 score on positive labeling of sentences with ISs of 0.603, with some high IS variability between folds still persisting.

We consider these performance results satisfactory given that we had a small and highly imbalanced data set (273 ISs out of 6,805 sentences). There are strong indications that increasing the positive examples in the training data set will further increase our classifier’s performance.¹⁰ Of course, it is challenging to ascertain if classifier performance varies across projects due to limited data size and expected differences in underlying distribution of institutional themes and policy mandates across projects.

¹⁰When we fine-tuned the classifier with only the 273 training email threads (i.e., without Institutional statements from the ASF policy documents), the F1 for positive label was found to be about 20% lower.

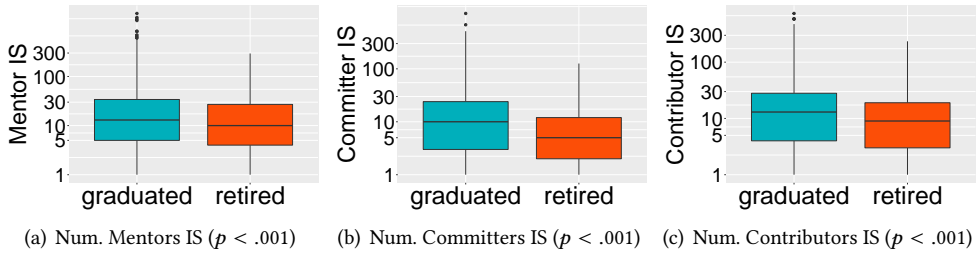


Fig. 1. Comparing graduated (in blue) vs retired (in red) projects along the number of Institutional Statements (IS) (color online). The Mann-Whitney U test p-val is sufficiently small (in brackets), suggesting significant differences in means between groups.

We ran our classifier on the full corpus of 1,201,746 emails (after bot email removal) across all ASF incubator projects. It identified 313,140 ISs in the emails, for an average of 0.261 sentence-level ISs per email. Table 2 shows descriptive statistics for both the socio-technical variables and the number of institutional statements from project mentors, committers, and contributors, calculated in monthly intervals, per project.

We find that the classifier’s errors are also informative. In one set of false positives, participants described plans for an event occurring outside of Apache and the relevant incubator project, not the kind of process or behavioral constraint typical of ISs. It was probably detected as an IS due to its semantic similarity to rules and guidelines which make up other positive examples. Conversely, the sentence ‘Send it to <EMAIL> and see what the reaction is’ was missed as an IS, despite appearing in the context of contributor agreements. This miss is likely due to the fact that many such recommendations are made in the emails that would not be considered institutional, because they indicate a particular individual as an individual, rather than in their institutional role.

Institutional Statements Over Roles and Sustainability Status. We turn to some exploratory analysis, to demonstrate the utility of our chosen features when reasoning about differences between graduated and retired projects. Comparing graduated and retired projects, we find a significant difference in the number of ISs. For example, in Figure 1(a), the number of IS sent by mentors graduated projects is statistically higher than retired projects (the Mann-Whitney U test is used for testing the difference in means). This, along with the fact that graduated projects tend to be more active socially overall compared to retired projects (i.e., more email exchanges), suggesting the mentors of retired project are concerned about the projects’ community progressing, thus, most of the email content is about rules and guidance. On the other hand, it is also plausible that mentors engage more socially and less institutionally with graduated projects, which may benefit those projects more. The numbers of ISs sent by committers and contributors, show similar patterns. We investigate them longitudinally in next section.

Topics Identification in Institutional Statements. We use Latent Dirichlet Allocation (LDA) model to study the token-level topics in institutional statements. By optimizing the LDA coherence score, we get the optimal number of topics of 12. The result further enables us to study which words are important to each topic. We present the clusters of top words for each topic in Table 3.

As this table reveals, words are well extracted from the institutional statements and are distinguished from each other. For example, in the first topic (i.e., ‘Progress Report’), there is the cluster of words – ‘review’, ‘board’ (which relates to ASF board), ‘submit’ and ‘report’ – all of which are associated with the important incubator rule that requires podlings (i.e., ASF’s referring to projects) to report regular progress reports. While in topic 7, words like ‘update’, ‘wiki’, ‘page’, ‘website’,

Table 3. Topics Identified in Institutional Statements.

