Large-Scale Analysis of New Employee Network Dynamics

Yulin Yu
University of Michigan
Ann Arbor, MI, USA
yulinyu@umich.edu

Longqi Yang
Microsoft
Redmond, WA, USA
longqi.yang@microsoft.com

Siân Lindley
Microsoft
Cambridge, United Kingdom
sianl@microsoft.com

Mengting Wan
Microsoft
Redmond, WA, USA
mengting.wan@microsoft.com

ABSTRACT

The COVID-19 pandemic has accelerated digital transformations across industries, but also introduced new challenges into workplaces, including the difficulties of effectively socializing with colleagues when working remotely. This challenge is exacerbated for new employees who need to develop workplace networks from the outset. In this paper, by analyzing a large-scale telemetry dataset of more than 10,000 Microsoft employees who joined the company in the first three months of 2022, we describe how new employees interact and telecommute with their colleagues during their "onboarding" period. Our results reveal that although new hires are gradually expanding networks over time, there still exists significant gaps between their network statistics and those of tenured employees even after the six-month onboarding phase. We also observe that heterogeneity exists among new employees in how their networks change over time, where employees whose job tasks do not necessarily require extensive and diverse connections could be at a disadvantaged position in this onboarding process. By investigating how web-based people recommendations in organizational knowledge base facilitate new employees naturally expand their networks, we also demonstrate the potential of web-based applications for addressing the aforementioned socialization challenges. Altogether, our findings provide insights on new employee network dynamics in remote and hybrid work environments, which may help guide organizational leaders and web application developers on quantifying and improving the socialization experiences of new employees in digital workplaces.

CCS CONCEPTS

• **Applied computing**; • **Information systems** → *Data mining*;

KEYWORDS

Future of Work, Onboarding, Organizations, Social Networks, Temporal Dynamics

1 INTRODUCTION

New employee onboarding often refers to the process where new hires are integrated into the organization. This process is widely acknowledged to be vital for organizations and positively related to employee satisfaction, productivity and retention [13–15, 25]. A successful new employee onboarding process not only involves knowledge acquisition, skills training, function understanding, but also encourages new hires to adapt to the organizational culture and socialize with their authentic selves, therefore establishing the sense of belonging [14]. Although there exist extensive studies on conceptualizing this organizational socialization process [25, 30, 34], empirical evidence about how new hires communicate and socialize with other employees throughout their onboarding period is very limited to date.

The COVID-19 pandemic has caused a fundamental shift to remote work for many information workers [4]. It has accelerated the digital transformation of many workplaces, where employees became reliant on telecommunications such as remote meeting and instant messaging. Despite the flexibility introduced by this working paradigm change, it also brings new onboarding challenges, largely due to the lack of spontaneous and informal in-person conversations [9, 11, 39]. Studies reveal that the collaboration network of information workers became more static and siloed with the shift to firm-wide remote work, which makes it more difficult for employees to acquire and share novel information across the network [58]. This struggle for organizational socialization can be aggravated on new employees who need to not only acquire necessary knowledge and skills in a short time, but also develop their workplace networks from scratch. For example, recent research on software developers indicates that new hires struggled to ask for help while working remotely (due to the isolation between teammates, scheduling difficulties, the lack of hallway conversations, etc.) and communicated with fewer people overall compared to those who joined pre-pandemic [43, 50].

This implies the need for different tools and approaches to supporting onboarding in remote settings. Additionally, the shift to remote work enables new opportunities to explore the impact of such tools, by empirically quantifying the socialization behavior of

new employees from their workplace digital traces, and the extent to which they can be augmented or improved by web applications. In this paper, we empirically study the socialization behavior of new employees on a large-scale post-pandemic workplace communication dataset from Microsoft, which includes rich telemetry data of employees on their remote video/audio meetings, emails, and instant messages. In addition, we measure the usage patterns of a web-based people recommendation engine in the workplace and demonstrate its potential in addressing onboarding challenges. The network characteristics we measure include:

1) The number of distinct connections (or social ties) each employee maintains in each week;
2) The frequency in which each individual communicates with other employees every week;
3) The extent to which an employee bridges communities within an organization.

The first two dimensions approximate the number of potential information sources and the total amount of information exchanged with the focal employee. The third dimension captures one's network structural diversity, which has been shown to play a key role for multiple individual and organizational outcomes [3, 10, 29, 42, 54]. Specifically one's neighborhood may consist of several disparate clusters, reflecting their co-workers from different organizational contexts. Prior studies indicate that having diversified connections to bridge different ("well-seperated") organizational groups allows individuals to access novel and non-redundant information [10, 20, 29, 42, 52].

The contributions of this paper are summarized as follows.

- We present the first large-scale, empirical study of the communication network dynamics of new employees in a multinational technology organization in the *remote* work environment.
- Results from our analysis 1) reveal the general time trends of new employee network dynamics across different communication media during remote onboarding, 2) identify the possibly disadvantaged communities in the onboarding network development process, and 3) demonstrate the potential of web-based applications to address employee socialization challenges.
- Our work points to several important insights, which may help guide organizational leaders and web application developers on how to quantitatively evaluate and improve the employee socialization experiences at digital workplaces.

