First Monday
PEER-REVIEWED JOURNAL ON THE INTERNET

Open source software field research: Spanning social and practice networks for re-entering the field

by Sean Goggins, Matt Germonprez, Kevin Lumbard, James Howison, Karthik Ram, and Caifan Du

Abstract

Sociotechnical research increasingly includes multitude of social and practice networks that emerge from large-scale sociotechnical infrastructure, including the infrastructure for building open source software. This paper addresses the investigation of these numerous networks as advantageous for researchers. It provides a methodological synthesis focusing on how researchers can best span adjacent social and practical networks during engaged scholarship. For example, spanning between corporate open source networks and scientific open source networks. Specifically, we describe practices and artifacts that aid movement from one social and practice network within a more extensive technical infrastructure to another. To surface the importance of spanning networks, we incorporate a discussion of social capital and the role of technical infrastructure in its development for sociotechnical researchers. We characterize a five-step process for spanning social and practice networks during engaged scholarship: commitment, context mapping, jargon competence, returning value, and bridging. We then present our experience studying corporate open source software projects and the role of that experience in accelerating our work in open source scientific software research as described through the lens of social capital. Based on our analysis, we offer recommendations for engaging in work in adjacent social and practice networks that share a technical context and a discussion of how the relationship between social and technically acquired social capital is a missing but critical methodological dimension for research on large-scale sociotechnical infrastructures.

Contents

1. Introduction

The domains of study in social computing, including open source software, social media, crowdsourcing, and citizen science, are increasingly centered on large-scale platforms with vague boundaries between the associated sub-groups of people and their constituent networks. Understanding how and to what extent knowledge transfers between networks in large domains occur is critical for understanding how information, knowledge, and practices are shared between different forms of technologically mediated engagement (Introne and Goggins, 2019). Social computing research is often centered on a single technology and bounded by a particular study's social networks and subnetworks. To advance the field, it is necessary to understand how social capital developed in one socio-technical network can be leveraged as an entryway into another network for scholarship with broader academic and applied outcomes.

Boundaries between networks within open source software, for example, are defined by practices of technology use and the movement of knowledge between networks (Harrison and Tatar, 2008). Yet, there are limited examples of social computing research that span between social computing networks, with notable exceptions in crisis informatics (Hughes and Palen, 2009; Palen and Liu, 2007; Palen, *et al.*, 2010, 2009; Sarcevic, *et al.*, 2012; Starbird, *et al.*, 2010; Starbird and Palen, 2012, 2011; Sutton, *et al.*, 2008; Vieweg, *et al.*, 2010). In response, we argue in this paper that engaged scholarship (Van de Ven, 2007) along with the analysis of technical artifacts can improve how researchers can conduct research across network boundaries (Introne, *et al.*, 2012; Introne and Goggins, 2019; Introne, *et al.*, 2020, 2016; Introne and Goggins, 2015, 2012; Goggins, *et al.*, 2013a; Goggins and Petakovic, 2014; Howison and Bullard, 2016). In this paper, we contribute a case of network spanning research as an exemplary process for discovering and navigating research across network boundaries in large-scale sociotechnical systems in open source software. We illustrate and argue for the unattended role of social capital as an essential dimension of engaged scholarship that incorporates social computing artifacts and data. Our contribution draws on our experiences entering the field (Chughtai and Myers, 2017) as we cross boundaries to study and engage with different networks working on open source software.

Star (1990) identified the boundary-crossing experience for sociotechnical researchers as a choice between being part of a network (singular) or part of networks (plural). In many cases, engaged scholarship takes on significant long-term investment as researchers simultaneously engage deeply with multiple networks (Van de Ven, 2007). These time commitments draw scholars away from their primary network of practice in sociotechnical systems research. Our academic communities impart a kind of invisible work regarding the standards for accepting work. When working across networks, the likelihood of experiencing marginality grows. This experience is especially true if a few scholars build bridges between any two scholarly or practice domains, a challenge that Star (1990) and others (Asdal, *et al.*, 2007) noted as critically important.

The area of concern our research centers on is open source software (OSS), which has seen tremendous growth in the last decade. The landscape of OSS is vast, and research access to one network within this landscape (*i.e.*, corporate OSS) no longer necessarily includes access to another network (*i.e.*, scientific OSS) (Germonprez, *et al.*, 2018). Here, Star's (1990) tension between networks and the resulting marginalization for the scholar becomes apparent through the common conceptualization of OSS as a single context. This oversimplification of open source software in sociotechnical and social computing discourse exists in HCI's larger tradition of studying complex sociotechnical phenomena from the perspective of being "in here" (in our scholarly network) and, simultaneously "out there" (outside our scholarly network) (Taylor, 2011). Regardless, engagement in outside networks is called for as researchers build research programs that account for the practices and constructs across distinct networks.

OSS scholarship has a long arc of influence on culture (Kelty, 2008), often at the intersection between the ideals of open culture and its reality (Coleman, 2013). Kelty (2008) described, in depth, the state of those intersections in the middle of the 2000s, and Coleman (2013) took a deep dive into the practices and ideals of the Debian community during the same period. Dunbar-Hester (2020) provided more contemporary observations centering on open source software as a mechanism for increasing diversity and addressing problems of global political inequality through new technology, pointing out that new technologies often promise liberation but seldom advance the cause. Stemming from prior open source work, we know that OSS does not exist as a single network but exists uniquely as a combination of corporate, scientific, public health, academic, and research OSS networks. Techniques for sharing knowledge between OSS networks are critically important for all and our paper proposes methods for boundary spanning across open source networks to understand OSS better and put into practice how we express OSS as a multi-networked system.

This paper argues that each corner of OSS is a distinct network. Each corner of OSS may require a long-term investment by default because their sociotechnical networks are entirely different, something obscured because much technology and underlying philosophy remain the same. In this paper, we frame the social capital developed through long-standing engagement with one network of OSS, corporate open source, as having technical and practical utility to contributors in another network of OSS, open source scientific software (Star, 1999). We propose that insights gained, instruments defined, and tools built to garner an understanding of one OSS network may afford a researcher some social capital within an adjacent OSS network, as our case describes. The effects can be shorter periods of building relationships and conducting research in an adjacent OSS network. However, we anticipate that efforts in our prior contexts do not give us *carte blanche* access to new networks (Tajfel, 1982b). As we enter a new network, we return to Star (1990), who observed tensions within our expanding field of view.

We propose that an OSS researcher can benefit from the development of social capital (Coleman, 1988) within different OSS networks (corporate, scientific, and others), providing an ability to contribute to our scholarly outputs across networks. The following sections describe prior work illustrating the role of two types of social capital and explain how each is conceptually vital for engaged scholarship in our context of open source software. We then demonstrate how prior engaged scholarship alludes to social capital as a methodological concern and pull those discrete insights together, much like Agre consolidated insights about the missing social dimension in long-standing political theory (Agre and Schuler, 1997; Agre, 2004). We then ground our argument that social capital may be missing in boundary-spanning research by describing our ability to move between networks, as experienced in our move from corporate to scientific open source software. We conclude with brief recommendations for institutionalizing multi-disciplinary research into the networks of open source software, which from a distance may appear as a singular, monolithic field of sociotechnical materialities (Leonardi, *et al.*, 2012) that can be studied without considering their social, contextual, and network-specific elements.

2. Prior literature

Social capital applied methodologically, is much like moving from one home in one city to a new home in a new city; the transition is never smooth. When a person moves, they must make new friends and rearrange furniture often. The function of the kitchen reveals itself slowly over time, and it takes a year to learn when and where the sun shines in a garden. For large-scale sociotechnical research, the same fundamental skills and essential infrastructure (Star and Ruhleder, 1996; Bateson, 2000) apply (*i.e.*, lounging, cooking, and gardening), but specifying how those skills and objects are enacted and used in different homes and different neighborhoods is a metaphor for how researchers might approach spanning networks. Everything looks the same if observed from afar yet is experienced differently by the people involved depending on the context. Moving requires time to identify differences and similarities that help people adjust and earn the social capital necessary to, for example, borrow a neighbor's shovel. Accruing social capital can also be a cornerstone of engaged scholarship. It takes time to gather and is easy to lose, but it can be leveraged into research productivity.

It is safe to say that a researcher accrues social capital in a new sociotechnical context by contributing to the things that members of the context are working with — not in the form of ideas or suggestions but in the form of documented and applied contributions. It is like how, in the moving metaphor, shoveling our neighbor's walk helps to build the social capital we need to get a "yes" when we ask to borrow their shovel. As an engaged scholar, social capital can be built within a targeted context and with the greater research community as the credibility and validity of research insight are directly connected to social capital (Robert, *et al.*, 2008). This social capital is the sum of the researchers' experiences, network ties, technical skills, and resources (Bozeman and Corley, 2004). Accruing social capital emphasizes directly helping people with similar backgrounds and focal concerns that, over time, can provide privileged access that enables a deep understanding of how a community or organization functions.

2.1. The role of common infrastructure in obscuring boundaries

Common technical and practice infrastructure (Star, 1999; Star and Ruhleder, 1996) can make boundaries within open source software challenging to see. From a distance, most people with a passing interest can observe that nearly all open source software today shares a common technical infrastructure that includes Git-based version control tools for managing distributed contributions to software repositories, issue trackers, and varying levels of documentation (McDonald and Goggins, 2013). Most people, however, even within open source, cannot state where, precisely, the boundaries between any two networks of open source software exist. The boundaries are perhaps more observable through structured open source software institutions like the Linux Foundation, which generally hosts and supports open source projects in the corporate context. Yet, in other contexts, like open source scientific software, the boundaries can be slightly vague and include some ambivalence toward boundary creation. Yet, routine work across the common technical infrastructure of open source software leads to questions about where one OSS network begins and another ends. Tools like Git, Git repository platforms like GitHub or GitLab, issue trackers, and mailing lists obscure distinctions. For example, scikit-learn is a Python library for machine learning that leverages many other libraries, like matplotlib and NumPy [1]. It is tempting to characterize scikit-learn's overlap as a trait that makes it a "boundary object" (Star and Griesemer, 1989) or an object that exists on the boundary between OSS networks. Lee (2007), and others have suggested this is an oversimplification in some cases and that boundaries between networks can be vague and subject to constant negotiation through what Lee (2007) described as *boundary negotiating artifacts*.

The well-established common technical and practice infrastructure across open source software is highly ordered and broadly distributed. It has all the critical marks of routine work: self-explanation, inclusion, compilation, structuring, and borrowing (Star and Griesemer, 1989). Nonetheless, open source software researchers question, with some regularity, the boundaries of the networks they choose to define and explain. Therefore, we will stop short of characterizing our questions as *boundary identification artifacts* and instead focus our attention on how social capital built by people in one open source network (*i.e.*, corporate OSS) and applied as significant for bridging engaged scholarship into a new and distinct open source network (*i.e.*, scientific OSS).

2.2. Bridging and bonding social capital

The construct of social capital holds two forms in prior literature: bonding and bridging (Ren, *et al.*, 2007). Bonding social capital accrues based on the researcher's engagement in a domain, and the bond deepens as a researcher and a community get to know one another over time. Bridging social capital is developed through a need within a bonded community for outside relationships, expertise, or perspectives. At first, there is less clarity about how the new community can apply a researcher's knowledge. Bridging social capital from one community to another can provide the foundation for developing bonding social capital in that new community (Robert, *et al.*, 2009). As such, bridging social capital means that without first-hand knowledge, social capital from one network can provide sufficient bridging social capital for introduction into a new network. However, it will be replaced over time by observational knowledge and new bonding social capital (Robert, *et al.*, 2009).