ID	Heuristic Topic	Top Sample Words
1	Progress Report	review, require, meeting, board, submit, report
2	Collective Decision	vote, start, proposal, thread, close, day, bind
3	Project Release	release, issue, think, fix, branch, policy
4	Community	project, email, send, community, behalf, incubation, talk
5	Report Review	board, report, time, meeting, prepare, reminder, review
6	Mailing List Issues	list, mailing, discussion, question, issue, comment, request
7	Documentation	update, wiki, page, website, documentation, link, doc
8	Software Testing	release, source, build, test, note, artifact, check
9	licensing Policy	license, file, software, version, copyright, compliance
10	Routine Work	project, committer, help, work, way, code
11	Mentorship	podling, report, form, mentor, know, sign, month, wish
12	Software Distribution	work, repository, information, file, distribute, commit



Fig. 2. Topics Evolution for graduated projects (in blue) compared to retired projects (in red). X-axis indicates the i -th month from their incubation start, y-axis represents the relative volume of the topics. Mann-Whitney U test found 10 out of 12 topics are significantly different in their means between graduated and retired projects (p -val < 0.01). Not significant were topic 9 (licensing policy) and topic 12 (software distribution).

and ‘documentation’ emerge, all related to requirements projects need to address related to their website or documentation requirements. The results advance the institutional theory under the software engineering domain, arguably that the IS is associated with OSS sustainability, suggesting us to dive deeper into the connections between social-technical system and institutional analysis.

RQ₁ Summary: We demonstrated that institutional analysis methodologies can capture differences between graduated projects and retired projects. We also showed that we can effectively identify meaningful institutional statements, and common topics, from ASF incubator projects’ emails.

5.2 RQ₂: Is OSS project evolution toward sustainability observable through the dual lenses of institutional and socio-technical analysis? And how do such temporal patterns differ?

In this section, our goal is to contrast graduated and retired projects over time in both IS space and socio-technical space. Projects exit the ASF incubator at different times. In effect, there will be larger variance during the end of the incubation month. Therefore, we restrict ourselves to the first 24 months for all projects (more than 60% projects stayed within 24 months in the incubator).

Topic Evolution Over Time. After identifying the words that contribute to various identified topics, by aggregating over all projects, we get the volume, which is measure by the number of tokens contributing to that topic, of each topic in each month. Moreover, since there exist trends of the number of IS, we subtract the mean volume for each month, separately for the graduated and retired projects. We present them in Figure 2, where the x-axis is the number of months after their incubation start, and y-axis indicates the relative volume compared to the mean.

The results of Mann-Whitney U test show 10 out of 12 topics are significantly different in their means between graduated and retired projects ($p\text{-val} < 0.01$). Not significant were topic 9 (licensing policy) and topic 12 (software distribution). Additionally, the augmented Dickey-Fuller test suggests that over time, 9 out of 12 topics are not stationary (i.e., temporal trends exist, with $p\text{-val} < .01$), except for topic 2 (collective decision), topic 6 (mailing lists), and topic 12 (software distribution). The testing results prompt us to analyze the difference in project-level dynamics between graduated and retired projects.

We observe an increasing trend of Topic 1 ‘Progress Report’ with a small seasonal effect, suggesting the projects are learning the ‘Apache Way’ and more actively discussing their regular project reporting over time. And such seasonal effect is found to be more significant in the Topic 5 (‘Report Review’). Project releases, documentation, and software testing, are all connected to the number of people participating regularly. The retired projects are on average smaller than the graduated ones, which is the likely explanation for the differences. E.g., in Figure 3(f), we show that graduated projects, on average, have more source files than the retired projects. Moreover, we find that Topic 9, ‘license policy’, has an increasing trend in the earlier stages of incubation (e.g., months 1-7) which makes sense in that the shift from one OSS license to the license required by ASF is an important discussion that projects would want to address earlier on.

On the contrary, the longitudinal pattern of IS language related to software testing is relatively rare in the beginning project incubation. It suggests that in earlier stages of incubation, developers are more likely focused on the transition to the incubator and perhaps less on new code development and testing. On the other hand, such transitions were implemented in fast manner, with testing discussions increasing rapidly in incubation months 3, 4 and 5.

By comparing graduated and retired projects, we find that, Topic 10, ‘Routine work’, to be the dominant topic for both types of projects, almost through all projects’ incubation (i.e., remain high volume compared to other topics). We also find that graduated projects tend to be more active on Topic 7 ‘Documentation’ and Topic 3 ‘Project Release’. Interestingly, on the other hand, mentorship related ISs (Topic 11) are found to be more active for the retired projects rather than graduated projects. One possible reason is that retired projects did seek help from their mentors when their projects were experiencing downturns, and further issuing project-wise statements.

Metric Evolution. We continue by exploring the evolution of our metrics over time. Looking at the mentors ISs, shown in Figure 3(a), we can see that even in the beginning of their incubation, mentors email a greater number of ISs to projects that eventually graduate compared to ones that eventually retire.