2 HYPOTHESES

Prior studies suggest that new employee socialization is an assimilation process where network development is arguably a key component [18, 25, 41]. This process may unfold over a newcomer's first several months (or years) with an intense ramp-up period at the beginning [25]. On the other hand, recent studies suggest that asynchronous communications have largely increased with the broad shift to remote work [58]. Therefore we speculate the network development patterns of new employees can be different across communication media. Together these inform us the following two hypotheses.

HYPOTHESIS (**H1: Time Trends Across Communication Media**). *There exists an onboarding period where communication*

networks of new employees are continually developing. These network developing trends may vary across different communication media (i.e., remote meetings, emails, and instant messages).

HYPOTHESIS (**H2: New Hires vs. Tenured Employees**). *New employees are less "connected" than tenured employees during their onboarding period in the organization (i.e., smaller network size, lower communication intensity, lower structural diversity). However, these differences shall reduce as the tenure of a new employee increases.*

We speculate there may exist heterogeneity among new employees because of the intrinsic differences of their job duties and social status, thus proposing the following hypothesis.

HYPOTHESIS (**H3: Heterogeneity Among New Hires**). *There exists heterogeneity among new hires (across different managerial positions and different job functions) in how they develop their workplace networks during the onboarding period.*

In addition to the existing investment of human capital on the new employee onboarding process [26, 37, 47], with the great workplace digital transformations, we are interested in if new web-based experiences can be enabled to help new employees develop their networks at digital workplaces. Previous studies demonstrate that knowledge acquisition and information seeking are key components in the onboarding process and positively relate to new employee socialization outcomes [5, 25]. Therefore we start our exploration by investigating if embedding "people recommendations" into organizational knowledge management systems can facilitate new employees in connecting with colleagues they otherwise wouldn't discover (i.e., bridging connections). In this way, knowledge itself serves as an information broker and triggers natural interactions between employees through knowledge sharing, thus improving the network structural diversity.

HYPOTHESIS (**H4: Web-Based People Recommendations**). *New employees who have engaged with the "knowledge-based people recommendations" are likely to have more bridging connections.*

We leverage the user engagement data on an enterprise knowledge management product—Microsoft Viva Topics[1]—to verify the above hypothesis. Microsoft Viva Topics is a commercial product which leverages machine learning techniques to compile and surface knowledge within an organization. It not only surfaces the textual description, but also suggests people in the organizations who specialize in this given knowledge entity. By analyzing how new employees utilize the "people recommendation" feature ("Pinned People" and "Suggested People") in this application, we are able to compare the potential differences of their network trends.

3 DATA AND METHODS

3.1 Data Description

We collected a large-scale dataset from Microsoft's full-time employees describing their communication activities from December 2021 to September 2022. The collected data contains anonymized events on major workplace communication platforms: Microsoft Teams for *remote meeting* and *instant messaging*, and Microsoft Outlook for *email services*. Meanwhile, we collected a user profile

---

[1]https://www.microsoft.com/en-us/microsoft-viva/topics

dataset containing the organizational group, job function, managerial status, location as well as the local team size of each individual user. Combining these two datasets, we identified two groups of users and conducted statistical analysis on their communication networks.

- **New Hires.** We use the first date of each user's outbound communication activities (i.e., meetings joined, emails sent, instant messages sent) as the estimated employee start date. To mitigate potential biases introduced by the planned absences (e.g., planned vacations) for these start date estimations, December 2021 is used as a burn-in period to allow users populating in the dataset. After that we identified 11,083 users whose communication activities first started between 2022/01/01 and 2022/03/27 (12 weeks) as the new hire group in our analysis. For each individual user, we then construct weekly snapshots of their communication networks for 24 consecutive weeks from their start week.[2]
- **Matched Tenured Employees.** In order to assess the network differences between new hires and tenured employees, we leveraged the above user profile information and identified a matched sample of tenured users in our dataset. For each identified new hire, we matched a tenured user who started before 2022/01/01 on the same profile features including organizational group (63 distinct values), job function (engineer or non-engineer), managerial status (individual contributor or manager), location (headquarter or non-headquarter), local team size (50 distinct values). Only 89 out of 11083 newcomers cannot be matched and are removed in the subsequent analysis. We then extracted their communication networks within exactly the same 24-week observation time window as the one for the paired new hire.

3.2 Constructing Ego Networks

We're interested in the ego network of each individual in the above identified user groups and how it changes over time. The notation and definitions are formally introduced as the follows.

Given an observation week t, a communication medium (meeting, email, or instant messaging), for each individual k, we define the undirected *ego network* as $\mathcal{G}_k^{(t)} = \{\mathcal{V}_k^{(t)}, \mathcal{E}_k^{(t)}, \mathcal{W}_k^{(t)}\}$. Here $\mathcal{V}_k^{(t)} = \{k\} \cup \mathcal{N}_k^{(t)}$ represents node k as well as its immediate neighbors[3] ($\mathcal{N}_k^{(t)}$). $\mathcal{E}_k^{(t)}$ is an undirected edge set which denotes the observed reciprocal connections among nodes in $\mathcal{V}_k^{(t)}$ within the given time window t. $\mathcal{W}_k^{(t)}$ represents the corresponding edge weights.