Social capital is not fungible like currency (Coleman, 1988) but is a critical dimension of engaged scholarship. Researchers earn, develop, and leverage social capital during long-term, committed research engagements. Effectively balancing the accumulation and leveraging of social capital combines the bridging and bonding types of social capital in different ways over time. Commitment to and the development of bonding social capital in one community comes first, and opportunities for the outcomes of that research to evolve into a bridging form of social capital come second. Bonding social capital provides access to rich data and insights within one community and enables bridging social capital necessary to connect with adjacent communities. To the extent our methodological approaches enable adjacent bridging across contexts, we are more likely to avoid the disconnection of scholarly insight from practice.

3. Methods: Bridging from corporate open source to scientific open source

Getting epistemologically valid, domain-specific access is a piece of hidden work in social computing scholarship, which we seek to make visible. While doing that, our research uses a type of engaged scholarship specified through soft systems methodology (Checkland and Poulter, 2010, 2006) and trace ethnography (Geiger and Ribes, 2011; Goggins and Mascaro, 2013; Goggins, *et al.*, 2013b). While conducting our research, we are attentive to transparency with our informants (Etherington, 2007) and the complexity and ambiguity of the OSS context (Clarke, 2003). We deploy strategies for ensuring that the perspective presented in our research is more resonant than dichotomous (Haraway, 1988), and reports from our engagements are not as though it's an outside place (Taylor, 2011), but instead ensuring we are reporting as authentic a view in the research that we produce (Haraway, 1988; Barad, 2007).

3.1. Our ethical guideposts for open source software engagement and social computing

Our work in corporate OSS is guided ethically by our values as researchers and people to be our best selves. Our work in this space has led us to contribute to numerous corporate open source projects, including co-founding and managing the CHAOSS project hosted by the Linux Foundation. Through our deep open source engagement, we recognize and strive to create humane documents and technologies [2], perhaps most clearly visible in documents and technologies that directly impact individuals. Fritz Lang's observation about the moral of his film, *Metropolis* (1927), that the mediator between the brain and the hand must be the heart, as underscored by Nardi and O'Day [3] is at the center of the information ecology we are building with our collaborators in OSS. Our respect for the potential vulnerabilities of individuals we work closely with has, to date, prohibited the use of real names and personal identifiers in our research publications. However, many individuals identify themselves as contributors to the open source project that we manage [4].

3.2. Moving from one network to another

We believe that research in complex sociotechnical systems contains "different stakeholders, with different worldviews, acting purposefully" in situations where demarcating boundaries is difficult [5]. This belief, as derived from soft systems methodology (Checkland and Poulter, 2006), conceptualizes a lens for understanding engaged scholarship as positioned within and across collaborative networks, composed of various stakeholders that drive actions leading to more effective practices from their point of view. As such, we approach our work with scientific OSS as an act of moving from one network to another. The results of our work in corporate OSS described below are a connection point to scientific OSS.

Our history of accruing bonded social capital necessary to be effective in the corporate open source context is deliberate. We recognized that institutions like the Linux Foundation, a registered trade organization, help coordinate between open source projects and the thousands of member technology firms under their umbrella. For example, the Linux Foundation originated to ensure the Linux Kernel would be consistent across distributions because market differentiating software runs on Linux, and managing different product builds for different Linux kernels would be costly. Our work with the Linux Foundation began in its early stages, and our recognized bonded social capital accrued is based on our prior contributions to open source software projects (Germonprez, *et al.*, 2012, 2011; Graves, *et al.*, 2016; McDonald and Goggins, 2013; McDonald, *et al.*, 2014).

Beginning with an NSF VOSS grant, one author from our research team progressively accrued their bonded social capital within the Linux Foundation toward more of the insider views that emerged through project-specific contributions and a series of studies that produced both scholarship and valuable insight for practice (Germonprez, *et al.*, 2016; Kendall, *et al.*, 2016). Work with the SPDX and FOSSology projects was beneficial for building a common language between the researcher and the respective communities.

Further, the researcher leveraged early open source engaged scholarship into the co-founding of the Linux Foundation brokered CHAOSS open source project to provide meaningful insights regarding the health of open source projects. Likewise, a second author from our research team leveraged their NSF VOSS grant and funding from the U.S. Navy and U.S. Department of Education to build bonded social capital in open source more broadly construed, and ultimately, the Linux Foundation. The researcher's multi-disciplinary and systematic integration of qualitative methods with trace data analysis accrued social capital with Linux Foundation member corporations through their ability to quickly develop and deploy metrics regarding open source community health in a corporate context. Collectively, our work in the corporate OSS network resulted in an accrual of bonded social capital, contributing to an ability to use standard open source technologies and philosophies to provide swift, helpful information and provides bridging social capital as we move into other networks (*i.e.*, scientific OSS).

3.3. Trace ethnography

As we sought to move from one network (corporate open source) to another (scientific open source), we used trace ethnography as a second method of inquiry (Geiger and Ribes, 2011), which aims to build context around the OSS trace data we gather by analyzing Git logs and Git Platform API'='s like those on GitHub and GitLab. These traces of individual interactions with and contributions to collections of repositories identified by informants bound each collection as a distinct OSS network. Trace data were triangulated with previously identified patterns in OSS project structure and tuned to align the estimation of connection between a contributor and a project with how our informants describe their engagement with projects during our prior work (Goggins, *et al.*, 2013a; Goggins and Petakovic, 2014). Trace ethnography provided a succinct, empirical validation that corporate and scientific OSS are manifest as distinct networks. To identify the network structure of each, we operationalized commit and platform message metrics as weighted connections between individuals and projects in two scientific OSS network types and one corporate OSS network. We expect that our described use of trace ethnography is applicable across OSS networks but that the results from that application will vary as described in Figure 1.

Our trace ethnography reflects our prior engaged scholarship efforts and a connection weighting method (Goggins, *et al.*, 2013a) that directly incorporates insights gleaned from our previous research. For example, our informants noted that code contributions reflect a more robust connection than platform messages around issues and merge requests. We wove this insight with trace data to value connections derived from code contributions more heavily than connections derived from messages. We also used the time distance between messages to decrease connection weight as the time distance becomes more significant. In addition, the message content was evaluated in our weighting, with topically and linguistically similar text in sequential messages being given more weight than less congruent messages (Dodsworth and Benton, 2019). Finally, we accounted for longitudinal patterns (Duxbury, 2022) by weighting actions from the most recent 18 months of activity more heavily than older interactions. The result was an expression of network centrality measures (Carrington, *et al.*, 2005) that produce visually succinct structural models of the communities identified within each network using a combination of the Leiden (Traag, *et al.*, 2019) and Louvain computational models (Murniyati, *et al.*, 2023).

3.3.1. Structural contrast of corporate and scientific OSS

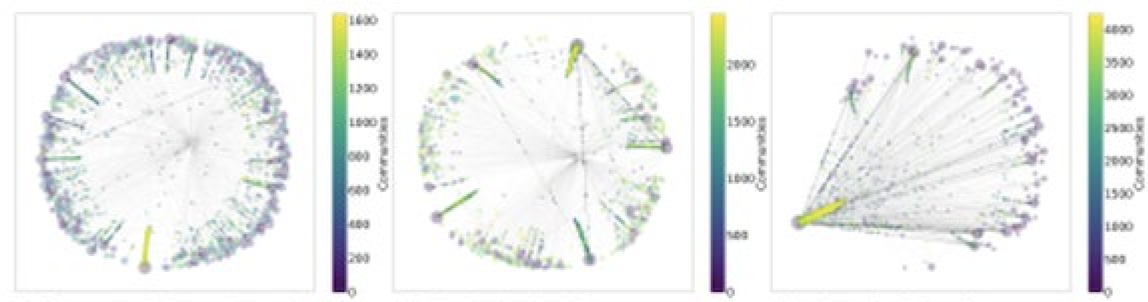
Reflexive, qualitative analysis of metrics, weighted using insight from research notes, illustrates structural differences between the networks. This method surfaced different patterns of structural organization, which, combined with our research notes, made it clear that there are, in fact, material differences in how each network's contributors engage with different repositories. This process is productive for us, as researchers, and engaging to our research sites because they, too, do not want to have statistics without underlying context. Figure 1 illustrates a doughnut-shaped structure for the corporate OSS network, a clock-shaped structure for a funded scientific OSS network, and a funnel-shaped structure for an organic scientific OSS network. Node colors identify individuals in distinct communities, with all repositories in "community 1" and purple. The connecting lines and sometimes linear alignment of contributors around a particular community indicate levels of dominance of one collection of projects over others in an OSS network.

Dozens of these communities are visible in the doughnut-shaped corporate OSS network.

In contrast, the clock-shaped funded scientific OSS network has five to six major community clusters, and the funnel-shaped organic scientific OSS network is visibly dominated by a singular community around a single, dominant project (in this case, Bioconductor). Larger versions of these sociograms can be examined in the Appendix. Other specific visual characteristics to note include:

1. In-degree centrality is reflected in node size. Since only the repositories have in-degree centrality above 1 in our data organization, they are the only nodes that appear larger.
2. The distance between nodes in each graph is determined by the similarity of contributor patterns toward each of the repositories from the three visualized networks.
3. The shapes of each sociogram reflect the structure of each expert-identified network.

These differences illustrate different practices and interconnections that define these networks and succinctly illustrate that the social and contribution structures of scientific and corporate OSS networks are empirically different. The individual and collective motivations and rewards of engagement in each network are distinct. As such, these differences in structure require different approaches and uses of social capital for engaged scholarship between networks.



(a) Corporate: Donut-shaped structure.

(b) Funded scientific: Clock-shaped structure.

(c) Organic Scientific: Funnel-shaped structure.

Figure 1: These are sociograms of the most prominent communities within a corporate (a), funded scientific (b), and naturally occurring scientific (c) open source software contribution networks. These differences in structure require different approaches for engaged scholarship, while the common central artifact of a software repository and motivation to understand the relative health and sustainability of individual repositories within each ecosystem frame the methodological importance and points of transferability of social capital from one network to the next.

Note: Larger version of Figure 1 available here.

4. Findings: Social capital in spanning networks

Our approach to working across OSS networks relies on our experience in helping open source communities and organizations understand the OSS health indicators (*i.e.*, success factors, risk, and sustainability) of projects they rely on. Failures of projects within open source software networks can have negative consequences for people and organizations relying on their output. Developing metrics and evaluating the sustainability of OSS projects in a corporate context enabled us to communicate similar concerns later in a scientific OSS network. However, entering the domain of scientific OSS without this experience would, ultimately, not be very different from entering a new research site. As such, we were centered on a long game of understanding and contributing to OSS in different networks, enabling a deeper understanding of OSS health across all of them through the revelation and explanation of shared and distinct practices. In each network, we focused on learning the language, contributing, and assuming leadership through a series of systematic steps. In this section, we highlight the approach by which we successfully moved from the corporate OSS network to the scientific OSS network. Finally, we reflect on the role of technical utility in our experiences and present a synthesized summary of how technical utility acted as a bridge between OSS networks that exist in technically parallel but sociotechnically orthogonal worlds.

4.1. Engaged scholarship to establish bonding social capital in corporate open source

Our understanding of corporate OSS accrued a deep knowledge of the social engagement and technical practices within that network. Through this work, our ability to provide insight and value to the corporate OSS communities we engaged with grew over time. This corporate engagement culminated in the creation of a new Linux Foundation open source project — the Community Health Analytics Open Source Software (CHAOSS) project. The CHAOSS project originated from an “unconference” session at the Open Source Leadership Summit 2017. The session, funded through a collaborative grant from the Alfred P. Sloan Foundation, started our deep investigation of OSS health.

The session revealed how corporations actively viewed and evaluated engagements with open source projects and how the health of these projects and the communities of people who build and maintain them were at the top of their minds. Following the session, CHAOSS was established at the Linux Foundation to combine efforts to understand the health of open source communities and create software tools to extract and visualize sustainability data. Members of our research team were instrumental in creating CHAOSS and remain active contributors to the project as board members, maintainers, and contributors.