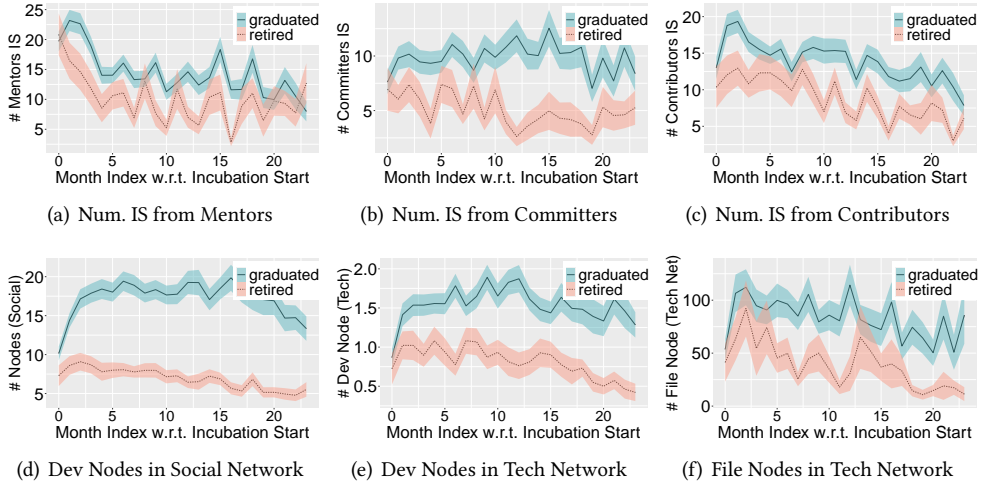


Fig. 3. The averaged monthly IS and ST variables between graduated projects and retired projects. On the top are the IS measures; On the bottom are ST measures. Shades indicate one st. error away from the mean. Month index 0 indicates the incubation starting month (color online).

Next, we see that the number of ISs in mentor emails decline for both graduated projects and retired projects before month 5, suggesting that ASF mentor activity may decrease after incubating projects work through the first steps of the incubation process.

Then, we visually identify an increasing trend of IS from mentors around month 6 for graduated while 5 for retired projects. One possible reason is the fact that mentors start helping projects when they are experiencing difficulties or downturns. It is consistent with ASF mentorship that during the early stage of the incubation, developers are required to make institutional-related decisions, e.g., voting for reports, discussing the ASF required licensing, and the community-related issues, and it is in these kinds of areas where mentors come to help.

On the Socio-Technical networks side, shown in Figure 3(d), for the first 6 months, we can see the graduated projects have a clear increasing trend of the number of nodes in social networks, while retired projects seems to be constant. We can see a slight decrease around month 10 to month 12 for both types of projects, suggesting 10 months might be a good timing for mentors to intervene/motivate their projects, if they are experiencing some difficulties.

RQ₂ Summary: We identify socio-technical and institutional signatures of OSS project evolution, and evidence that it differs between graduated and retired projects, and that these patterns can even be distinguished by institutional heuristic topics. On the institutional side, both graduated and retired projects have more stable institutional topics during their first 3 months. On the Socio-Technical network side, graduated projects keep attracting community over their first 6 months, while retired projects are unstable during their first 3 months.

5.3 Case Study: Association Between Institutional Governance and Organizational Structure

To communicate concretely how the institutional and socio-technical dimensions interact within ASFI ecosystem, we showcase four diverse instances of their mutual interrelationship.

Case A. In July 2011, the *HCatalog* project announced a vote for its first Release Candidate (RC), the first officially distributed version of its code ¹¹. Because a project's RC's reflect on the whole ASF, they require approval from the foundation after project contributors have given their approval. In preparation for the first vote, developers double-checked the installation process and reported missing files and features. This drove contributions to the code and documentation, e.g., release notes were added after being reported missing. The contributors then cast their votes. With four people's votes, the product was approved and a proposal was forwarded to Apache Incubator leadership for approval.

Case B. In December 2010, an independent developer emailed the *Jena* project community to share their idea for a new feature, and asking how to proceed toward contributing it ¹². Their query includes policy questions, such as whether they must obtain an Individual Contributor License Agreement (ICLA). A developer responds that the policy does not require an ICLA for the type of smaller contribution that the volunteer is proposing. The developer then guides the volunteer through established project processes for contributing to the code, including what mailing lists to use and how to submit their feature as a patch.

Case C. In December 2016, a developer in the *Airflow* project community raised concerns over the integration testing infrastructure offered by Apache, citing unnecessary obstacles it imposes on volunteer contributors¹³. The developer offers their own resources as an alternative, with the caveats that they will administer it and control access. This triggers a discussion on the technical merits of the developer's concerns, and a policy discussion as to whether ASF permits the use of unofficial alternative infrastructure options. Several developers conclude that a transition is technically advisable and institutionally sound, and the community transitions to the alternative integration testing framework.

Case D. In September 2015, the *Kalumet* project received a proposal that it be retired from ASFI after its code had been languishing for several months. Contributors agreed upon retirement almost unanimously. One contributor, identifying features of the project that could be of use to other ASF and ASFI projects, suggests distributing key parts of its functionality to other active projects. The retirement vote is ultimately followed by developer effort distributing *Kalumet's* assets.