We first focus on each individual's *one-on-one* networks where $\mathcal{E}_k^{(t)}$ is restricted on the communication interactions with only two participants. Following previous studies [20, 36], we define the edge weight $w_{ij}^{(t)} = \sqrt{w_{i\rightarrow j}^{(t)} \times w_{j\rightarrow i}^{(t)}}$ between two individuals i and j as the geometric mean of the number of directed interactions during the observation time period t. For example, for each employee's

email network, $w_{i\rightarrow j}^{(t)}$ denotes the number of emails that i sent to j in week t. To ensure the reciprocity of the observed connections, we define the ego network $\mathcal{G}_k^{(t)}$ comprising the edges with only non-zero weights, i.e., $w_{ij}^{(t)} > 0$. We consider an alternative group interaction network for robustness test, which accounts for employees' group communications with less than 10 participants.[4]

Following prior studies [20, 24, 52], we consider network metrics to characterize each individual's ego network from three different perspectives: network size, intensity, and structural diversity.

- **Number of distinct connections.** This metric is defined as the total number of employees connected to the focal individual k within the observation time window t, which directly measures the *size* of one's neighborhood, i.e., $|\mathcal{N}_k^{(t)}|$.
- **Sum of edge weights.** We sum over the weights of all edges connected to the ego node k to assess its overall communication *intensity*, i.e., $W_k^{(t)} = \sum_{j \in \mathcal{N}_k^{(t)}} w_{ij}^{(t)}$. Note the units of edge weights vary in communication media, where the weights for meeting, email, and IM networks are calculated based on the number of joint meetings, emails exchanges, and joint chat sessions respectively.
- **Number of ego components.** We leverage the number of ego components ($C_k^{(t)}$) to measure how much each individual k is bridging surrounding communities. This measure is defined as the number of connected components that remain in the ego network $\mathcal{G}_k^{(t)}$ when the focal node k as well as its incident edges are removed. This metric is frequently adopted in prior network science studies to describe the *structural diversity* of one's ego network [20, 52].

3.3 Organizational Attributes

In addition to the above network metrics, we also consider the following organizational attributes as control variables or predictors in our analysis.

- **Organizational Group.** This attribute represents the highest-level organizational group within Microsoft that each individual employee belongs to. In our dataset, it is defined as the parent nodes one step below the CEO (the root node) in the formal organizational chart.
- **Job Function.** We consider a binary variable $x^{(eng)}$ to represent if an individual is at the engineering function role, i.e., engineer ($x^{(eng)} = 1$) versus non-engineer ($x^{(eng)} = 0$).
- **Managerial Status.** This attribute is defined as an indicator $x^{(mngr)}$ to reflect if the employee is in the management position, i.e., manager ($x^{(mngr)} = 1$) versus individual contributor ($x^{(mngr)} = 0$).
- **Location.** We introduce a binary variable to represent if each employee's office (whether it is a home office or not) is located in the same metropolitan area as Microsoft's headquarter office.
- **Local Team Size.** We calculate the number of employees who directly report to the same direct manager as the target individual and use this variable to assess the number of peer co-workers in one's direct team.

---

[2]This onboarding window is selected because a) the telemetry data collection has to comply with our data retention policy and thus being restricted to a limited timeframe; and b) this 24-week window is where most onboarding activities take place.

[3]Note $\mathcal{N}_k^{(t)}$ includes all Microsoft full-time employees who have connected with individual i in the given week t. The neighboring employees here are not restricted in the user groups identified in Section 3.1.

[4]Edge weight ($w_{i\rightarrow j}^{(t)}$) in the group interaction network is relaxed as the number of directed interactions, normalized by the number of recipients within these interactions.

3.4 Models

As described in Section 3.1, we calculate the network statistics of each identified new hire and the matched tenured employee's weekly ego networks across different communication platforms over 24 consecutive weeks. To examine our proposed hypotheses, we follow the mixed effect framework and perform a series of longitudinal ordinary least squares (OLS) regressions. The regression framework can be described as below

$$y_k^{(t)} \sim \textit{variables of interests} + \textit{control variables} + \eta_t + \epsilon_k. \quad (1)$$

For each individual k in the observation week t, the dependent variable $(y_k^{(t)})$ encodes one of the following three network metrics—the number of distinct connections $(|\mathcal{N}_k^{(t)}|)$, the sum of edge weights $(W_k^{(t)})$, and the number of ego components $(C_k^{(t)})$. In order to control for the potential time unit-specific bias (e.g., the lack of activeness during the holiday week), a time-dummy variable η_t is included to capture these time fixed effects. To model the potential variations across different individuals, ϵ_k is introduced to capture the random effects. For each individual new hire k, we also introduce *Onboarding Week* as a numeric time-varying variable. For each new hire k within the given observation week t, the corresponding *Onboarding Week* can be defined as $t - \tilde{t}_k$, where $\tilde{t}_k$ denotes the start week of the individual k. We discuss how we operationalize this framework to examine each hypothesis as the following. A summary of the detailed model designs are included in Table 1.

To test whether new hire's network connectivity metrics increase over the monitored 24-week onboarding period and how these metrics differ across various communication platforms (**H1**), we perform the regression analysis on new hires' meeting, email as well as instant messaging (IM) ego networks. We examine the *Onboarding Week* as the variable of interests in Eq. (1), and the organizational attributes (as defined in Section 3.3) are utilized as control variables (**M1**). The estimated coefficient of *Onboarding Week* allows us to assess the direction as well as the speed of network connectivity changing over the new hire onboarding period.