Since its inception, the CHAOSS work of our research team has had a strong presence at open source conferences: Open Source Summit North America (2017, 2018, 2019, 2021, 2022, 2023, 2024), Open Source Summit Europe (2017, 2018, 2019, 2022, 2023, 2024), CHAOSScon Europe at FOSDEM (2018, 2019, 2020, 2022, 2023, 2024), CHAOSScon North America (2018, 2019, 2020, 2021, 2022, 2023, 2024), Open Source Leadership Summit (2018, 2019, 2021, 2022, 2023, 2024), and the Community Summit (2018, 2019, 2021, 2022, 2023). Attendance at CHAOSS-related sessions at these conferences and contributions to the CHAOSS project has continually grown. Interest has come from corporations, open source foundations, project maintainers, and open source project members, indicating that this topic is essential to many. This work applies traditional research methods to construct academic research outcomes (Germonprez, *et al.*, 2019, 2018), yet considerable time was also spent engaged in community building and community contributions. The CHAOSS project’s principle aims to make open source project health and sustainability more transparent and actionable for community members, providing direct value to the open source community members interested in understanding project sustainability.

The CHAOSS metrics developed for measuring open source community health and sustainability generated indicators describing how groups around any open source project function. For example, the CHAOSS project’s published metrics include indicators of how many different individuals contribute to the project over time through standard open source technical components as software commits, merge requests, issues, documentation, and communication in the forms of comments on each of those artifacts, as well as participation in mailing lists. We developed the Augur software to examine open source software projects, and we can quickly generate CHAOSS metrics for any software project or collection of projects. In this way, the infrastructure of the CHAOSS project affords us the technical foundation and bridging social capital outside the corporate OSS network, where our work originated.

Within the CHAOSS project, we deeply understand the social arrangements and technical practices manifest across corporate OSS as members, contributors, and researchers. Outside of the CHAOSS project, our capacity to understand adjacent open source networks, derived from understanding elemental tasks like issue creation and code creation, provided a bridge. While we have extensive sociotechnical expertise in our corporate OSS network, our ability to perfectly translate that expertise to a new OSS network was inhibited, not because of our technical skills but because of how identical OSS technologies are operationalized and socialized in scientific software practice. Our comprehension of social and technical phenomena is not zero in these new contexts (horizontally) because they use identical OSS technologies. Therefore, our move to new contexts requires a commitment to (1) effectively building bonded social capital by demonstrating technical utility, and then (2) adapting our sustainability expertise to account for the distinct patterns of engagement found in open source scientific software.

As directly crucial to our deepening understanding of social engagement and technical practices, in our time at the CHAOSS project, we found that making OSS health data transparent and actionable was applicable in the corporate open source network and other OSS networks. Open source community health is relevant to many different networks, and the tools and methods deployed to understand community health can provide insight across many different contexts, as illustrated in Figure 2 (Germonprez, *et al.*, 2019, 2018) CHAOSS provides utility to OSS stakeholders by defining a taxonomy of community health metrics. That same taxonomy is consistent, if not perfect, and the CHAOSS software, (Goggins, *et al.*, 2021), implements concrete analysis using CHAOSS metrics served as bridging social capital to cross into new OSS networks, with scientific OSS being a most prominent example.

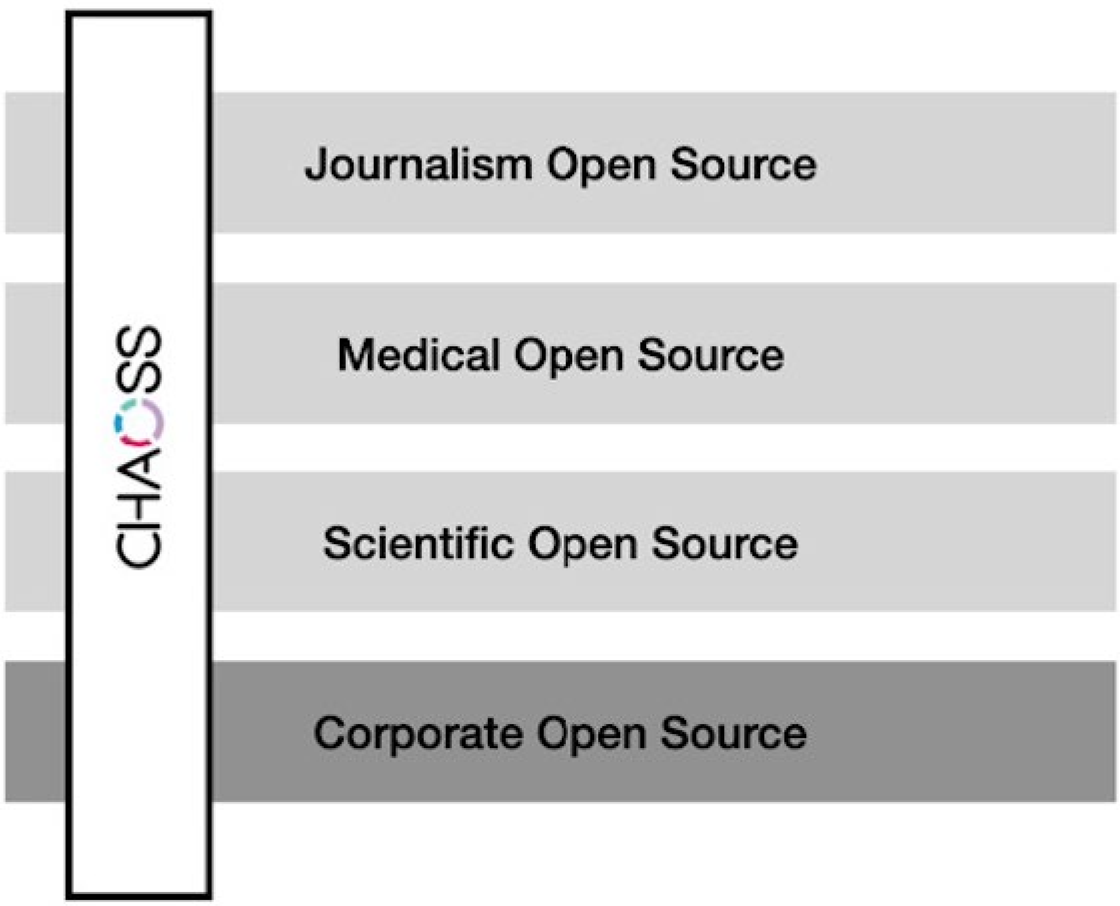


Figure 2: The various OSS networks to which the CHAOSS Project can provide insights.

4.2. Engaged scholarship to apply bridging social capital to scientific open source

To some extent, our research focus on open source across networks relies on the barely fungible currency of social capital (Coleman, 1988). Without first-hand social knowledge, bonding social capital from one network can introduce social capital in new networks (Ren, *et al.*, 2007). This initial social capital is bridging social capital. However, over time, bridging social capital must be replaced by bonding social capital as new observational knowledge is made available to community members (Robert, *et al.*, 2009). Scientific OSS exists at the intersection of four focal context areas — software development, open source communities, scientific research, and domain-specific areas (*e.g.*, biology). In working with scientific OSS, our research team brought bonding social capital from other networks (*e.g.*, corporate open source), providing us a point of entry. However, we knew we needed to understand this new, scientific OSS network and develop the language to ultimately contribute to the goals and knowledge of new communities of interest (*e.g.*, scientific OSS and domain-specific areas). Our bridging social capital accrued via CHAOSS was fleeting (Robert, *et al.*, 2009), and it needed to be replaced in this new context by bonding social capital to maintain and grow and establish ourselves in a new network.

It was only at this point where we began to see the role of both bonding social capital and bridging social capital (Coffé and Geys, 2007; Patulny and Svendsen, 2007; Goggins, *et al.*, 2013a; Ren, *et al.*, 2007) to construct the relationships necessary for effective work in new networks. Our social capital, developed through bonds in corporate OSS, granted us some degree of bridging social capital into scientific OSS because of a shared goal of understanding open source community health. Our bridging social capital was derived, in part, from the dimensions of our long-term engaged scholarship with corporate OSS and emerged from an opportunity to examine scientific OSS projects and their expressed needs, vision, and heuristic definitions of open source health and sustainability.

4.3. Five stages of building social capital in the scientific OSS network

First, researchers must be committed to orienting themselves to understand adjacent networks (Rich, 2013). It is reasonable for researchers to engage in programs that do not require much investment in accruing domain knowledge and still produce high-quality research. Knowledge of new networks demands intentionality motivated by a researcher's long-term goals. While each personal research program informs the size of this investment, some significant investment is necessary because building knowledge and credibility in an adjacent network depends on two parties knowing each other well over time (Lacity, *et al.*, 2009). Open source engagement can mean joining an open source project as a newcomer — downloading source code documentation, following the community discussion on mailing lists and issues forums, and posting questions and comments relevant to the community (Jensen, *et al.*, 2011). As relational ties strengthen, bonding social capital is accrued through a commitment to the norms, values, and beliefs of a community (Chua, *et al.*, 2012).

Second, a researcher must accurately map the new network to accrue an understanding of the specific idiosyncrasies necessary for engaging with nuances found in any new setting. Context mapping is not an exercise returned to the network in question. Instead, it is an activity for a researcher alone. Dimensions of mapping can include a review of relevant literature germane to the context, familiarity with the actors and activities present in the context, and computational mapping representing how projects within a network or tied together. However, bonding social capital can be accrued with understanding and sharing new findings that may come of this effort with a network (Lacity, *et al.*, 2009; Coffé and Geys, 2007). While a researcher can advance research and a community's goals that they engage with, the transaction cost of mapping a context is high. It is a knowledge-intensive activity requiring the accumulation of domain knowledge, experience, and intensive learning (von Krogh, *et al.*, 2003). These costs hinder many researchers from moving to a new network because context mapping does not immediately result in key research or practice contributions.

Third, a researcher must build language skills to understand and contribute to a new network. Communities often develop specialized terminologies or jargon that signal community or "in-group" membership (Turner, *et al.*, 1979). Language skills signal "in-group" membership, develop over time, and ensure the researcher's familiarity with key concepts and phrases that strengthens engagement in a new network (Agerfalk and Fitzgerald, 2008). Over time, like mapping the new network, developing language skills is a transaction cost for the researcher. However, it leads to greater knowledge of the new network, which, in turn, enables increased access to more significant insights, which is likely impossible for researchers who do not understand the specialized language of a new network.

Fourth, a researcher must contribute tangible value that the new network members recognize. Deeper insights, derived from such contributions to a network (Turner, *et al.*, 1979) enable detailed narrative description, case studies, empirical analysis, and theory development beyond mere tertiary or secondary means of engagement (Coffé and Geys, 2007). Providing tangible value to the network builds a researcher's credibility and utility (Lacity, *et al.*, 2009), providing researchers with access to rich data (Robert, *et al.*, 2008). By contributing to the new network, researchers acquire information that can lead to more significant insights than previously gained. This accumulates access, insight, and previously unattainable knowledge within a network.

Fifth, a researcher can then leverage their utility beyond their immediate network to bridge to yet other adjacent network. Moving beyond a first adjacent network helps to build relationships in a growing number of networks of varying heterogeneity (Coffé and Geys, 2007), where, in our case, understanding different OSS networks provides a more cohesive view of OSS as an overarching concept. But this is derived from long-term contribution and acknowledgment in more than one network where the engaged scholar has contributed, allowing for the construction of larger narratives than just that of a single network or a small collection of networks. The materiality of what generates such connection may be deep knowledge of adjacent networks (Nardi, 1996; Blincoe, *et al.*, 2012; Goggins and Mascaro, 2013), capacity to measure outputs from a common infrastructure (Blincoe, *et al.*, 2012), or individually formed trust bonds between researchers and practitioners working in mutually interested fields (Valetto, *et al.*, 2012; Goggins, *et al.*, 2013b; Goggins and Valetto, 2014).

