BREAKING DOWN REMOTE SILOS: HOW WORKER CHARACTERISTICS AFFECT REMOTE INTERACTIONS IN OPEN-SOURCE SOFTWARE (OSS) COMMUNITIES

PIYUSH GULATI
INSEAD
1 Ayer Rajah Avenue
Singapore 138676

INTRODUCTION

Remote work creates communication silos, impeding information exchange between organizational members. This finding has consistently appeared within the stream of literature developing around the implications of remote work (Bloom, Liang, Roberts, & Ying, 2015; Chauvin, Choudhury, & Fang, 2020; Choudhury, Foroughi, & Larson, 2021; Yang et al., 2021). Prior research finds evidence for a reduction in interaction among a firm's employees that are *structurally* or *geographically* distant from each other, caused by a pandemic-induced remote shift (Yang et al., 2021) and timezone-distance induced reduction in time overlap (Chauvin et al., 2020) respectively. Theoretically, these silos may not have negative performance implications if interdependencies could be perfectly organized within them. But achieving such a perfect organizational design is rarely feasible (Clement & Puranam, 2018; Ethiraj & Levinthal, 2004) – which makes silo-intensification an important problem for organizations to address.

This paper builds upon the current remote work literature by starting to explore what firms can do to overcome remote work challenges. The specific research question that this paper addresses is: what worker-level characteristics are associated with higher cross-silo distant (timezone) interactions? For this, we explore open-source software (OSS) communities as standards of all-remote organizations (He, Puranam, Shrestha, & Krogh, 2020; Lakhani & Von Hippel, 2004; Lerner & Tirole, 2002; von Hippel & von Krogh, 2003). Contributors to open-source are (1) geographically dispersed across timezones, (2) rely primarily on digital communication modes to interact with each other, and (3) are involved in complex, interdependent software development work.

We leverage prior OSS academic literature to gain an understanding of the motivations of two types of contributors, *volunteers* and *firm-sponsored contributors*, as well as speculate about how these motivations link to remote behavior. Moreover, to model their distant interaction behavior, we assume OSS contributors to be goal-driven boundedly rational agents – balancing between exploration and exploitation (March, 1991). This means that for code review, testing, and integration, contributors decide between reaching out to *distant* community members (exploration) or those *nearby* (exploitation).

We use the Linux-kernel OSS project as the context to test the hypotheses, developed from the OSS literature and the problemistic search decision model described above. Communications among contributors, within the linux-kernel project, are accessible through public mailing lists. From these emails, we construct a dataset of 2,879,298 observations of the count of emails sent by contributors to recipients within different timezones in a given month of the year. This panel dataset spans 3,734 senders, 25 timezones, 34 Linux-kernel sub-systems, and 130 months of software development work from Jan 2010 to Oct 2020.

Four main findings help summarize our results. First, we find that timezone distance is associated with a decrease in the frequency of interaction among community members, a finding that corroborates with extant remote work theory. Second, this decrease in the frequency of interactions with distance is lower for volunteers as compared to firm-sponsored contributors, preliminarily indicating that differences in motivations are associated with differences in distant interaction behavior. Third, the reduction in the negative effect is higher for those volunteers that have a lower level of attainment on their recognition and knowledge goals – strengthening support for the goals-driven argument and the problemistic search mechanism we propose. Lastly, further lending support, we find that the decrease in the frequency of interactions with distance is higher for firm-sponsored contributors subjected to a higher firm-level influence.

Primarily, the results of this paper contribute to the remote work literature by advancing the discussion from silos-intensification implications to possible restorative actions.

THEORY AND HYPOTHESES

Open-source software (OSS) communities as remote organizations

One key finding of the remote work literature (Bloom et al., 2015; Choudhury et al., 2021; Yang et al., 2021) is that going remote negatively affects interactions within an organization due to the intensification of communication silos. Under remote work, two underlying mechanisms operate, apart from the elimination of face-to-face interaction. First, the shift to remote makes workers focus on close, strong ties at the expense of those linked indirectly (Yang et al., 2021). Second, when workers are geographically dispersed, they also experience a reduction in time overlap with each other, which further causes a decline in communication (Chauvin et al., 2020).