These cases illustrate how institution-side policy discussion and sociotechnical-side project contributions interact, with developments on the artifact motivating policy discussions, and policy constraints steering developer effort. With longitudinal data on both institutional and socio-technical variables, we now transition to a quantitative investigation of these relationships.

5.4 RQ₃: Are periods of increased Institutional Statements frequency followed by changes in the project organizational structure, and vice-versa?

In the previous RQs, we conducted exploratory and qualitative studies of the IS extraction technology, and of IS and socio-technical variable changes over time. In this section, we investigate the temporal relationship between our measures of institutional governance and organizational structure, as OSS projects progress on their incubation trajectories. As predicted by contingency theory, our hypothesis is that during project evolution, developers and mentors must make time

¹¹<https://lists.apache.org/thread/p88lpxn3gtprc11xw20jbnrmp3f8fmpw>

¹²<https://lists.apache.org/thread/8dgbvjkwosbvvg6oj6ptyssn0m4rdto1>

¹³<https://lists.apache.org/thread/twsg2h72d2025jc6nc1ltzgbmwoxh7cx>

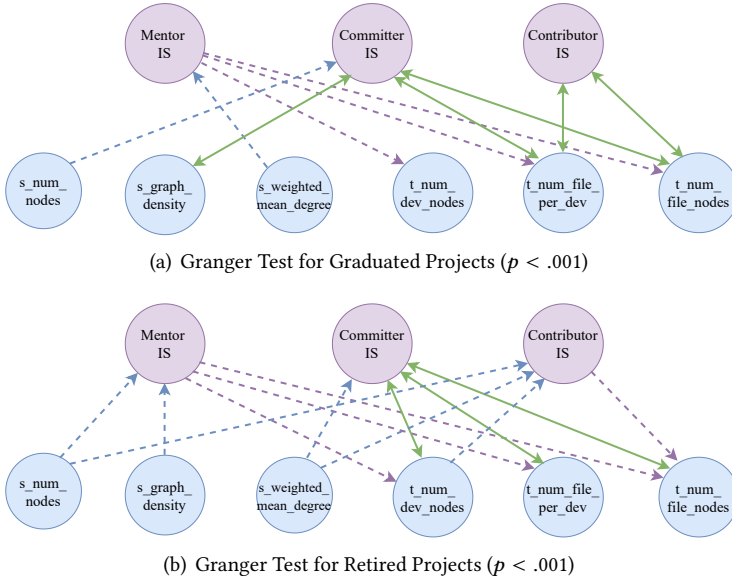


Fig. 4. The Granger Causality between Institutional Statements and Socio-Technical networks. The blue/purple directed links indicate Granger causality from ST/IS measures, respectively. A green bi-directional link indicates that there is two-way significant temporal relationship (p -val $< .001$). Graduated projects seem to have fewer links from ST variables to IS variables, suggesting a more unidirectional flow from institutional to sociotechnical changes in successful projects (color online).

for decisions related to their organizational structure, contingent on ASF required institutional arrangements and governance. That is, incubating projects change their organizational structure based on the institutional norms and rules being discussed, as required of them as potential new member of the ASF community. And vice versa, organizational changes can incite followup discussion about institutional processes. To test for RQ₃, here we use the pair-wise Granger causality test with lagged order of 2. We run the test for all pairs between the institutional statements and socio-technical variables, resulting in 36 separate tests for the graduated projects set and 36 for the retired ones. We adjust our p -values for multiple hypothesis testing to control false discovery rate, using the Benjamini-Hoschberg procedure [14]. We only consider significant with p -val < 0.001 .

The results are summarized in Figure 4, where a directed edge from node X to node Y indicates that X Granger-causes Y , i.e., change in X is the precursor to the change in Y . Also, as discussed in Section 4.6, Granger is not a complete test of causality, but does yield and effect and its directionality, although without effect size or sign.

We observe a large number, 31 (out of 72 total), of Granger-causal relationship between the measures of institutional governance and the organizational structure. Of those 31 Granger-causal relationships, 15 are from the graduated set and 16 from the retired set, and 8 of the relationships are shared between the sets. We conclude that there is a significant Granger-causality between changes in institutional governance discussions and organizational structure of the projects. We note 8 bidirectional relationships¹⁴, the remaining 15 are unidirectional.

¹⁴Bidirectional causality indicates a feedback of some sort. E.g., supply causes demand and demand, in turn, causes supply; a conversation with a child is another example, where a question causes a response which in turn incites another questions.

We look at graduated projects first. Interestingly, Figure 4, top, shows that the number of ISs from mentors, committers and contributors have effects on the technical network, and vice-versa for the latter two. Namely, IS from all roles (mentors, committers, and contributors) Granger-cause changes in the technical networks, i.e., on developer file productivity (`t_num_files_per_dev`), and total number of coding files changed (`t_num_file_nodes`) variables. Mentor IS, additionally, Granger-cause changes to number of developer (`t_num_dev_nodes`). This is consistent with ASFI expectations that a mentor's emails provide advice and engage people, and conversely, that a drop in engagement may elicit a mentor engagement. Mentors usually do not code, which is presumably why they Granger-cause but do not appear in feedback relationships with any of the technical network variables.