In some ways, it may seem we are arguing that commitment, context assessment, language skills, contributions, and an ongoing bridging to new networks are five stages through which one might begin to doubt Coleman's (1988) assertion that social capital is not fungible. Our argument, instead, is that treating knowledge of adjacent networks as fungible neglects to recognize the invisible work of building knowledge of new networks. Can knowledge of adjacent networks be accrued and expended in transactional ways? We think the literature suggests the answer to these questions is "no". Although adjacent networks often share a common technical infrastructure, knowledge of adjacent networks may appear more transactional. However, credibility in an adjacent network is inherently necessary for negotiating third-order infrastructural challenges (Star and Ruhleder, 1996). The deep insights sought by long-term engaged scholars and the accrued knowledge of adjacent networks aim to theoretically and practically reveal the complex tensions in large-scale sociotechnical arrangements.

4.4. Case example: Establishing ourselves in the scientific OSS network

4.4.1. Case example: Orientation

Our entry into the scientific OSS network began with the examination of detailed roadmaps of nearly 300 scientific open source projects. Our approach centered on learning the landscape and language through a detailed qualitative content analysis of funding proposal documents submitted by scientific OSS projects. In these documents, open source project leaders identified their challenges, prior successes, and individual goals. Insight from our prior experiences in corporate OSS allowed us to synthesize these concepts into recurring themes in scientific OSS. The documents' open and axial coding identified concerns associated with developing scientific OSS. We constructed several themes relevant to the health of scientific OSS to which the following issues emerged consistently around the theme of contributor constraints in scientific

OSS:

1. Having sufficient contributors available to maintain and improve the software.
2. Ensuring a large enough set of core contributors for the project to onboard new contributors promptly.
3. Overcoming challenges associated with poor documentation impeded new contributor on-boarding.
4. Building a large enough community of contributors to recruit and train new contributors in more corporate environments.

Our findings at this early stage indicated that scientific OSS projects were concerned about a lack of contributors and have difficulty building communities of contributors. Compared with more traditional volunteer and corporately engaged open source projects, there are fewer contributors to the scientific OSS projects we examined. The low contributor counts likely drive concerns about growing demands from software users for new features, complete software testing, and research lab-specific data integration support. Many scientific open source software projects expressed risks like collapsing under their weight, resulting from growing user demands and a shortage of contributors and time. Interestingly, while scarcity of contributors was an identified and known concern for scientific OSS projects, most of the community outreach and recruiting activity proposed by these projects was focused on building user bases rather than contributor bases.

To triangulate our qualitative observations, we empirically validated concerns about contributor levels. Our question was: Are there really fewer contributors on open source scientific software projects than more corporate projects? To do this, we compared the 2,035 publicly available open source repositories from nearly 300 scientific OSS projects that we reviewed, against 6,744 corporate OSS projects sampled from companies that work in open source. We looked at several factors and hypothesized the challenges articulated by scientific OSS project leaders' differences in participation diversity and size of the contributor base for those projects compared with corporate OSS. Specifically, scientific OSS repositories have nine or fewer committers in 60 percent of the 2,035 repositories examined.

This effort enabled us to identify similarities and differences between corporate and scientific OSS networks, helping us establish what we know and don't know and how to best move forward. We knew that our knowledge of corporate OSS could not, by fiat or process, automatically transform into a well-fitted utility for this new scientific context. To move forward, we used the five phases that we applied in our corporate open source efforts as guides to help in our entry into this new network. As existing engaged scholarship methodologies do not speak explicitly about the role of social capital in engaged scholarship, we suggest that these five phases lead one to a conceptual framework of the steps for moving between networks. Additional orientation efforts included regular participation in two scientific software organizations following our initial deep engagement with a collection of scientific projects we described earlier.

First, we participated with "FAIR for Research Software" [6], whose aims are focused on helping scientists, who happen to build scientific OSS, develop a shared path for recognition of those contributions within their institutions and the scientific community. In addition, FAIR for Research Software helps scientists who develop open source software learn about five critical characteristics for their projects to strive for in the interest of sustainability: 1) have a version-controlled repository, 2) include an open source license, 3) register the software in a community registry like CRAN, for R-based statistical software, for example, 4) enable citation of scientific software, using publications like the *Journal of Open Source Software* [7] or using markers like those available through CiteAs [8], and 5) use a software quality checklist [9]. Second, we participated in the Research Software Alliance (ReSA) [10], whose mission aligns with FAIR for Research Software (FAIR4RS) but aims more directly at the general advancement of research software quality in OSS. Our collaboration focused on developing a large, in person workshop in mid–2022 and seeking specific mechanisms for applying, sharing, and informing members about using CHAOSS metrics to advance scientific OSS.

4.4.2. Case example: Mapping open source scientific software

We built our map of the open source scientific software landscape from two perspectives to focus on accruing knowledge of the open source scientific software network. First, we expanded our analysis of the aforementioned set of 2,035 scientific OSS repositories using an integration of six different computational modeling techniques: Clustering, topic modeling, computational linguistic analysis (CLA) of discussion novelty, CLA of sentiment, predictive modeling of pull request acceptance likelihood using our specialized combination of statistical models, and a random forest algorithm to identify anomalies in the statistical rates of participation for each repository in our previously noted corpora. We combined those results into a grouping of 14 collections in the original corpora and are using the results as we continued our engaged scholarship with FAIR4RS and ReSA. Second, we analyzed 92 scientific OSS repositories from the U.S. National Science Foundation-funded projects. In this work, we are collaborating with CiteAs to mutually understand the characteristics of funded projects that resulted in sustained scientific OSS projects and those that did not, using CHAOSS metrics. We also (1) seeded our list with projects we were familiar with through our initial engagements and prior research and (2) generated CHAOSS metrics for larger sets of scientific OSS projects, generating a high-level understanding of the health of those projects and how the scientific OSS network is structured. Our mapping activity emerged as a critical step in developing knowledge of an adjacent network (*i.e.*, scientific OSS) develops through a combination of the bridging social capital derived another network (*i.e.*, corporate OSS).

4.4.3. Case example: Building language skills

From our outreach and mapping efforts, we focused learning the language of the scientific OSS network. We built our language in this new network by attending meetings, conferences, and workshops aimed at scientific OSS. We met with people directly in their communities to learn about their challenges with respect to the development of their scientific software and the management of their communities. In building language skills, we found that much of our ability to bridge between corporate OSS and scientific OSS stemmed from how fundamental open source concepts are used. This included, for example, language around community responsiveness, licensing and copyright, newcomer experiences, and contributor retention. What was unique when bridging between networks is how these concepts are drawn together in different ways to achieve meaningful goals for a respective network. Knowing how to assemble such concepts in meaningful ways for the scientific OSS network turned out to be the primary focus of our language building, developing an awareness of how members attend to open source concepts as applicable to goals of community growth and decline, stewardship and sustainability of communities, software maintenance, and ecosystem management. While we could rely on fundamental open source concerns that we learned in our engagement within the corporate OSS network, we were required to reassemble our use of the concepts in meaningful ways in the scientific OSS network.

4.4.4. Case example: Contributing value

From our engagement with FAIR4RS and ReSA we identified a subset of exemplar projects in each organization to contribute to. Our work here focused on the impact and scope of those projects within our larger corpora of the several thousand scientific OSS projects described earlier. We analyzed connections between the projects by identifying contributors who participate in the exemplar projects and other projects inside and outside our larger corpora, with preliminary findings suggesting the existence of programming language centered “contributor groups” within scientific OSS. The exemplar projects helped explain scientific OSS, which is less organizationally bound than we find in corporate OSS. We hypothesize this is because scientific OSS is more diffuse and varied. These findings proved to be helpful information to project owners, who contrast projects they view as successful with their projects. This previously unavailable awareness helped ground the efforts of scientific OSS project maintainers to improve their software and expand their contributor communities.

Working orthogonally with our analysis and contextualization of scientific OSS within the exemplar projects, we discovered relationships between the “contributor groups” identified and scientific papers resulting from funded projects in the 92 project CiteAs corpora. This work included identifying related papers through systematic literature reviews focused on criteria of scientific impact in concert with scientific OSS usage. We helped a community of scientific OSS builders tease out the networks and technical mechanisms that elevate the impact of their software on science and when, where, and if that impact is visible in publications.

From our conversations with scientists who build open source software, their use of any project can be idiosyncratic to their labs, available equipment, and the team member skills working there in any given year. We are investigated a small number of individual cases where this is apparent, and our work showed these practices to be at once widespread and labor-intensive to understand. Unlike the relationship between scientific OSS and scientific publications, these assemblages existed in isolation. To fully understand how individual scientific OSS software pipelines follow patterns or validate an early hypothesis that they follow some pattern, we described how the type of long-term engaged scholarship experienced through our work in CHAOSS can generate further insight. Through these processes, we recognized that the knowledge of adjacent OSS networks will help build understanding in a broad sense of the health of scientific OSS by:

1. Improving information resources for stakeholders,
2. Identifying critical yet under-resourced open source scientific software and
3. Identifying local collections of open source scientific software as meaningful for scientific scholarship.

Throughout our work, bridging knowledge of corporate OSS to scientific OSS was possible because we contributed resources and knowledge to the communities, projects, and organizations we worked with in the scientific OSS network. Our preliminary findings point strongly to scientific and corporate OSS maintaining sharply distinct characteristics and practices. While work remains, we believe we are contributing knowledge to scientific OSS communities by helping identify areas of concern that may require attention to maintain the health and sustainability of their communities. Further, this information creates value by providing a glimpse of systematic and structured scientific OSS community work.

4.4.5. Case example: Bridging to new networks

From our work in the corporate and scientific OSS networks, we can now bridge to new networks (*i.e.*, university OSS) to even further understand the overarching domain of open source. We are now attending to the emerging network of university open source — focused on supporting university activities tied to open source, including research excellence, research translation, education, and community development. Like the relationship between scientific and corporate OSS networks, university open source shares fundamental open source similarities that we use to bridge to this new network. Having done similar bridging work between corporate and scientific OSS networks, we can bridge to the university network as we can focus more immediately on learning how university open source assembles fundamental open source concepts in ways relevant to their network. To date, our efforts have focused on developing relationships in understanding questions and challenges university open source members face. This has included meeting with university open source members from academic institutions and connecting with existing efforts in this newly emerging network.

5. Discussion: Building and using social capital for long term sociotechnical engagement

This paper explicates our methods for researching open source software across OSS networks and describes the role of social capital in this. Agre (2004) pointed out how centuries of political theory development missed the essential truth that social skills are a fundamental concern in political science (Agre, 2004; Agre and Schuler, 1997). We illustrated the essential function and utility of social capital for examining large-scale sociotechnical infrastructures, which is critical for developing more robust social computing theory

(Hammond, 1988; Giorgi, 1997).

Star and Ruhleder (1996) apply the steps outlined by Bateson (2000) to discuss three levels of issues in understanding large-scale sociotechnical infrastructures. First-order issues are easy to identify and can be addressed through the addition or redistribution of existing resources. Second-order issues are more challenging to locate and can arise from the interaction of two or more first-order issues. In the study of open source software, first- and second-order issues center around the sociotechnical aspects of these systems. When bridging from one OSS network to another, startup knowledge of adjacent networks is derived from a current understanding of common problems within these first two orders of issues. Our knowledge of shared technical infrastructure of open source software is a materiality (Leonardi, *et al.*, 2012) that is one component of our startup knowledge across networks.

Third-order issues are more challenging to understand, are broader in scope, and involve, for example, deep conceptual disagreements or political issues. Complexities arises in how different OSS networks enact third-order issues through differentiation among social groups that comprise a network's social constructs and practices, making this a form of systematic development across networks a vital methodological concern for researchers (Tajfel, 1982a, 1978). The knowledge of an adjacent network requires understanding the role of social infrastructure as an invisible yet vital aspect of large-scale collaborative systems (Star and Ruhleder 1996; Star, 1999). We found that research across OSS networks requires attention this third-order issue in the development of social knowledge both across and within networks (Coleman, 1988; Wenger, 1998).