In this paper, we explore solutions to remote work challenges by looking at OSS communities as standards of completely remote organizations (Lakhani & Wolf, 2003; Lerner & Tirole, 2002; von Hippel & von Krogh, 2003). Within software development, since their inception in the early 1990s, they have grown into an alternative innovation model to close-sourced innovation (Baldwin & von Hippel, 2011). While referred to as *communities*, they satisfy the properties of organizations – they are multi-agent goal-driven systems with well-defined norms for work division and aggregation (Puranam, 2018).

While working from home (or anywhere) is a recent phenomenon in traditional work settings, OSS community members have always worked remotely. Typically, contributors are located all over the world and rely exclusively on digital communication modes to interact among themselves (Faraj et al., 2011; He et al., 2020). Therefore, we posit that challenges of distant non-local interactions, i.e. absence of face-to-face frequent interaction and lack of time overlap, are expected to apply to OSS communities as well. As a baseline, we predict:

> *Hypothesis 1: In an open-source community, contributors have a lower frequency of interactions with recipients in distant timezones versus those in nearby timezones.*

OSS contributors and distant interactions: Exploration vs exploitation trade-off

To understand individual-level differences in the level of distant interactions undertaken, we delve into the software development process within OSS communities. Here, contributors

write new code in independent modules called patches (von Hippel & von Krogh, 2003). They share these patches with other members for review and testing. The decision of who to send the patch, and whether to reply/ignore, lies with the community member or the firm to which they are affiliated.

Based on the process described above, we argue that the decision of who to reach out to resembles the trade-off between exploration and exploitation (March, 1991). Conditional on H1 being true, members that are distant represent opportunities farther away from the locus of action of the focal contributor. This implies that, given the communication challenges, distant contributors are less likely to comply with a feedback request – but can be expected to offer radical benefits such as novel ideas, or connections. On the contrary, those nearby are more likely to comply but offer only incremental benefits. Hence, reaching out to those that are distant represents "exploration of new possibilities" while to those nearby represents "exploitation of old certainties".

The framing of distant interaction as an exploration vs exploitation trade-off implies that goals and accomplishment levels are relevant to the decision process (Cyert & March, 1963). We discuss these below.

Effect of goals on distant interactions

An extensive prior literature has looked into the goals and motivation of *volunteers* that participate in OSS communities. They are known to be motivated by intrinsic factors (von Krogh et al., 2012), as extrinsic pecuniary benefits are not available to them. It suggests that volunteers contribute as they seek recognition and knowledge. In terms of recognition, participation offers them reputational and career benefits (Lerner & Tirole, 2002) and increased diffusion of their innovation (Baldwin & von Hippel, 2011). In terms of knowledge, they learn from others' code (Lakhani & Wolf, 2003), as well as get feedback on their submission (Raymond, 1999).

In addition to volunteers, a second type – *firm-sponsored contributors* also participate. They are employed by their firms to contribute. Participation by firms benefits them commercially by providing opportunities to improve their closed-source proprietary software (Alexy et al., 2018). Given their affiliation, we expect firm-sponsored contributors to be influenced by firm-level commercial goals. However, as these are often ex-volunteers (Corbet & Kroah-Hartman, 2017), we expect them to be also motivated by recognition/knowledge goals.

Given the differences in their goals, we expect volunteers and firm-sponsored contributors to exhibit differences in their exploratory behavior. The inclusion of firm-level commercial goal for firm-sponsored contributors is expected to offer a "safety net" that limits risky exploratory actions towards individual-level goals. Moreover, volunteers are more likely to explore distant timezones as they are likely to find structural and knowledge brokering opportunities (Mell et al., 2021). Thus we posit:

> *Hypothesis 2: In an open-source community, the negative effect of distant timezone on interactions is lower for volunteer contributors versus firm-sponsored contributors.*

Effect of accomplishment levels on distant interactions

Apart from the effect of differences in goals, the problemistic search theoretical argument suggests that differences in accomplishment levels also affect the level of exploration (Greve,