To test the potential differences between new hires and tenured employees' networks (**H2**), we construct a combined dataset with both new hires as well as the matched tenured employees. We introduce a binary variable *Tenure Group* $x_k^{(tenure)}$, where $x_k^{(tenure)} = 1$ indicates the individual k belongs to the new hire user group and $x_k = 0$ otherwise. We use all organizational attributes as control variables and test the effect of *Tenure Group* (**M2**). The estimated effect then corresponds to the average difference between the new hire group and the tenured employee group within the observation time window.

To examine the potential differences of network connectivity metrics as well as their growing speeds between new managers and new individual contributors (**H3**), we consider the *Managerial Status*, *Onboarding Week* as well as their *interaction* as the variables of interests (**M3.a**). All other organizational variables are used as the control variables. In this model, the estimated effect of the *Managerial Status* reflects the average difference between new managers and new individual contributors while the interaction between *Managerial Status* and *Onboarding Week* indicates the difference of growth rates of their networks over time. By switching

the *Managerial Status* variable and the *Job Function* variable, we perform a similar regression analysis to examine the potential differences between engineers and non-engineers during their 24-week onboarding time (**M3.b**).

To test whether the adoption of the knowledge-based "people recommendation" application associates with a new hire's networking behavior (**H4**), we introduce an additional binary variable *People Recommendation* $x_k^{(rec.)}$ to categorize new hires based on their engagement behavior on these recommendations, where $x_k^{(rec.)} = 1$ indicates the individual k has clicked or viewed the "Pinned People" and "Suggested People" feature in Viva Topics within the tracked 24-week onboarding period and $x_k^{(rec.)} = 0$ otherwise. Similar to the previous setup, we test *People Recommendation*, *Onboarding Week* as well as their first-order *interaction* and include all organizational attributes as control variables (**M4**). By testing *People Recommendation*, we aim to identify its potential correlation with one's network connectivity metrics. By testing the interaction term (*People Recommendation × Onboarding Week*), we explore its potentials on one's network connectivity growth rate during the onboarding period.

4 RESULTS AND DISCUSSION

4.1 Communication Trends

We start by presenting model-free time trends for different network metrics. These trends describe how the communication networks of new hires change across different communication media over the 24-week onboarding period. By comparing these new hires' trends to the matched tenured employees within the same observation window, we highlight how new hires assimilate their communication behavior to tenured employees over time. Then we conduct a series of statistical regression analyses (as described in Section 3.4) to formally examine our hypotheses.

Communication Media. Figure 1 shows the average weekly time series for network connectivity statistics on different communication platforms across different tenure groups (new hires versus matched tenured employees). Focusing on the trends of new hires' network statistics, we observe different onboarding patterns from this figure with respect to the use of communication media (**H1**). We find new hires' one-on-one *meeting* and *instant messaging* (IM) networks are overall expanding over the 24-week onboarding period with respect to their network size, intensity, and structural diversity, where the size and the intensity of meeting networks rapidly converge in the first few weeks while the IM networks are continually growing on all three network metrics. On the other hand, we notice new hires generally started with larger and more intense *email* networks and these metrics quickly dropped in their first few onboarding weeks.[5] To validate these observations, we test the *Onboarding Week* variable in our statistical regression model (**M1**) over the 24-week onboarding window and present the corresponding coefficients with the 95% confidence intervals (CIs) in Figure 2. We find a statistically significant weekly growth of new hires' IM networks on all three network metrics over the 24-week observation time (+0.131 per week on network size, $p < 0.001$;

---

[5]A possible explanation for this observation is in the first few weeks, newcomers often engage with ad-hoc activities requiring formal written communications (e.g., introduction emails), thus leading to the bump in their email network trends.

Hypothesis	Model	Scope	Variable(s) of Interests	Control Variables
H1	M1	New hires only	Onboarding Week	Organizational Group, Job Function, Managerial Status, Location, Local Team Size
H2	M2	New hires + tenured employees	Tenure Group	Organizational Group, Job Function, Managerial Status, Location, Local Team Size
H3	M3.a	New hires only	Managerial Status, Onboarding Week Managerial Status × Onboarding Week	Organizational Group, Job Function, Location, Local Team Size
H3	M3.b	New hires only	Job Function, Onboarding Week Job Function × Onboarding Week	Organizational Group, Managerial Status, Location, Local Team Size
H4	M4	New hires only	People Recommendation, Onboarding Week People Recommendation × Onboarding Week	Organizational Group, Job Function, Managerial Status, Location, Local Team Size

Table 1: Summary of the model designs, where each model is conducted on meeting, email, and IM networks separately.
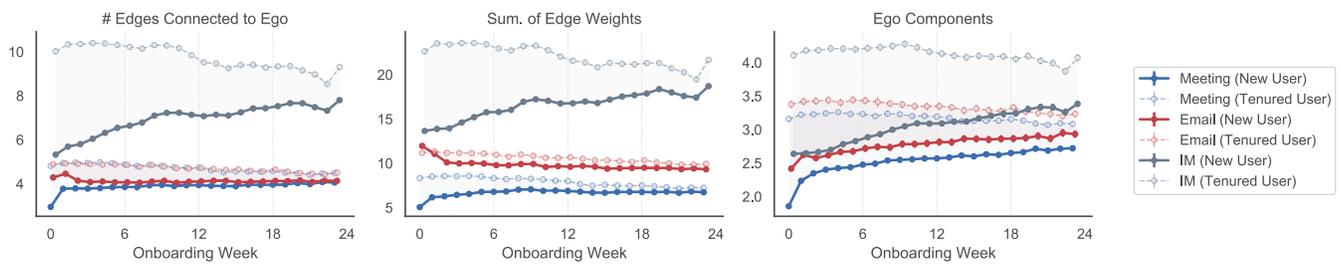