5.1. Methods of engaged scholarship

Our experience of gaining knowledge within and across OSS networks contributes to existing research methods in several ways. First, context matters when integrating social and technical processes (Nardi, 1996). To elaborate on the functional effects of acquiring knowledge across adjacent OSS networks, we assert a contextual hierarchy to recognize, assess, and adapt to each specific case of building social capital in a new network. Our approach includes attention to processes, outputs, and existing units of organization, both formal and informal (Alderfer and Smith, 1982; Peterson, 1998) where the researcher is frequently engaged as a facilitator or participant in transformative actions within a network (Stringer, 2014).

Second, and to reiterate a key idea from our work, from a distant perspective, open source software might appear to be a singular domain, with nominal contextual differences. However, our approach accommodates social distinctions across OSS networks with implications for the researcher's process to ensure effective engagement, leading to new knowledge and improved circumstances for each engaged network. Within the corporate OSS network, our long-term engagement and research program relied heavily on our deep understanding of the domain, which results in shared success for us as scholars and informants. One could think of that sort of knowledge as "top-down" because, like ethnographers, we fully embed ourselves with support from organizational leadership, and we are building a long-term record of effective organizational change.

As we learned in scientific OSS, bridging our knowledge of the corporate context to science required a more "bottom-up" approach to conducting engaged scholarship. Our established expertise in developing metrics for estimating the health and sustainability of corporate OSS (Germonprez, *et al.*, 2019, 2018) helped initiate our engagement with scientific OSS. As newcomers, our interactions focused on questions about our expertise and how we could help (Collins and Evans, 2007), ensuring to not drive skepticism of our expertise from another network (Collins, 2014). We think, in general, new research projects where a researcher aims to engage over the long term in adjacent networks are likely to begin as "bottom-up" rather than "top-down" relationships. However, the development and evolution of knowledge in adjacent networks is not necessarily linear or deterministic where one can expect progression from expertise acquired from an adjacent network into impact and commitment in a new network.

5.2. Constructions of OSS networks

The contrast between corporate OSS and scientific OSS illustrates the function of research context adjacency and distinction in the evolution of how researchers engage within and across networks. Corporate OSS is a context focused on increasing the systematic organizational structure of open source projects across its ecosystems for the strategic advantage of its members. The structure of the Linux Foundation, for example, is valuable to thousands of technology firms because it sustains essential infrastructure that is non-market-differentiating for its members. Moreover, maintaining this type of structured network of open source projects is actively facilitated by over a dozen annual conferences where all members or members with interests meet and build connections, ultimately serving the open source projects themselves and the Linux Foundation's mission.

In contrast, scientific OSS has two key characteristics that lead us to believe that effectively bridging between networks is not always routine. We observed two unique characteristics in the scientific OSS network. The first characteristic includes observable, technical connections between scientific OSS projects. For example, many Matplotlib projects import another open source project for data manipulation called Pandas. Dependency types, not observable except through work in individual labs, are constructed from a combination of software and process within each lab's "scientific pipeline." In the life sciences, this includes a bricolage of strategies that move from some biological sample through data carpentry, analysis, and aggregation using distinct open source projects in concert with localized processes to perform each step required to produce research findings that advance research aimed at, for example, curing disease. This is unique from corporate OSS in that scientific OSS does not build within a single foundation in the same way as the Linux Foundation supports work for corporations in the development of the non-differentiating technologies.

The second characteristic is the project leadership style in many scientific OSS projects. Leadership is more distributed (McDonald and Goggins, 2013; Gronn, 2009, 2000; Stewart, *et al.*, 2015) than it is "lightly centralized". We have not observed an innate or organic motivation to structure, survey, or organize the structural, technical, process, and social dependencies as potential leverage points in scientific OSS. In contrast, these factors significantly motivate corporate OSS contexts. For example, there is a divergence in the relationship between the social context of use for scientific OSS and corporate OSS. Small teams build open source software needed to do science in many domains, while corporate open source software coordinates, manages, and monitors across a much more extensive social network of developers and organizations.

Adapting knowledge to divergent materialities in scientific OSS and corporate OSS are built from and aimed at means that, methodologically, we must consider types of divergences to effectively adapt our knowledge to any new network (Leonardi, *et al.*, 2012). In more philosophical terms, this form of embedded sociotechnical research compels an explicit recognition of how our ontological choices for framing the method may hide our epistemological commitments (Goggins and Petakovic, 2014; Floridi, 2009, 2008) if our method is not clear about areas of divergence between sociotechnical networks.

A methodology focused on long-term engagement and building knowledge in adjacent networks needs to foreground the importance of recognizing the entanglement of, in our case, the open source software, the social processes, and the research artifacts produced when interpreting the health and sustainability of open source scientific software (Leonardi, *et al.*, 2012; Leonardi and Barley, 2010; Barad, 2007). The entanglement of people, process, and technologies creating scientific and corporate OSS software are different. It is more common in science for a small group of scientists to "build software they need," even if that is not their primary job. In contrast, corporate open source software includes many software developers paid by corporations to build and maintain software their employers need.

How we as researchers choose to make agential cuts (Barad, 2007) to distinguish the social, technical, and procedural objects we are interpreting will, effectively, result in a set of ontological commitments that define their relationship and constrain how we interpret observations across networks (Leonardi, *et al.*, 2012; Leonardi and Barley, 2010; Barad, 2007). For example, one might distinguish these two contexts through the lens of motivation. That agential cut could be, "corporate open source is motivated by profit, while scientific open source is motivated by the production of papers." Methodologically, it is not essential to make these "cuts" at the beginning but developing the necessary knowledge that is necessary to advance scientific OSS in the service of curing disease is already defining a critical agential cut we need to make to build and sustain relationships across networks during a long term engagement (Barad, 2007).

6. Conclusion

Our methodological attention to divergent materialities across networks is part of our method's reflexive nature. The metrics we defined and deployed through our work with CHAOSS regarding the interactions between software objects and developers, between developers and developers, and between developers and projects fit many open source networks well (Goggins, *et al.*, 2013a). This knowledge of social constructs and software practices in new open source networks results from our prior work in adjacent networks.

Some sociotechnical research questions require long-term engagement, and in these cases, a researcher bears some responsibility for facilitating the organizational structure in a new network adjacent to one they are already studying. In these cases, bringing value from an adjacent network and recognizing the new network as having some critical differences, are prerequisites to building a structured research program for long-term engagement. In our case, our long-term understanding how and to what extent the current configurations in scientific OSS influence, advance, or impede scientific progress also demands the development of social capital in this network. Researchers may begin building social capital in a network by making invisible dependencies between networks more visible or categorizing projects along dimensions a community thinks may distinguish a project as essential or promising in accelerating goals of that network.

While researchers have deeply inspected individual research methods, they rarely discuss what is involved in establishing and continuing research agendas in studying large-scale sociotechnical systems. Thus, fundamental knowledge for research is relatively tacit or written off as relatively unimportant practice work. This paper draws on our experiences and the literature on social capital to theorize and illustrate the work involved in long-term and cross-cutting research agendas. We argue that understanding the distinct social and practice realms in an adjacent network is an underappreciated aspect of building research over long periods. Engaged scholarship in a first network leads to accomplishments, making understanding and appreciating adjacent networks possible. Moreover, knowledge of a first network translates — albeit not without work — to utility in an adjacent network, shortening the time required to build relationships. This diligent work is somewhat invisible to evaluating academic productivity and almost absent from doctoral education, yet we feel it is knowledge worth knowing.

Future sociotechnical scholars may use this methodological approach to quickly advance their work into new worlds and contexts. Alternatively, scholars may use it to explain why their approach to diligent work is vital for their research agendas, careers, and sociotechnical systems research. Finally, we hope others will apply their own theoretical lenses to share, study, and clarify the long-term work involved in engaged scholarship.

About the authors

Sean Goggins is a professor of electrical engineering and computer science at the University of Missouri.
E-mail: s [at] goggins [dot] com

Matt Germonprez is a professor of information systems at the University of Nebraska at Omaha.
E-mail: Germonprez [at] gmail [dot] com

Kevin Lumbard is an assistant professor of computer science at Creighton University.
E-mail: kevinlumbard [at] creighton [dot] edu

James Howison is an associate professor in the Information School at the University of Texas at Austin.
E-mail: jhowison [at] ischool [dot] utexas [dot] edu

Karthik Ram is a research scientist at the Berkeley Institute for Data Science.
E-mail: karthik [dot] ram [at] berkeley [dot] edu

Caifan Du is a Ph.D. student in the Information School at the University of Texas at Austin.
E-mail: cfdu [at] utexas [dot] edu

Notes

1. https://scikit-learn.org/stable/.

2. Nardi and O'Day, 2000, pp. 67–68.

3. Nardi and O'Day, 2000, p. 67.

4. Bruckman, *et al.*, 2015, p. 256.

5. Checkland and Poulter, 2006, pp. 3–6.

6. https://www.rd-alliance.org/groups/fair-4-research-software-fair4rs-wg.

7. https://joss.theoj.org/.

8. https://citeas.org.

9. https://fair-software.nl/recommendations/checklist.

10. https://www.researchsoft.org/.

References

P.J. Agerfalk and B. Fitzgerald, 2008. "Outsourcing to an unknown workforce: Exploring opensourcing as a global sourcing strategy," *MIS Quarterly*, volume 32, number 2, pp. 385–409, and at https://misq.umn.edu/ourtsourcing-in-an-unknown-workforce-exploring-opensourcing-as-a-global-sourcing-strategy.html, accessed 13 January 2025.

P. Agre, 2004. "The practical republic: Social skills and the progress of citizenship," In: A. Feenberg and D. Barney (editors). *Community in the digital age: Philosophy and practice*. Lanham, Md.: Rowman & Littlefield, pp. 201–223.

P. Agre and D. Schuler (editors), 1997. *Reinventing technology, rediscovering community: Critical explorations of computing as a social practice*. Greenwich, Conn.: Ablex.

C.P. Alderfer and K.K. Smith, 1982. "Studying intergroup relations embedded in organizations," *Administrative Science Quarterly*, volume 27, number 1, pp. 35–65.
doi: https://doi.org/10.2307/2392545, accessed 13 January 2025.

K. Asdal, B. Brenna, and I. Moser (editors), 2007. Technoscience: The politics of interventions. Oslo: Unipub.

K. Barad, 2007. *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Durham, N.C.: Duke University Press.
doi: https://doi.org/10.2307/j.ctv12101zq, accessed 13 January 2025.

G. Bateson, 2000. *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology*. Chicago: University of Chicago Press.

K. Blincoe, G. Valetto, and S. Goggins, 2012. "Proximity: A measure to quantify the need for developers' coordination," *CSCW '12: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, pp 1,351–1,360.
doi: https://doi.org/10.1145/2145204.2145406, accessed 13 January 2025.

B. Bozeman and E. Corley, 2004. "Scientists' collaboration strategies: Implications for scientific and technical human capital," *Research Policy*, volume 33, number 4, pp. 599–616.
doi: https://doi.org/10.1016/j.respol.2004.01.008, accessed 13 January 2025.

A. Bruckman, K. Luther, and C. Fiesler, 2015. "When should we use real names in published accounts of Internet research?" In: E. Hargittai and C. Sandvig (editors). *Digital research confidential: The secrets of studying behavior online*. Cambridge, Mass.: MIT Press, pp. 243–258.
doi: https://doi.org/10.7551/mitpress/9386.001.0001, accessed 13 January 2025.

P.J. Carrington, J. Scott, and S. Wasserman (editors), 2005. *Models and methods in social network analysis*. Cambridge: Cambridge University Press.
doi: https://doi.org/10.1017/CBO9780511811395, accessed 13 January 2025.

P. Checkland and J. Poulter, 2010. "Soft systems methodology," In: M. Reynolds and S. Holwell (editors). *Systems approaches to managing change: A practical guide*. London: Springer, pp. 191–242.
doi: https://doi.org/10.1007/978-1-84882-809-4_5, accessed 13 January 2025.