Notably absent, however, are links from mentor and contributor ISs into social network variables. Only committer ISs (bidirectionally) Granger-cause changes in the social network density, which, perhaps, simply indicates that ISs from committers induce substantial traffic in the social network, which in turn get committers to discuss policy and rules issues. We have observed situations where mentors are likely to interrupt the projects when the projects become less active (either socially or technically)¹⁵. On the other hand, it could also be that a mentor is reacting to some particular broader discussion among developers, e.g., one on a monthly report.

Together, the above tells a story of the importance to the technical networks of changes in any IS variable. Surprisingly, mentor IS changes are not as consequential to the social network, seemingly at odds with the ASF community-first goals. Thus, there may be room to enhance community engagement with mentors and vice-versa.

RQ₃ Summary: In both graduated and retired projects, there are no inputs from the IS into the social network variables, even though there are IS inputs into all technical network variables. Retired projects exhibit less bidirectionality between ST and IS variables. Finally, and interestingly, among retired projects there are causal inputs into contributor ISs from both the social and technical variables. This is not the case for the graduated projects.

6 DISCUSSION

In this study, we use individual institutional prescriptions, Institutional Statements (IS), and the Socio-Technical (ST) network features to reason about OSS project sustainability. OSS projects are a form of digital public goods which, like other public goods (e.g., water, forest, marine, etc.), can be subject to degradation due to over-harvesting, e.g., in the form of free-riders who take advantage of OSS but do not contribute to the required resources for development and maintenance of the software. Ostrom's work illuminated the fact that many communities avoid the dreaded 'Tragedy of the Commons', and other collective action problems, through the hard work of designing and implementing self-governing institutions. In that context, the ASF is a nonprofit foundation that, through its incubation program, encourages nascent OSS projects to follow some ASF-guided operational-level rules or policies around their self-governance. The OSS projects that join the ASF incubator trade some of the freedom of unlimited institutional choice in exchange for incubator resources that increase their chances of enduring the collective action problems that characterize OSS development [38], and becoming sustainable in the long run.

We found that in the ASF Incubator, the amount of institutional statements and levels of socio-technical variables are associated with projects graduation outcome, suggesting that the measures of institutional governance and organizational structure can signal information on sustainability. In

¹⁵An example of mentor interrupting project *warble*: <https://lists.apache.org/thread/x6h8pzhmfwtty354ml1xm9sylq4y5r7l>

particular, in RQ₁, the Mann-Whitney U test shows that the graduated projects have significantly more ISs from all three types of participants: committers, contributors, and project mentors than retired projects. This, presumably, is indicative of more active or intentional self-governance. In theoretical and empirical work on commons governance, it is well documented that getting self-governing institutions ‘right’ is hard work and takes time and effort [33]. This is consistent with a narrative that participants in graduated projects debate and work harder on their project’s operational-level institutional design.

Recent work has shown that ASFI graduate and retired projects have sufficiently different socio-technical structures [49], so that graduation can be predicted early on in development at 85+% accuracy. The results in RQ₂, show that, for the first 3 months of incubation, developer nodes in the social networks of graduated projects increase at a higher rate (means increase from 10.1 to 17.1, and from 7.3 to 9.1 for graduated and retired projects, respectively), suggesting graduated projects were able to keep developers contributing more actively or recruit more new members. On the other hand, for the first 3 months, we also found that the amount of Institutional Statements by mentors increases in graduated projects, and decreases in retired projects (from 19.7 to 22.7 vs 22.6 to 14.6, for graduated and retired projects, respectively), suggesting that the initial help from project’s mentors is of importance.

To further study the effects of ISs, we performed a deep-dive into IS topics. We found the topics of institutional-relevance in the graduated projects differ from those of the retired projects, specifically, we find that the topic of documentation (topic 7) in graduated projects is more prevalent than in retired projects. On the other hand, we found that the topics of mentorship (topic 11) of retired projects are significantly higher than retired projects, signaling that the retired projects might be struggling during the incubation. Combined with the fact that there are more developer nodes in both the social and technical networks, together the findings suggest that graduated projects have more capacity and energy to attend to non-coding issues, like documentation, than retired projects do. However, even among graduated projects there is still diversity in the institutional statements. Thus, as predicted by contingency theory, as well as Ostrom’s theory of institutional diversity [34], a one-size-fits-all solution to a successful trajectory toward sustainability is not likely. Instead, future work should focus on gathering larger corpora of data, to be able to resolve individual or small-group differences in sustainable projects.