Figure 1: Time trend comparisons on different communication media—meeting, email, and instant messaging (IM)—which are highlighted using different colors. Data points represent the means of network statistics across users in the given observation week and error bars indicate the 95% confidence intervals (covered by the markers in some places). The differences between new hires and tenured employees on each platform are shaded using its corresponding color.
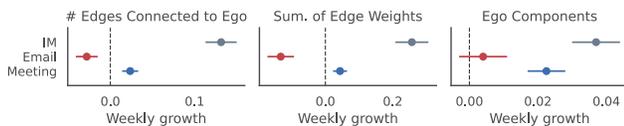

Figure 2: Weekly growth rates (with 95% CIs) of network statistics (M1).

+0.257 per week on network intensity, $p < 0.001$; +0.037 per week on structural diversity, $p < 0.001$). These effects on IM networks appear to be significantly larger than those on other communication media, which confirms the fast-growing patterns we observed about new hires' networking behavior via instant messaging.

In addition to our observations about communication media choice among new hires, we also notice a strong preference from tenured employees on one-on-one communications over instant messages (i.e., larger communication network size, intensity, and diversity). Although employees' choices on communication media can be affected by various factors, including both firm-wide guidelines and personal inclinations, this preference for instant messages is in fact consistent with the findings from recent studies where asynchronous and informal communications have increased among information workers with the shift to remote work [58]. The growing trend of communicating through instant messages may also reflect new hires' assimilation process, where they gradually adopt the communication media choice from tenured employees in the remote and hybrid work environment.

New Hires vs. Tenured Employees. Figure 1 also reveals clear differences between the communication networks of new hires and tenured employees (**H2**). Compared to new hires, tenured employees generally connect with more colleagues every week (larger network size), communicate more intensively with other colleagues (larger edge weights), and maintain more diversified connections within the company (higher structural diversity) across all communication media. Despite their time trend differences on different communication media in the first few onboarding weeks, we observe new hires' network statistics eventually trend towards the states of tenured employees at the later stage of the tracked 24-week onboarding period. We investigate this pattern by conducting finer-grained statistical regressions to test the *Tenure Group* variable (**M2**) on every 6 onboarding weeks and report the results in Figure 3. We notice the network statistics of new hires and tenured employees are significantly different ($p < 0.001$), while these differences are diminishing as the tenure of a new employee increases.

Through the above analysis, the socialization aspect of the organizational assimilation process can be reflected on these time trends of new hires' communication networks, where their network size, intensity, and diversity are approaching the states of tenured employees in the organization. Regardless we still find statistically significant gaps on these metrics between different tenure groups at the latest state of the tracked onboarding period, implying it may take longer than 24 weeks (6 months) to have new employees fully "onboarded", i.e., statistically indistinguishable from tenured employees on network size, intensity and diversity.
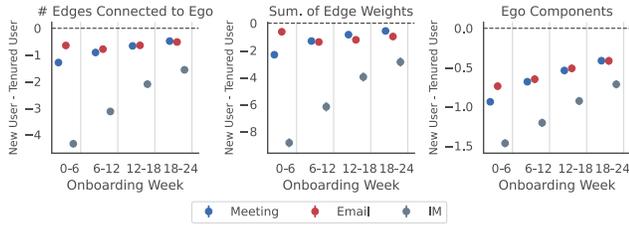
Figure 3: Differences of network statistics (with 95% CIs) between new hires and tenured employees within each 6-week testing window (M2).

## 4.2 Heterogeneity Among New Employees

While we observe a general trend that new hires are assimilating their networks to those of tenured employees, we hypothesize there may exist heterogeneity across new employees due to the fundamental differences of their job duties, social status, and day-to-day workflows [21, 44, 45, 55] (**H3**). We conduct descriptive analysis and statistical regressions to investigate their heterogeneity on the following two dimensions: managerial position (manager vs. individual contributor) and job function (engineer vs. non-engineer).

Managers vs. Individual Contributors. Among the 11,083 new hires we identified, we observe 6.37% of them are managers and 93.63% of them are individual contributors. Figure 4a presents the weekly trends of network statistics for employees with different managerial statuses, where we observe a distinct pattern that compared to individual contributors (ICs), managers connect with more employees in the organization and their communication networks tend to be more intense and more structurally diverse. Note these differences are consistent across both new hires and tenured employees, reflecting the intrinsic nature of their role types: manager roles are usually obliged to conduct people coordination while IC roles are often expected to focus on "hands-on" work practices. We further investigate the differences among new hires and notice that managers' communication networks generally grow and stabilize faster than individual contributors. To quantify the differences of their network growth rates, we test the first order interaction between *Managerial Status* and *Onboarding Week* (**M3.a**) and report these effects in Figure 5a. We find the growth rates on various network metrics are significantly different across managers and ICs. In particular, we find new managers' networks grow significantly and consistently faster than ICs on their structural diversity metrics across all communication media (+0.023 per week on remote meeting, $p < 0.001$; +0.032 per week on email, $p < 0.001$; +0.006 per week on instant messaging, $p = 0.002$).