P. Checkland and J. Poulter, 2006. *Learning for action: A short definitive account of soft systems methodology and its use for practitioners, teachers, and students*. Chichester: Wiley.

C.E.H. Chua, W.-K. Lim, C. Soh, and S.K. Sia, 2012. "Enacting clan control in complex IT projects: A social capital perspective," *MIS Quarterly*, volume 36, number 2, pp. 577–600, and at https://misq.umn.edu/enacting-clan-control-in-complex-it-projects-a-social-capital-perspective.html, accessed 13 January 2025.

H. Chughtai and M.D.Myers, 2017. "Entering the field in qualitative field research: A rite of passage into a complex practice world: Entering the field in qualitative field research," *Information Systems Journal*, volume 27, number 6, pp. 795–817.
doi: https://doi.org/10.1111/isj.12124, accessed 13 January 2025.

A.E. Clarke, 2003. "Situational analysis: Grounded theory mapping after the postmodern turn," *Symbolic Interaction*, volume 26, number 4, pp. 553–576.
doi: https://doi.org/10.1525/si.2003.26.4.553, accessed 13 January 2025.

H. Coffé and B. Geys, 2007. "Toward an empirical characterization of bridging and bonding social capital," *Nonprofit and Voluntary Sector Quarterly*, volume 36, number 1, pp. 121–139.
doi: https://doi.org/10.1177/0899764006293181, accessed 13 January 2025.

E.G. Coleman, 2013. *Coding freedom: The ethics and aesthetics of hacking*. Princeton, N.J.: Princeton University Press.

doi: https://doi.org/10.2307/j.ctt1r2gbj, accessed 13 January 2025.

J.S. Coleman, 1988. "Social capital in the creation of human capital," *American Journal of Sociology*, volume 94, pp. 95–120.
doi: https://doi.org/10.1086/228943, accessed 13 January 2025.

H. Collins, 2014. "Rejecting knowledge claims inside and outside science," *Social Studies of Science*, volume 44, number 5, pp. 722–735.
doi: https://doi.org/10.1177/030631271453601, accessed 13 January 2025.

H. Collins and R. Evans, 2007. *Rethinking expertise*. Chicago: University of Chicago Press.

R. Dodsworth and R.A. Benton, 2019. *Language variation and change in social networks: A bipartite approach: A bipartite approach*. New York: Routledge.
doi: https://doi.org/10.4324/9781315642000, accessed 13 January 2025.

C. Dunbar-Hester, 2020. *Hacking diversity: The politics of inclusion in open technology cultures*. Princeton, N.J.: Princeton University Press.
doi: https://doi.org/10.2307/j.ctvhrd181, accessed 13 January 2025.

S.W. Duxbury, 2022. *Longitudinal network models*. Los Angeles, Calif.: Sage.
doi: https://doi.org/10.4135/9781071938904, accessed 13 January 2025.

K. Etherington, 2007. "Ethical research in reflexive relationships," *Qualitative Inquiry*, volume 13, number 5, pp. 599–616.
doi: https://doi.org/10.1177/1077800407301175, accessed 13 January 2025.

L. Floridi, 2009. "Web 2.0 vs. the semantic Web: A philosophical assessment," *Episteme*, volume 6, number 1, pp. 25–37.
doi: https://doi.org/10.3366/E174236000800052X, accessed 13 January 2025.

L. Floridi, 2008. "The method of levels of abstraction," *Minds and Machines*, volume 18, number 3, pp. 303–329.
doi: https://doi.org/10.1007/s11023-008-9113-7, accessed 13 January 2025.

R.S. Geiger and D. Ribes, 2011. "Trace ethnography: Following coordination through documentary practices," *2011 44th Hawaii International Conference on System Sciences*, pp. 1–10.
doi: https://doi.org/10.1109/HICSS.2011.455, accessed 13 January 2025.

M. Germonprez, J. Lipps, and S. Goggins, 2019. "The rising tide: Open source's steady transformation," *First Monday*, volume 24, number 8.
doi: https://doi.org/10.5210/fm.v24i8.9297, accessed 13 January 2025.

M. Germonprez, G.J. Link, K. Lumbard, and S. Goggins, 2018. "Eight observations and 24 research questions about open source projects: Illuminating new realities," *Proceedings of the ACM on Human-Computer Interaction*, volume 2, number CSCW article number 57, pp. 1–22.
doi: https://doi.org/10.1145/3274326, accessed 13 January 2025.

M. Germonprez, T. Gurney, S.U.S. Korlimarla, and R. Gandhi, 2016. "DoSOCS: A system for SPDX 2.0 document creation and storage," *Journal of Open Source Software*, volume 1, number 7, 38.
doi: https://doi.org/10.21105/joss.00038, accessed 13 January 2025.

M. Germonprez, B. Young, L. Mathiassen, J.E. Kendall, and K.E. Kendall, 2012. "Risk mitigation in corporate participation with open source communities: Protection and compliance in an open source supply chain," *eProceedings of the 7th International Research Workshop on Information Technology Project*

Management (IRWITPM), pp. 8–100, and at http://aisel.aisnet.org/irwitpm2012/3, accessed 13 January 2025.

M. Germonprez, J. Kendall, K. Kendall, B. Warner, and L. Mathiassen, 2011. "Organizational participation in open communities: Conceptual framing and early findings," *Proceedings of the Seventeenth Americas Conference on Information Systems*, at http://aisel.aisnet.org/amcis2011_submissions/242, accessed 13 January 2025.

A. Giorgi, 1997. "The theory, practice, and evaluation of the phenomenological method as a qualitative research procedure," *Journal of Phenomenological Psychology*, volume 28, number 2, pp. 235–260.
doi: https://doi.org/10.1163/156916297X00103, accessed 13 January 2025.

S. Goggins and E. Petakovic, 2014. "Connecting theory to social technology platforms: A framework for measuring influence in context," *American Behavioral Scientist*, volume 58, number 10, pp. 1,376–1,392.
doi: https://doi.org/10.1177/0002764214527093, accessed 13 January 2025.

S.P. Goggins and G. Valetto, 2014. "Assessing the structural fluidity of virtual organizations and its effects," In: K. Zweig, W. Neuser, V. Pipek, M. Rohde, and I. Scholtes (editors). *Socioinformatics — The social impact of interactions between humans and IT*. Cham, Switzerland: Springer International, pp. 121–137.
doi: https://doi.org/10.1007/978-3-319-09378-9_8, accessed 13 January 2025.

S. Goggins and C. Mascaro, 2013. "Context matters: The experience of physical, informational, and cultural distance in a rural IT firm," *Information Society*, volume 29, number 2, pp. 113–127.
doi: https://doi.org/10.1080/01972243.2012.758212, accessed 13 January 2025.

S. Goggins, K. Lumbard, and M. Germonprez, 2021. "Augur: Architecture for capturing, reshaping, and socially contextualizing open source software communities," *ACM SoHeal Conference*.

S.P. Goggins, C. Mascaro, and G. Valetto, 2013a. "Group informatics: A methodological approach and ontology for sociotechnical group research," *Journal of the American Society for Information Science and Technology*, volume 64, number 3, pp. 516–539.
doi: https://doi.org/10.1002/asi.22802, accessed 13 January 2025.

S.P. Goggins, G. Valetto, C. Mascaro, C., and K. Blincoe, 2013b. "Creating a model of the dynamics of socio-technical groups using electronic trace data," *User Modeling and User-Adapted Interaction*, volume 23, number 4, pp. 345–379.
doi: https://doi.org/10.1007/s11257-012-9122-3, accessed 13 January 2025.

I. Graves, N. McDonald, and S. Goggins, 2016. "Sifting signal from noise: A new perspective on the meaning of tweets about the 'big game'," *New Media & Society*, volume 18, number 2, pp. 293–312.
doi: https://doi.org/10.1177/1461444814541783, accessed 13 January 2025.

P. Gronn, 2009. "Leadership configurations," *Leadership*, volume 5, number 3, pp. 381–3394.
doi: https://doi.org/10.1177/1742715009337770, accessed 13 January 2025.

P. Gronn, 2000. "Distributed properties: A new architecture for leadership," *Educational Management & Administration*, volume 28, number 3, pp. 317–333.
doi: https://doi.org/10.1177/0263211X000283006, accessed 13 January 2025.

P.E. Hammond, 1988. "The relationship between method and theory," *Soundings*, volume 71, numbers 2–3, pp. 365–371.

D. Haraway, 1988. "Situated knowledges: The science question in feminism and the privilege of partial perspective," *Feminist Studies*, volume 14, number 3, 575–599.

doi: https://doi.org/10.2307/3178066, accessed 13 January 2025.

S. Harrison and D. Tatar, 2008. “Places: people, events, loci — The relation of semantic frames in the construction of place,” *Computer Supported Cooperative Work*, volume 17, numbers 2–3, pp. 97–133.
doi: https://doi.org/10.1007/s10606-007-9073-0, accessed 13 January 2025.

J. Howison and J. Bullard, 2016. “Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature,” *Journal of the Association for Information Science and Technology*, volume 67, number 9, pp. 2,137–2,155.
doi: https://doi.org/10.1002/asi.23538, accessed 13 January 2025.

A.L. Hughes and L. Palen, 2009. “Twitter adoption and use in mass convergence and emergency events,” *International Journal of Emergency Management*, volume 6, numbers 3–4, pp. 248–260.
doi: https://doi.org/10.1504/IJEM.2009.031564, accessed 13 January 2025.

J. Introne and S. Goggins, 2019. “Advice reification, learning, and emergent vollective intelligence in online health support communities,” *Computers in Human Behavior*, volume 99, pp. 205–218.
doi: https://doi.org/10.1016/j.chb.2019.05.028, accessed 13 January 2025.

J. Introne and S. Goggins, 2015. “Taming a menagerie of heavy tails with skew path analysis,” *WebSci ’15: Proceedings of the ACM Web Science Conference*, article number 19, pp. 1–5.
doi: https://doi.org/10.1145/2786451.2786484, accessed 13 January 2025.

J. Introne and S. Goggins, 2012. “Tracing knowledge evolution in online forums,” *Proceedings from ACM Web Science Conference: Words and Networks Workshop* (Evanston, Ill.).

J. Introne, B. Semaan, and S. Goggins, 2016. “A sociotechnical mechanism for online support provision,” *CHI ’16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 3,559–3,571.
doi: https://doi.org/10.1145/2858036.2858582, accessed 13 January 2025.

J. Introne, I. Erickson, B. Semaan, and S. Goggins, 2020. “Designing sustainable online support: Examining the effects of design change in 49 online health support communities,” *Journal of the American Society for Information Science and Technology*, volume 71, number 4, pp. 379–394.
doi: https://doi.org/10.1002/asi.24250, accessed 13 January 2025.

J. Introne, K. Levy, S. Munson, S. Goggins, R. Wash, and C. Aragon, 2012. “Design, influence, and social technologies: Techniques, impacts, and ethics,” *CSCW ’12: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion*, pp. 9–10.
doi: https://doi.org/10.1145/2141512.2141518, accessed 13 January 2025.

C. Jensen, S. King, and V. Kuechler, 2011. “Joining free/open source software communities: An analysis of newbies’ first interactions on project mailing lists,” *HICSS ’11: Proceedings of the 2011 44th Hawaii International Conference on System Sciences*, pp. 1–10.
doi: https://doi.org/10.1109/HICSS.2011.264, accessed 13 January 2025.

C.M. Kelty, 2008. *Two bits: The cultural significance of free software*. Durham, N.C.: Duke University Press.
doi: https://doi.org/10.1215/9780822389002, accessed 13 January 2025.

J.E. Kendall, K.E. Kendall, and M. Germonprez, 2016. “Game theory and open source contribution: Rationale behind corporate participation in open source software development,” *Journal of Organizational Computing and Electronic Commerce*, volume 26, number 4, pp. 323–343.
doi: https://doi.org/10.1080/10919392.2016.1228360, accessed 13 January 2025.