1998; Heath et al., 1999). The theory suggests that lower the level of goal attainment, as compared to a reference, the higher the level of risk-taking behavior individuals are expected to exhibit. This should apply to distant interactions as well, as they represent novel and risky opportunities. For firm-sponsored contributors, with low accomplishment levels, these mechanisms would operate at a lower intensity due to the inclusion of firm-level commercial goals. Therefore, assuming the community maximum as the reference point, we posit:

> *Hypothesis 3: In an open-source community, the reduction in the negative effect of distant timezone on interactions, from firm-sponsored to volunteers, is higher for contributors that have (a) low recognition (low indegree centrality), or (b) low knowledge (low knowledge variety).*

DATA AND METHODS

We leverage the Linux-kernel project as the context to test our hypotheses. Linux is the primary operating system (60-99 percent market share) used within the commercial computing infrastructure (Corbet & Kroah-Hartman, 2017). We obtained archived emails from the public mailing lists for the period 2010-2020. We limit our analysis to 34 sub-systems where emails are completely archived (Duda, 2019). From the email address and metadata, we identify the affiliation of contributors as well as the timezone in which they are located. To be able to generalize results to knowledge work settings, we exclude contributors that participate sporadically. Based on the above, we construct a dataset of 2,879,298 observations of the count of emails sent by contributors to recipients within different timezones in a given month.

Variables

Log (Interactions). The main dependent variable is calculated as the log of the total number of interactions a contributor has initiated. Specifically, for each sender i in month t, we count the number of emails sent to recipients in timezone k.

TZ Difference. The baseline independent variable is calculated as the absolute value of the difference between the sender and the recipient timezone. We also construct a dummy variable *Distant TZ*, which takes the value 0 if sender and recipient are at a timezone distance of less or equal to 1 hour, and takes the value 1 otherwise.

Sender Volunteer. The main independent variable is a dummy variable that takes the value 0 if the contributor is firm-sponsored, and takes the value 1 if the contributor is a volunteer. We identify this distinction based on the email address domain.

Indegree Centrality. We use normalized in-degree centrality as a measure for recognition, i.e. how sought after an individual is within the community. We dichotomize the score to construct *Low Indegree Centrality* which takes the value of 1 if the contributor's in-degree centrality score in a month is lower than average.

Knowledge Variety. We use a text-based measure for knowledge based on email content. We treat each email as a "bag of words", to identify the topical content using a Latent Dirichlet Allocation (LDA) model. We calculate the intrapersonal topic variety for each contributor-month by calculating the mean Herfindahl score. We take the log-inverse such that a higher value represents a higher variety score. We also construct a dummy variable *Low Kw Variety* which takes the value 1 if the contributor's variety score in a month is lower than average.

Control variables. We use individual-level and timezone level time-varying control variables, to control for the effect that the geographical distribution of contributors and their past interactions have on their current period interactions.

RESULTS

For the results reported below, we use a pooled OLS specification linear model with standard errors clustered at the sender level. We include sender timezone fixed effects, recipient timezone fixed effects, and month fixed effects in our main models. For robustness, we have also verified results using sender fixed effects specification models, developed through segmented regression, separately for volunteers and firm-sponsored contributors.

Baseline hypothesis results

We find that OSS communities, like organizations in the prior remote work research (Chauvin et al., 2020; Yang et al., 2021), exhibit a decline in interactions with distance. For a sender, an increase in timezone distance by 1 hour is associated with a 1.2 percent decline in the frequency of interaction with recipients in timezones that are farther away. Interestingly, we find that this effect is not linear. The largest drop is observed when comparing email recipients in the same timezone versus those one hour away. The subsequent decline is lower and becomes insignificant when comparing among those that are 2+ hours away. Based on this, and to make the inference simpler for subsequent analysis, we dichotomize the *TZ Difference* to *Distant TZ* which takes the value 1 when the difference is greater than 1, and 0 otherwise.