A possible explanation here is in comparison with individual contributors, extensive communications and coordinations are required in managers' daily work (e.g., host routine one-on-one meetings with their direct reports, maintain frequent check-ins with representatives from other groups), which creates both demands and opportunities for new hires who are in manager roles to rapidly develop their networks. On the contrary, individual contributors may be put on a disadvantaged position at their onboarding process regarding their social network building. Such a disadvantage can be worsen with the current shift to remote work due to the absence of
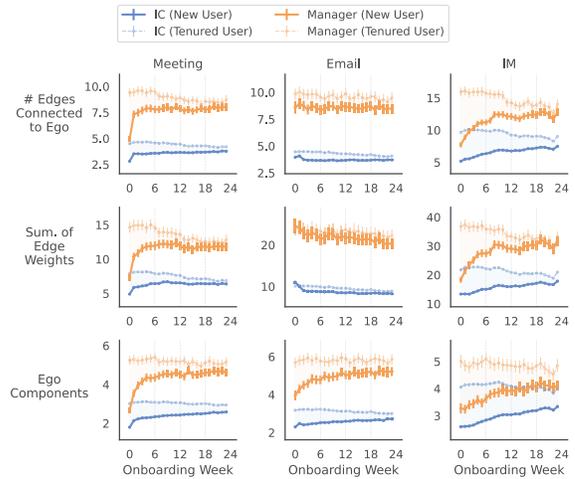
(a) Heterogeneity w.r.t. Managerial Status (Manager vs. Individual Contributor).
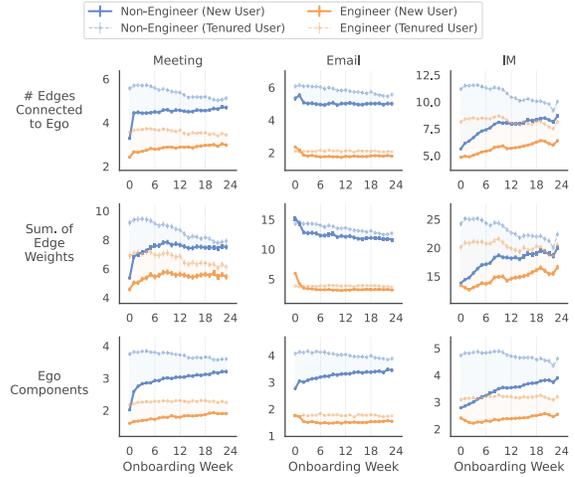
(b) Heterogeneity w.r.t. Job function (Engineer vs. Non-Engineer).

Figure 4: Time trend comparisons across different new hire groups on their meeting, email, and instant messaging networks (average network statistics with 95% CIs).

spontaneous in-person connections at workplace (e.g., impromptu lunches, hallway conversations) [6, 19, 28, 40, 58].

Engineers vs. Non-Engineers. We observe that 35.67% of our identified new hires are engineers while the other 64.33% are in non-engineering roles (e.g., program managers, sales representatives). Figure 4b presents the weekly trends of network statistics for employees with different job functions. Similarly we observe a clear pattern that engineers connect with fewer employees, with less total communication intensity, and their networks tend to be less structurally diverse compared to non-engineers. We also notice that new engineers' communication networks in general expand slower than non-engineering new hires, especially in terms of their network structural diversity. By testing the interaction between *Job Function* and *Onboarding Week* through regression analysis (**M3.b**), we are able to validate these observations (as shown in Figure 5b). Notably

(a) **Manager vs. Individual Contributor.**

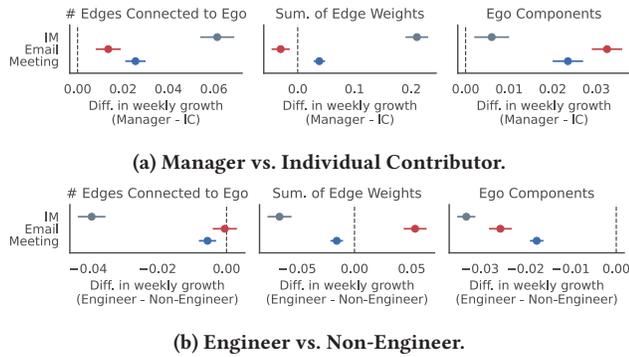(b) **Engineer vs. Non-Engineer.**

Figure 5: Heterogeneity of network growth rates (with 95% CIs) across different new hire groups.

we find engineering new hires' network growth rates on structural diversity are significantly lower than non-engineers across all communication platforms (-0.017 per week on remote meeting, $p < 0.001$; -0.026 per week on email, $p < 0.001$; -0.033 per week on instant messaging, $p < 0.001$).

Relative to non-engineers (e.g., program managers, sales representatives), we conjecture new employees in the engineering role may be less in need of large and diverse communication networks in order to perform their daily jobs tasks, therefore lacking opportunities to organically grow their networks (especially on weak and bridging ties, which is reflected on the number of ego components).