Our framework allowed us to combine the IS and STS structures and study them together over time. With it, in RQ₃, we found two-way, causal correlations between socio-technical variables and ISs over time, arguably indicating that OSS project socio-technical structure and their governance structure evolve together, as a coupled system. In addition, our methods point to a way to study possible interventions in underperforming projects. Specifically, the finding that in retired projects there are bi-directional links from committer’s ISs to all three features of technical networks (i.e., *t_num_dev_nodes*, *t_num_file_per_dev*, *t_num_file_nodes*), suggest that increase in committer’s IS are interleaved with changes in features of the socio-technical networks.

As for the design implications, in addition to the current categories of mailing lists in ASF incubator (e.g., ‘commit’, ‘dev’, ‘user’, etc.), there can be a benefit to creating a separate mailing list, for institutionally-related discussions to help committers (and also for mentors and contributors) participate faster in those discussions in a timely manner. This could be made more useful using technology for self-monitoring, with which project participants could monitor a project’s digital traces and discussions in order to more quickly react to episodic events. Some such tools have already been created for socio-technical networks in ASFI projects [35], and could be extended to include ISs as well. Such tools can help identify entry points and targets for interventions, whereby underperforming projects could be leaned on, internally or externally, via rules or advice to adjust their trajectories.

Contributions to Institutional Analysis and Socio-technical System Theory. Making a full circle, our findings also point to ways in which the theories we started from can be refined or extended. We find, in Sect. 5.4, evidence that the features of OSS projects' socio-technical systems co-change together with the amount of Institutional Statements in them, and that the co-change relationships are sparse. This evidence of co-change implies that OSS project structure and their governance form a (loosely) coupled system. From a controllability point of view, a dynamic coupled system refines Smith et al.'s mechanistic binary notion of 'inside' and 'outside' interventions [42].

Our findings also suggest that for OSS projects, adopting additional rules and norms (e.g., by joining ASFI) can be worth the loss of some freedoms, as the Institutional Statements (Sect. 5.2, 5.3, 5.4) seem to serve to organize the project's actions and discussions, as predicted by Siddiki et al. [41] and Crawford and Ostrom [10]. Thus, our findings tie in into, and potentially extend the Institutional Analysis Design (IAD) view, suggesting that the feedback between the socio-technical system structure and institutional governance analysis is sufficiently direct and significant, and should be considered as unitary in further studies.

More practically, our institutional statement predictor, although still a work in progress, can effectively predict atomic elements of self-governance. As such, it can be used as a tool to provide quantitative data for applying institutional analysis and design (IAD) more generally, e.g., to OSS projects that are outside of ASF, or self-governed systems with public documents and discussion forums.

7 THREATS TO VALIDITY

First, our commit and email data is from only hundreds of projects ASF incubator projects. Thus, generalizing the implications beyond ASF, or even beyond the ASF Incubator projects carries potential risks, for example, OSS projects in other incubator programs may not have mentors. Expanding the dataset beyond ASF incubator, e.g., with additional projects from other OSS incubator program could lower this risk. Second, we do not consider communication channels other than the ASF mailing lists, e.g., in-person meetings, website documentation, private emails, etc. However, ASF mandates the use of the public mailing lists for most project discussion, a policy that ensures particularly low risk of missing institutional or socio-technical information¹⁶. Annotations of the Institutional Statements (IS) can be biased by individual annotators, while we gave the annotators sufficient training and a reference documentation which lowers the risk. We expect the performance of the classifier as we increase the size of the training set and better incorporate contextual information, and we plan in the future to distinguish types of ISSs. In OSS projects, developers may use their different emails or aliases, which in turn complicates the identification of distinct developers, while assigning and insisting using an unique apache.org domain email address reduces such risks¹⁷. Finally, as noted in Sect. 4, there are likely cases where OSS projects that have retired from the ASF Incubator program still go on to become sustained over time. In these instances, some OSS projects entering the ASFI may simply not be a good fit to the ASF culture and institutional requirements or policies and ultimately retire as a result. In this paper we explicitly use graduation as a measure of sustainability given that this is an ultimate goal of the ASFI – to create projects that can indeed be sustainable. But we want to recognize the point that few retired projects still could become sustainable following a different path than association with ASF.

¹⁶The Apache Way: <http://theapacheway.com/on-list/>

¹⁷ASF committer emails: <https://infra.apache.org/committer-email.html>

8 CONCLUSION

Understanding why OSS projects cannot meet the expectations of nonprofit foundations may help others improve their individual practice, organizational management, and institutional structure. More importantly, understanding the relationship between institutional design and socio-technical aspects in OSS can bring insights into the potential sustainability of such projects. Here we showed that quantitative network science features can capture the organizational structure of how developers collaborate and communicate through the artifacts they create. Combining the two perspectives, socio-technical measures and institutional analysis, we leverage the unique affordances of the Apache Software Foundation’s OSS Incubator project to extend modeling of OSS project sustainability, leveraging a novel longitudinal dataset, a vast text and log corpus, and extrinsic labels for the success and failure of project sustainability. Future work is needed to offer validation of this or similar strategies experimentally.