Main results

Our primary hypothesis (hypothesis 2) states that the negative effect of distant timezone on interactions is lower for volunteers as compared to firm-sponsored contributors. We find support for this hypothesis in the data. As compared to recipients in a nearby timezone (0/1 hr difference), firm-sponsored contributors have 29.0 percent fewer interactions with recipients in distant timezones (>1 hr difference). The comparative drop for volunteer contributors is only 12.4 percent. These results hold even if we use the variable *TZ Difference*, instead of the dichotomized *Distant TZ*. They indicate that volunteers are more exploratory as they exhibit a more balanced nearby-distant interaction distribution.

Hypothesis 3 helps us test the theoretical mechanisms proposed here. It states that the reduction in the negative effect of distant timezone on interactions, from firm-sponsored to volunteers, is higher for contributors that have low recognition and/or low knowledge. We find support for this hypothesis in the data. For firm-sponsored, we find that those with high recognition (high knowledge) exhibit a 56.7 percent (41.4 percent) drop in interaction frequency with distance. The comparative drop for those with low recognition (low knowledge) is only 14.8 percent (15.3 percent). Similarly, for volunteers, we find that those with high recognition (high knowledge) exhibit a 49.8 percent (31.1 percent) drop in interaction frequency with distance. Those with low recognition (low knowledge) instead exhibit an increase of 3.9 percent (2.4 percent). These results indicate that accomplishment levels are relevant, for both volunteers and firm-sponsored – and those with lower accomplishment are more exploratory.

Moreover, looking at the values in the previous paragraph – at high recognition (high knowledge), the reduction in the drop percentage from firm-sponsored to volunteers is only 6.8 percentage points (10.3 percentage points). The comparative reduction at low recognition (low knowledge) is significantly higher at 18.7 percentage points (17.8 percentage points). As the difference in the nearby-distant interaction balance between the two groups gets amplified at low accomplishment levels, these results, along with those for hypothesis 2, support our theoretical argument that volunteers and firm-sponsored differ in their goals and motivation. For firm-sponsored, recognition and knowledge accomplishment levels correlate positively with distant interactions, but not as much as volunteers – indicating the presence of a goal dimension that suppresses timezone distance exploration.

Additional analysis

To test if firm-level influence affects firm-sponsored contributors, we carry out another test. For the linux-kernel project, contributors use "signed-off-by:" or "acked-by:" tags in their code to denote if any contributor apart from the submitter is involved in the development process (Corbet & Kroah-Hartman, 2017). We tag a firm-sponsored contributor as being subject to *high firm-level influence* if the code submitted is signed off by another individual from the same firm. We find that firm-sponsored contributors subject to high firm-level influence exhibit a 54.2 percent drop in interaction frequency with distance. The comparative drop for those subject to low influence is only 22.6 percent. This indicates that firm-level influence plays a role in distant interactions undertaken, lending further support to our theoretical arguments.

DISCUSSION AND CONCLUSIONS

In this paper, we explore what individual-level factors are associated with a higher frequency of distant silo-spanning interactions. In general, it is an important question for organizations as perfect silos can rarely be designed. In the context of remote work, this question becomes more pertinent. Here, we leverage OSS communities (all-remote organizations) as the context and timezone difference as the distance dimension to explore answers. We find that worker-level factors, their goals and accomplishment levels, are relevant to the level of distant interactions undertaken. When they have agency in choosing partners, those with goals better aligned to the distance dimension and those with lower accomplishments are more likely to undertake risky distant interactions. The underlying mechanisms are expected to be similar to the exploration-exploitation trade-off proposed in March (1991).

The findings from this study help advance the remote work literature (Bloom et al., 2015; Chauvin et al., 2020; Choudhury et al., 2021), by suggesting that recruitment and incentivization are important levers within a firm's remote-work strategy. With a better understanding of the silo-intensification challenges of remote, providing emphasis on understanding workers' goals and motivation and offering those recruited the right incentives to accomplish those goals – are important subsequent steps for firms. These would be especially pertinent to places where workers self-select where (and with whom) they work.

REFERENCES AVAILABLE FROM THE AUTHOR(S)