4.3 Opportunities for Web-Based Applications

The above results reveal that it may take a long time (> 24 weeks) to fully onboard new employees from the socialization perspective, and certain groups of new hires could be at a disadvantaged position in this process because of the limited opportunities to expand networks when performing day-to-day requirements of the job. Nevertheless, workplace digital transformations during COVID-19 pandemic also introduced opportunities to address this connectivity crisis by enabling new experiences on web-based digital applications. In addition to provide disruptive interventions such as intentionally hosting virtual "watercooler" sessions, we also investigate if web-based nudges can be seamlessly embedded into an employee's day-to-day workflows, thus triggering natural interactions with colleagues whom they otherwise wouldn't connect with and improving their network structural diversity.

Knowledge-Based People Recommendation. Given knowledge acquisition is a critical component of the onboarding process for information workers, we start by investigating if providing "people recommendations" in organizational knowledge management systems can help new employees connect with colleagues otherwise they wouldn't know. Specifically, we test if there are differences of network dynamics between new employees who have or not engaged with (viewed or clicked) these "people recommendations" (i.e., the "Pinned People" and "Suggested People" feature in Microsoft Viva Topics) during their onboarding period. These statistical tests (**M4**) are performed on every 6 onboarding weeks and the results are presented in Table 2, where we observe statistically significant differences regarding network diversity and the weekly growth rate

Onboarding Week	Email	IM	Meeting
0-6	0.194***	0.141***	0.028
6-12	0.200*	0.033	-0.024
12-18	0.120	0.140	0.051
18-24	0.242	0.097	0.088

(a) **Differences of network structural diversity ($^*p < 0.05$, $^{**}p < 0.01$, $^{**}p < 0.001$).**

Onboarding Week	Email	IM	Meeting
0-6	0.026**	0.018**	0.029***
6-12	0.013	0.026***	0.019**
12-18	0.015	0.011	0.009
18-24	0.005	0.011	0.005

(b) **Differences of the weekly growth of network structural diversity ($^*p < 0.05$, $^{**}p < 0.01$, $^{**}p < 0.001$).**

Table 2: The relationship between new hires' engagements with the "Pinned People" and "Suggested People" feature in Microsoft Viva Topics and their network structural diversity (M4), i.e., the number of ego components.

of this metric in the first 12 onboarding weeks, indicating the potential positive effects of presenting "people recommendations" on new employees' network diversity. However we find such effects become less significant in the late onboarding stage. A possible explanation is new hires in our dataset are normally leverage this organizational knowledge base (Microsoft Viva Topics) as an onboarding tool to explore the organizational knowledge taxonomy in their early days. As the onboarding process progresses, their workflows start shifting from exploring organizational knowledge to executing job tasks. This observation also implies future research directions on fully investigating potential canvases and approaches to present "people recommendations" based on the employee's working context.

4.4 Additional Analysis on Group Interaction

Our analysis thus far has considered the one-on-one communication patterns of newcomers during the virtual onboarding period. We also conducted a separate analysis for group communications (with less than 10 participants in each group interaction). Our findings remain consistent on the significant gaps of all network statistics between new hires and tenured employees after the six-month onboarding phase, the heterogeneity across different managerial positions and job functions, and the positive association between the usage of web-based people recommendations and the network structural diversity of new hires. More details are included in the appendix.

5 RELATED WORK

Remote work. Remote work is an important topic to study across academia and industry. Remote work-related studies have covered various areas, such as communication challenges [31, 43], well-being [11, 39], management [49], etc. While most work focuses on distributed work or remote work due to geographical distance [8, 32, 35, 38, 40], ours focuses on virtual trending caused by COVID-19. Related to our study, a stream of research emphasizes

the impact of remote work on organizational workers' connectivity in general. Most of them feature the strength of existing or new interpersonal ties [17, 56–58]. For example, scholars find that firm-wide remote work triggered by COVID leads to a more siloed network among information works with the decrease weak and bridging ties [58] and in general, ties are more difficult to form in remote work settings [17].

This could have significant ramifications, as connections between workers in in-person settings have been associated with organizational outcomes such as building successful collaborations [33], relationships, and with knowledge transfer [2]. Organizational knowledge is inherently social, and is dependent on shared language, narratives, identification, norms, commitment and trust [1]. Additionally, the ability to build connections enables workers to self-organize and share expertise [1]. Research prior to the pandemic indicates that while face-to-face conversation is preferred for relationship building, ideation, and problem solving, communication technologies such as IM and social network sites perform complementary purposes including building and maintaining connections [51]. In this paper, we focus on virtual onboarding rather than studying full-time tenured organizational workers, and study how use of a range of technologies, including one that highlights other organization members to the user, are related to workers' connectivity.

Virtual onboarding. While traditional, in-person onboarding has been frequently studied in the context of organization outcomes [12, 18, 27, 46, 48], virtual onboarding is still an open area. Most studies focus on virtual onboarding challenges related to specific job functions such as engineering [8, 22, 27, 43]. For example, a very relevant qualitative study by Rodeghero et al. indicates that remote work harms social connection, but changing the mode of interaction (such as turning on the camera during meetings) could potentially improve socialization [43]. Previous work also discusses AI tools (e.g., voice assistant tools) [23, 53], or human effort (e.g., hosting virtual watercooler events [7], or providing checklists or mentorship [16, 27]) that can be used to improve onboarding. Our work further explores the remote onboarding process in the context of connectivity and provides potential web solutions that connect people with the expert knowledge needed to improve employee connectivity.