ACKNOWLEDGEMENTS

We are grateful to the National Science Foundation for funding this project, under GCR Grant #2020751 and #2020900. We greatly thank the reviewers for their constructive comments.

REFERENCES

- [1] BARCLAY, D. W. Interdepartmental conflict in organizational buying: The impact of the organizational context. *Journal of Marketing Research* 28, 2 (1991), 145–159.
- [2] BENKLER, Y. *The wealth of networks*. Yale University Press, 2008.
- [3] BIRD, C., GOURLEY, A., DEVANBU, P., GERTZ, M., AND SWAMINATHAN, A. Mining email social networks. In *Proceedings of the 2006 international workshop on Mining software repositories* (2006), pp. 137–143.
- [4] BIRD, C., NAGAPPAN, N., GALL, H., MURPHY, B., AND DEVANBU, P. Putting it all together: Using socio-technical networks to predict failures. In *2009 20th International Symposium on Software Reliability Engineering* (2009), IEEE, pp. 109–119.
- [5] BIRD, C., PATTISON, D., D’SOUZA, R., FILKOV, V., AND DEVANBU, P. Latent social structure in open source projects. In *Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of software engineering* (2008), pp. 24–35.
- [6] BLOMQUIST, W., ET AL. *Dividing the waters: governing groundwater in Southern California*. ICS Press Institute for Contemporary Studies, 1992.
- [7] CHEUNG, Y.-W., AND LAI, K. S. Lag order and critical values of the augmented dickey–fuller test. *Journal of Business & Economic Statistics* 13, 3 (1995), 277–280.
- [8] COHAN, A., BELTAGY, I., KING, D., DALVI, B., AND WELD, D. S. Pretrained language models for sequential sentence classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (Hong Kong, China, 2019), Association for Computing Machinery, p. 3693–3699.
- [9] COOKE-DAVIES, T. The “real” success factors on projects. *International journal of project management* 20, 3 (2002), 185–190.
- [10] CRAWFORD, S., AND OSTROM, E. A grammar of institutions. *American Political Science Review* 89, 3 (1995), 582–600.
- [11] DEVLIN, J., CHANG, M.-W., LEE, K., AND TOUTANOVA, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [12] DUCHENEAUT, N. Socialization in an open source software community: A socio-technical analysis. *Computer Supported Cooperative Work (CSCW)* 14, 4 (2005), 323–368.
- [13] DUMITRESCU, E.-I., AND HURLIN, C. Testing for granger non-causality in heterogeneous panels. *Economic modelling* 29, 4 (2012), 1450–1460.
- [14] FERREIRA, J., AND ZWINDERMAN, A. On the benjamini–hochberg method. *The Annals of Statistics* 34, 4 (2006), 1827–1849.
- [15] FISCHER, G., AND HERRMANN, T. Socio-technical systems: a meta-design perspective. *International Journal of Sociotechnology and Knowledge Development (IJSKD)* 3, 1 (2011), 1–33.
- [16] FLEISCHMAN, F., LOKEN, B., GARCIA-LOPEZ, G., AND VILLAMAYOR-TOMAS, S. Evaluating the utility of common-pool resource theory for understanding forest governance and outcomes in indonesia between 1965 and 2012. *International Journal of the Commons* 8, 2 (2014).
- [17] FRISCHMANN, B., MADISON, M., AND STRANDBURG, K. *Governing Knowledge Commons*. Oxford University Press, 2014.
- [18] GONZÁLEZ-BARAHONA, J. M., LOPEZ, L., AND ROBLES, G. Community structure of modules in the apache project. In *Proceedings of the 4th International Workshop on Open Source Software Engineering* (2004), IET, pp. 44–48.