M.C. Lacity, S.A. Khan, and L.P. Willcocks, 2009. “A review of the IT outsourcing literature: Insights for practice,” *Journal of Strategic Information Systems*, volume 18, number 3, pp. 130–146.
doi: https://doi.org/10.1016/j.jsis.2009.06.002, accessed 13 January 2025.

C.P. Lee, 2007. “Boundary negotiating artifacts: Unbinding routine of boundary objects and embracing chaos in collaborative work,” *Computer Supported Cooperative Work*, volume 16, pp. 307–339.
doi: https://doi.org/10.1007/s10606-007-9044-5, accessed 13 January 2025.

P.M. Leonardi and S.R. Barley, 2010. “What’s under construction here? Social action, materiality, and power in constructivist studies of technology and organizing,” *Academy of Management Annals*, volume 4, number 1, pp. 1–51.
doi: https://doi.org/10.1080/19416521003654160, accessed 13 January 2025.

P.M. Leonardi, B.A. Nardi, and J. Kallinikos, 2012. *Materiality and organizing: Social interaction in a technological world*. Oxford: Oxford University Press.
doi: https://doi.org/10.1093/acprof:oso/9780199664054.001.0001, accessed 13 January 2025.

N. McDonald and S. Goggins, 2013. “Performance and participation in open source software on github,” *CHI EA ’13: CHI ’13 Extended Abstracts on Human Factors in Computing Systems*, pp. 139–144.
doi: https://doi.org/10.1145/2468356.2468382, accessed 13 January 2025.

N. McDonald, K. Blincoe, E. Petakovic, and S. Goggins, 2014. “Modeling distributed collaboration on GitHub,” *Advances in Complex Systems*, volume 17, numbers 7–8, 1450024.
doi: https://doi.org/10.1142/S0219525914500246, accessed 13 January 2025.

A.B.M. Murniyati, S. Wirawan, T. Yusnitasari, and D. Anggraini, 2023. “Expanding Louvain algorithm for clustering relationship formation,” *International Journal of Advanced Computer Science and Applications*, volume 14, number 1, pp. 701–708, and at https://thesai.org/Downloads/Volume14No1/Paper_77-Expanding_Louvain_Algorithm_for_Clustering_Relationship.pdf, accessed 13 January 2025.

B.A. Nardi (editor), 1996. *Context and consciousness: Activity theory and human-computer interaction*. Cambridge, Mass.: MIT Press.
doi: https://doi.org/10.7551/mitpress/2137.001.0001, accessed 13 January 2025.

B.A. Nardi and V. O’Day, 2000. *Information ecologies: Using technology with heart*. Cambridge, Mass.: MIT Press.
doi: https://doi.org/10.7551/mitpress/3767.001.0001, accessed 13 January 2025.

L. Palen and S.B. Liu, 2007. “Citizen communications in crisis: Anticipating a future of ICT-supported public participation,” *CHI ’07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 727–736.
doi: https://doi.org/10.1145/1240624.1240736, accessed 13 January 2025.

L. Palen, S. Vieweg, S. Liu, and A. Hughes, 2009. “Crisis in a networked world: Features of computer-mediated communication in the April 16, 2007, Virginia Tech event,” *Social Science Computer Review*, volume 27, number 4, pp. 467–480.
doi: https://doi.org/10.1177/0894439309332302, accessed 13 January 2025.

L. Palen, K.M. Anderson, G. Mark, J. Martin, D. Sicker, M. Palmer, and D. Grunwald, 2010. “A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters,” *ACM-BCS ’10: Proceedings of the 2010 ACM-BCS Visions of Computer Science Conference*, article number 8, pp. 1–12.

R.V. Patulny and G.L.H. Svendsen, 2007. “Exploring the social capital grid: Bonding, bridging, qualitative, quantitative,” *International Journal of Sociology and Social Policy*, volume 27, numbers 1–2, pp. 32–51.

doi: https://doi.org/10.1108/01443330710722742, accessed 13 January 2025.

M.F. Peterson, 1998. "Embedded organizational events: The units of process in organization science," *Organization Science*, volume 9, number 1, pp. 16–33.
doi: https://doi.org/10.1287/orsc.9.1.16, accessed 13 January 2025.

Y. Ren, R. Kraut, and S. Kiesler, 2007. "Applying common identity and bond theory to design of online communities," *Organization Studies*, volume 28, number 3, pp. 377–408.
doi: https://doi.org/10.1177/0170840607076007, accessed 13 January 2025.

M.A. Rich, 2013. "'From coal to cool': The creative class, social capital, and the revitalization of Scranton," *Journal of Urban Affairs*, volume 35, number 3, pp. 365–384.
doi: https://doi.org/10.1111/j.1467-9906.2012.00639.x, accessed 13 January 2025.

L.P. Robert, A.R. Denis, and Y.-T.C. Hung, 2009. "Individual swift trust and knowledge-based trust in face-to-face and virtual team members," *Journal of Management Information Systems*, volume 26, number 2, pp. 241–279.
doi: https://doi.org/10.2753/MIS0742-1222260210, accessed 13 January 2025.

L.P. Robert, A.R. Dennis, and M.K. Ahuja, 2008. "Social capital and knowledge integration in digitally enabled teams," *Information Systems Research*, volume 19, number3, pp. 314–334.
doi: https://doi.org/10.1287/isre.1080.0177, accessed 13 January 2025.

A. Sarcevic, L. Palen, J. White, K. Starbird, M. Bagdouri, and K. Anderson, 2012. "'Beacons of hope' in decentralized coordination: Learning from on-the-ground medical Twitterers during the 2010 Haiti earthquake," *CSCW '12: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pp. 47–56.
doi: https://doi.org/10.1145/2145204.2145217, accessed 13 January 2025.

S.L. Star, 1999. "The ethnography of infrastructure," *American Behavioral Scientist*, volume 43, number 3, pp. 377–391.
doi: https://doi.org/10.1177/00027649921955326, accessed 13 January 2025.

S.L. Star, 1990. "Power, technology and the phenomenology of conventions: On being allergic to onions," *Sociological Review*, volume 38, number S1, pp. 26–56.
doi: https://doi.org/10.1111/j.1467-954X.1990.tb03347.x, accessed 13 January 2025.

S.L. Star and K. Ruhleder, 1996. "Steps toward an ecology of infrastructure: Design and access for large information spaces," *Information Systems Research*, volume 7, number 1, pp. 111–134.
doi: https://doi.org/10.1287/isre.7.1.111, accessed 13 January 2025.

S.L. Star and J.R. Griesemer, 1989. "Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39," *Social Studies of Science*, volume 19, pp. 387–420.
doi: https://doi.org/10.1177/030631289019003001, accessed 13 January 2025.

K. Starbird and L. Palen, 2012. "(How) will the revolution be retweeted? Information diffusion and the 2011 Egyptian uprising," *CSCW '12: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, pp. 7–16.
doi: https://doi.org/10.1145/2145204.2145212, accessed 13 January 2025.

K. Starbird and L. Palen, 2011. "'Voluntweeters': Self-organizing by digital volunteers in times of crisis," *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1,071–1,080.
doi: https://doi.org/10.1145/1978942.1979102, accessed 13 January 2025.

K. Starbird, L. Palen, A. Hughes, and S. Vieweg, 2010. "Chatter on the red: What hazards threat reveals about the social life of microblogged information," *CSCW '10: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, pp. 241–250.
doi: https://doi.org/10.1145/1718918.1718965, accessed 13 January 2025.

C.A. Stewart, W.K. Barnett, E.A. Wernert, J.A. Wernert, V. Welch, and R. Knepper, 2015. "Sustained software for cyberinfrastructure: Analyses of successful efforts with a focus on NSF-funded software," *SCREAM '15: Proceedings of the 1st Workshop on the Science of Cyberinfrastructure: Research, Experience, Applications and Models*, pp. 63–72.
doi: https://doi.org/10.1145/2753524.2753533, accessed 13 January 2025.

E.T. Stringer, 2014. *Action research*. Fourth edition. Thousand Oaks, Calif.: Sage.

J. Sutton, L. Palen, and I. Shklovski, 2008. "Backchannels on the front lines: Emergent uses of social media in the 2007 southern California wildfires," *Proceedings of the Fifth International ISCRAM Conference*, pp. 624–631, and at https://idl.iscram.org/files/sutton/2008/985_Sutton_etal2008.pdf, accessed 13 January 2025.

H. Tajfel, 1982a. *Social identity and intergroup relations*. Cambridge: Cambridge University Press.

H. Tajfel, 1982b. "Social psychology of intergroup relations," *Annual Review of Psychology*, volume 33, pp. 1–39.
doi: https://doi.org/10.1146/annurev.ps.33.020182.000245, accessed 13 January 2025.

H. Tajfel, 1978. *Differentiation between social groups: Studies in the social psychology of intergroup relations*. New York: Academic Press.

A.S. Taylor, 2011. "Out there," *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp 685–694.
doi: https://doi.org/10.1145/1978942.1979042, accessed 13 January 2025.

V.A. Traag, L. Waltman, and N.J. Van Eck, 2019. "From Louvain to Leiden: Guaranteeing well-connected communities," *Scientific Reports*, volume 9, number 1, article number 5233.
doi: https://doi.org/10.1038/s41598-019-41695-z, accessed 13 January 2025.

J.C. Turner, R. Brown, and H. Tajfel, 1979. "Social comparison and group interest in ingroup favouritism," *European Journal of Social Psychology*, volume 9, pp. 187–204.
doi: https://doi.org/10.1002/ejsp.2420090207, accessed 13 January 2025.

G. Valetto, K. Blincoe, and S.P. Goggins, 2012. "Actionable identification of emergent teams in software development virtual organizations," *RSSE '12: Proceedings of the Third International Workshop on Recommendation Systems for Software Engineering*, pp. 11–15.
doi: https://doi.org/10.1109/RSSE.2012.6233401, accessed 13 January 2025.

A.H. Van de Ven, 2007. *Engaged scholarship: A guide for organizational and social research*. New York: Oxford University Press.
doi: https://doi.org/10.1093/oso/9780199226290.001.0001, accessed 13 January 2025.

S. Vieweg, A.L. Hughes, K. Starbird, and L. Palen, 2010. "Microblogging during two natural hazards events: What Twitter may contribute to situational awareness," *CHI '10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1,079–1,088.
doi: https://doi.org/10.1145/1753326.1753486, accessed 13 January 2025.

G. von Krogh, S. Spaeth, and K.R. Lakhani, 2003). "Community, joining, and specialization in open source software innovation: A case study," *Research Policy*, volume 32, number 7, pp. 1,217–1,241.

doi: https://doi.org/10.1016/S0048-7333(03)00050-7, accessed 13 January 2025.

E. Wenger, 1998. *Communities of practice: Learning, meaning and identity*. New York: Cambridge University Press.
doi: https://doi.org/10.1017/CBO9780511803932, accessed 13 January 2025.

Appendix

Through our collaborations with myriad scientific open source software groups, we identified and mapped comparisons between the corporate and scientific open source software structure. Figures 3, 4, and 5 in this appendix provide larger form visualizations of the summary level views in Figure 1. Notable characteristics of these three sociograms include:

1. Network nodes are either contributors or repositories.
2. Repositories are purple.
3. Contributors are along the other color spectrum.
4. In Degree centrality is reflected in node size. Since only the repositories have In Degree centrality above 1 in our data organization, they are the only nodes that appear larger.
5. The distance between nodes in each graph is determined by the similarity of contributor patterns toward each of the repositories from the three ecosystems.
6. The shapes of each sociogram reflect the structure of each expert-identified ecosystem.
7. We can see that in a large corporate ecosystem, the ecosystem's overall social structure is balanced across many projects.
8. In the two scientific ecosystems, there are defined centers, with funded scientific software showing several centers, and the organically identified ecosystem showing one center in a kind of funnel shape.
9. These differences illustrate different practices and interconnections that define these ecosystems, succinctly illustrating that the social and contribution structures of scientific and corporate open source ecosystems are empirically different.