- [19] GRUBY, R. L., AND BASURTO, X. Multi-level governance for large marine commons: politics and polycentricity in palau's protected area network. *Environmental science & policy* 33 (2013), 260–272.
- [20] HARDIN, G. The tragedy of the commons: the population problem has no technical solution; it requires a fundamental extension in morality. *science* 162, 3859 (1968), 1243–1248.
- [21] HERRMANN, T., HOFFMANN, M., KUNAU, G., AND LOSER, K.-U. A modelling method for the development of groupware applications as socio-technical systems. *Behaviour & Information Technology* 23, 2 (2004), 119–135.
- [22] HESS, C., AND OSTROM, E. *Understanding knowledge as a commons: From theory to practice*. JSTOR, 2007.
- [23] HISSAM, S., WEINSTOCK, C. B., PLAKOSH, D., AND ASUNDI, J. Perspectives on open source software. Tech. rep., Carnegie Mellon Univ Pittsburgh PA - Software Engineering Inst. 2001.
- [24] JOBLIN, M., AND APEL, S. How do successful and failed projects differ? a socio-technical analysis. *ACM Trans. Softw. Eng. Methodol.* (dec 2021).
- [25] JOSLIN, R., AND MÜLLER, R. The impact of project methodologies on project success in different project environments. *International Journal of Managing Projects in Business* (2016).
- [26] LEHTONEN, P., AND MARTINSUO, M. Three ways to fail in project management and the role of project management methodology. *Project Perspectives* 28, 1 (2006), 6–11.
- [27] LINDKVIST, E., BASURTO, X., AND SCHLÜTER, M. Micro-level explanations for emergent patterns of self-governance arrangements in small-scale fisheries—a modeling approach. *PLoS one* 12, 4 (2017), e0175532.
- [28] LOPEZ, J. H. The power of the adf test. *Economics Letters* 57, 1 (1997), 5–10.
- [29] NARDUZZO, A., AND ROSSI, A. The role of modularity in free/open source software development. In *Free/Open source software development*. Igi Global, 2005, pp. 84–102.
- [30] OLSON, M. The logic of collective action [1965]. *Contemporary Sociological Theory* 124 (2012).
- [31] O'REILLY, T. Lessons from open-source software development. *Communications of the ACM* 42, 4 (1999), 32–37.
- [32] OSTROM, E. *Governing the commons: The evolution of institutions for collective action*. Cambridge university press, 1990.
- [33] OSTROM, E. *Understanding institutional diversity*. Princeton university press, 2009.
- [34] OSTROM, E., JANSSEN, M., AND ANDERIES, J. Going beyond panaceas. *Proceedings of the National Academy of Sciences* 104, 39 (2007), 15176–15178.
- [35] RAMCHANDRAN, A., YIN, L., AND FILKOV, V. Exploring apache incubator project trajectories with apex. In *2022 IEEE/ACM 19th International Conference on Mining Software Repositories (MSR)* (2022), IEEE, p. Accepted.
- [36] ROPOHL, G. Philosophy of socio-technical systems. *Techné: Research in Philosophy and Technology* 4, 3 (1999), 186–194.
- [37] SANDLER, T. *Global collective action*. Cambridge University Press, 2004.
- [38] SCHWEIK, C. M., AND ENGLISH, R. Tragedy of the foss commons? investigating the institutional designs of free/libre and open source software projects. *First Monday* (2007).
- [39] SCHWEIK, C. M., AND ENGLISH, R. C. *Internet success: a study of open-source software commons*. MIT Press, 2012.
- [40] SEN, A., ATKISSON, C., AND SCHWEIK, C. M. Cui bono: Do open source software incubator policies and procedures benefit the projects or the incubator? *Available at SSRN* (2021).
- [41] SIDDIKI, S., HEIKKILÄ, T., WEIBLE, C. M., PACHECO-VEGA, R., CARTER, D., CURLEY, C., DESLATTE, A., AND BENNETT, A. Institutional analysis with the institutional grammar. *Policy Studies Journal* (2019).
- [42] SMITH, A., AND STIRLING, A. Moving outside or inside? objectification and reflexivity in the governance of socio-technical systems. *Journal of Environmental Policy & Planning* 9, 3-4 (2007), 351–373.
- [43] SURIAN, D., TIAN, Y., LO, D., CHENG, H., AND LIM, E.-P. Predicting project outcome leveraging socio-technical network patterns. In *2013 17th European Conference on Software Maintenance and Reengineering* (2013), IEEE, pp. 47–56.
- [44] TRIST, E. *The evolution of socio-technical systems: A conceptual framework and an action research program*. Ontario Ministry of Labour, 1981.
- [45] TURNER, J. R., AND MÜLLER, R. Communication and co-operation on projects between the project owner as principal and the project manager as agent. *European management journal* 22, 3 (2004), 327–336.
- [46] ŘEHŮŘEK, R., SOJKA, P., ET AL. Gensim—statistical semantics in python. *Retrieved from gensim.org* (2011).
- [47] WEARNE, S., AND STANBURY, A. A study of the reality of project management: Wg morris and gh hough. john wiley, uk (1987)£ 29.95, isbn 0471 915513 pp 295. *International Journal of Project Management* 7, 1 (1989), 58.
- [48] WU, J., GOH, K.-Y., AND TANG, Q. Investigating success of open source software projects: A social network perspective. *ICIS 2007 Proceedings* (2007), 105.
- [49] YIN, L., CHEN, Z., XUAN, Q., AND FILKOV, V. Sustainability forecasting for apache incubator projects. In *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (New York, NY, USA, 2021), Association for Computing Machinery, p. 1056–1067.
- [50] YIN, L., ZHANG, Z., XUAN, Q., AND FILKOV, V. Apache software foundation incubator project sustainability dataset. In *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)* (2021), IEEE, pp. 595–599.
- [51] YU, H., AND YANG, J. A direct lda algorithm for high-dimensional data—with application to face recognition. *Pattern recognition* 34, 10 (2001), 2067–2070.