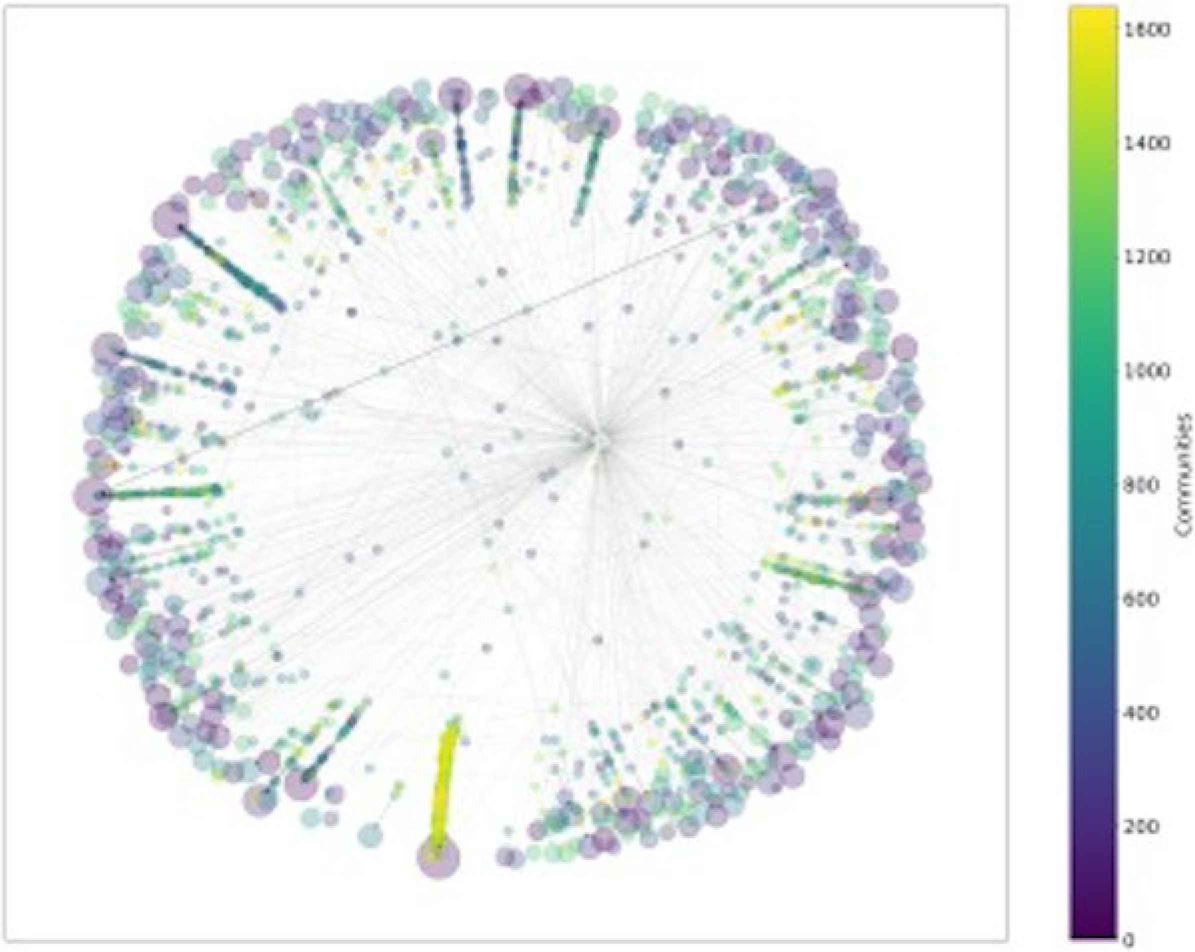


Figure 3: Dozens of these communities are visible in the donut-shaped corporate OSS ecosystem.

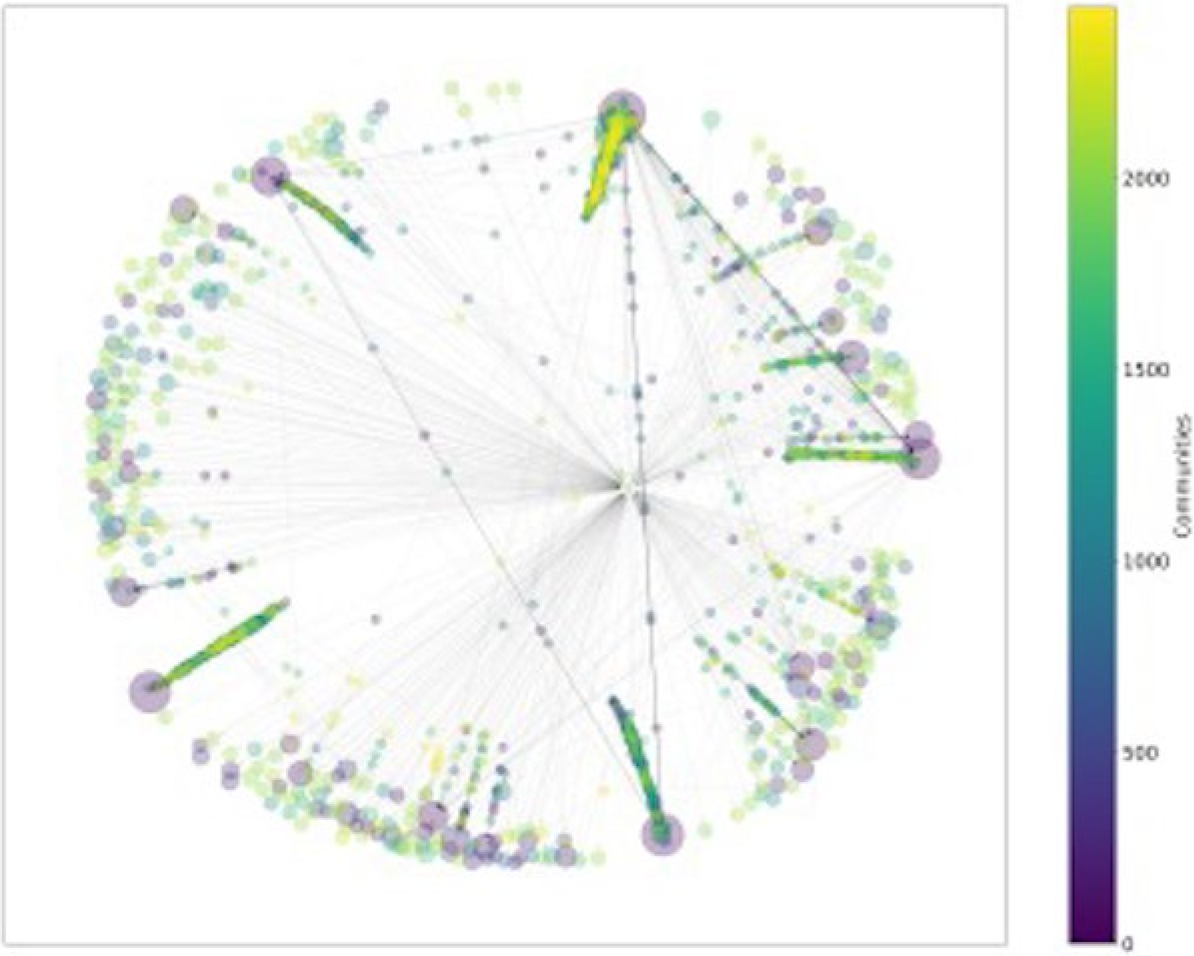


Figure 4: The clock-shaped funded scientific OSS ecosystem has 5-6 major community clusters.

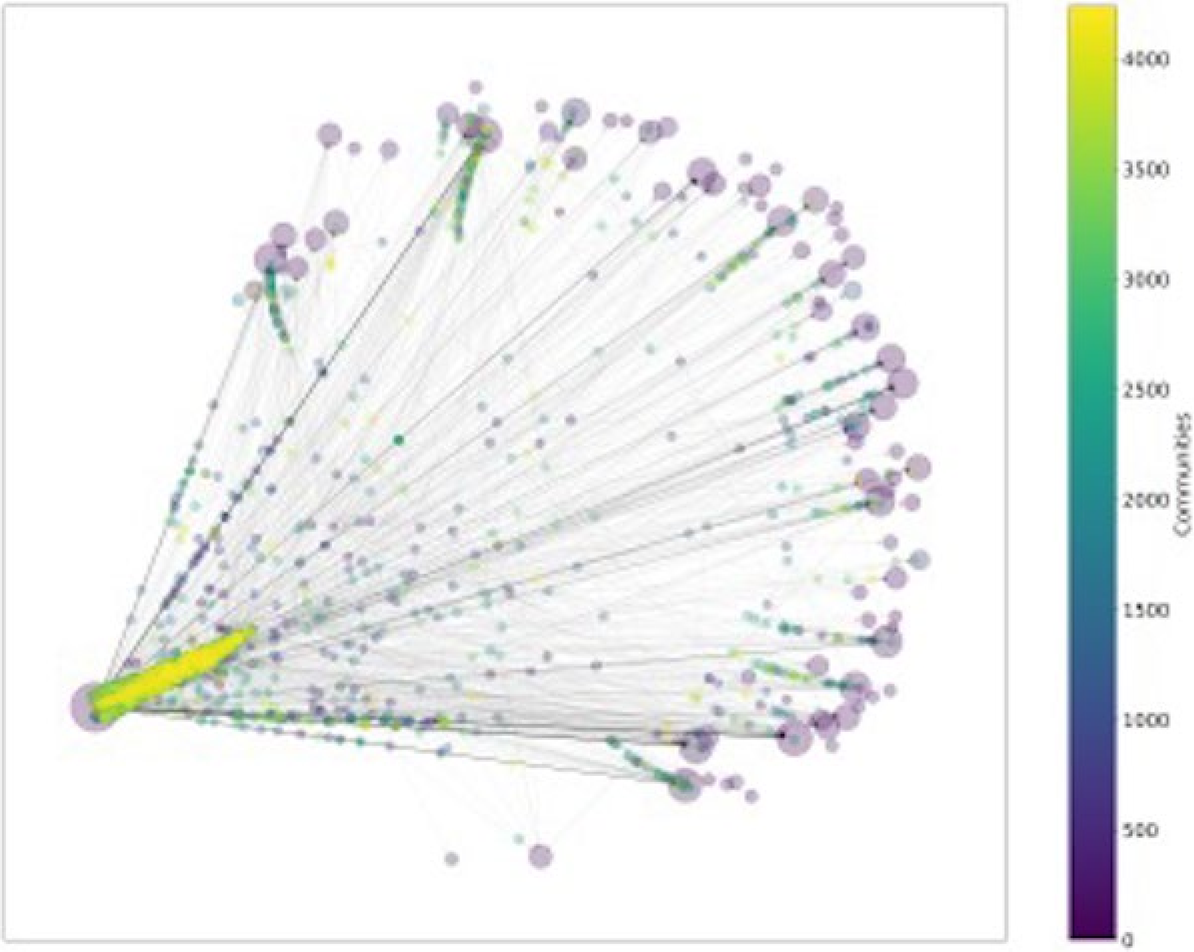


Figure 5: The funnel-shaped organic Scientific OSS ecosystem is visibly dominated by a singular community around a single, dominant project (in this case, Bioconductor).

Editorial history

Received 13 May 2024; revised 31 October 2024; accepted 13 January 2025.

Open source software field research: Spanning social and practice networks for re-entering the field
by Sean Goggins, Matt Germonprez, Kevin Lumbard, James Howison, Karthik Ram, and Caifan Du.
First Monday, volume 30, number 2 (February 2025).
doi: https://dx.doi.org/10.5210/fm.v30i2.13629