6 CONCLUSION AND FUTURE WORK

In this paper, we investigated the dynamics of new employees' workplace communication networks throughout a 24-week onboarding period. We also explored the potential web-based applications to address the socialization challenge and summarize our findings as follows.

- While new hires are less connected with colleagues in the organizations compared to tenured employees in the 24-week onboarding period (smaller network size, lower communication intensity and lower structural diversity), their networks are gradually expanding over time.
- Time trends of new hires' network statistics vary across different communication media, managerial positions, and job functions. We observed that new employees, whose day-to-day job tasks

are not necessarily reliant on extensive and diverse collaborations (e.g., individual contributors, software engineers), can be put into a disadvantaged position regarding organizational socialization in the onboarding process.
- We explored the opportunities for web-based applications to address the new employee socialization challenges. Early evidence revealed that presenting "people recommendations" within employees' workflows may benefit their workplace network development during the early onboarding period (first 12 weeks).

This study, to the best of our knowledge, presents the first large-scale empirical investigation on communication network dynamics of new employees in the technology industry. This study is not without limitations. First, given the availability and retention policy of the dataset, our study was restricted to a 24-week onboarding window. This constraint can be further relaxed if more data become available since we empirically revealed that it may take longer than 24 weeks to observe new hires' network statistics fully converge to the states of tenured employees. In this way, one can evidently document the convergence time of new employees with different job roles and empirically understand the dynamics of new hires' networks in the long term. These quantitative analyses can also be complemented with qualitative studies to deeply understand new employees' networking behavior. Second, new employees studied in this work mostly followed the remote onboarding guidelines within the organization, thus limiting our scope within the remote work context. We acknowledge that the current and future hybrid work policies could impact new employees' socialization experiences. Therefore a future research direction is to extend the empirical analysis and investigate the potential onboarding challenges as well as opportunities induced by the emerging trend of hybrid work. Third, our study is limited to a single firm and, for the most part, information workers. A future direction could be expanding the analysis scope and investigating the potential heterogeneity regarding new employee network dynamics across different firms and industries.

We quantitatively described how new employees develop their workplace network over a 6-month onboarding period in this study. Our results confirmed the struggle and the need of "staying connected" from new employees in the remote work environment. Our study revealed there is no single network development path for all job functions. We also want to highlight that overlooking the socialization needs from disadvantaged employee groups may result in negative effects on individuals, organizations, and social fairness. In addition to the investments of human capitals on new employee onboarding experiences, we demonstrated that recent workplace digital transformations have introduced possibilities for web-based technologies to address the new employee socialization challenges. Here, we'd like to call for both human capital investments (e.g., host social events) and technical solutions (e.g., people recommendations, virtual "watercooler" sessions) on addressing the socialization needs of new hires. Network metrics in this study can also be leveraged as a toolkit to identify/self-identify newcomers who may need help. We hope our study can shed some lights on these opportunities and encourage future research efforts from the web community to keep advancing digital workplaces.

REFERENCES

[1] Mark S Ackerman and Christine Halverson. 2004. Sharing Expertise: The Next Step for. *Social capital and information technology* (2004), 273.

[2] Tammy D Allen, Timothy D Golden, and Kristen M Shockley. 2015. How effective is telecommuting? Assessing the status of our scientific findings. *Psychological science in the public interest* 16, 2 (2015), 40–68.

[3] Linda Argote and Paul Ingram. 2000. Knowledge transfer: A basis for competitive advantage in firms. *Organizational behavior and human decision processes* 82, 1 (2000), 150–169.

[4] Jose Maria Barrero, Nicholas Bloom, and Steven J Davis. 2021. *Why working from home will stick.* Technical Report. National Bureau of Economic Research.

[5] Talya N Bauer, Todd Bodner, Berrin Erdogan, Donald M Truxillo, and Jennifer S Tucker. 2007. Newcomer adjustment during organizational socialization: a meta-analytic review of antecedents, outcomes, and methods. *Journal of applied psychology* 92, 3 (2007), 707.

[6] Iavor Bojinov, Prithwiraj Choudhury, and Jacqueline N. Lane. 2021. Virtual Watercoolers: A Field Experiment on Virtual Synchronous Interactions and Performance of Organizational Newcomers. (May 2021).

[7] Iavor Bojinov, Prithwiraj Choudhury, and Jacqueline N Lane. 2021. Virtual watercoolers: A field experiment on virtual synchronous interactions and performance of organizational newcomers. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* 21-125 (2021).

[8] Ricardo Britto, Daniela S Cruzes, Darja Smite, and Aivars Sablis. 2018. Onboarding software developers and teams in three globally distributed legacy projects: A multi-case study. *J. Softw. (Malden)* 30, 4 (April 2018), e1921.

[9] Melanie S Brucks and Jonathan Levav. 2022. Virtual communication curbs creative idea generation. *Nature* 605, 7908 (2022), 108–112.

[10] Ronald S Burt and Jennifer Merluzzi. 2016. Network Oscillation. *AMD* 2, 4 (Dec. 2016), 368–391.

[11] Jenna Butler and Sonia Jaffe. 2021. Challenges and gratitude: A diary study of software engineers working from home during covid-19 pandemic. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. IEEE, 362–363.

[12] Daniel M Cable, Francesca Gino, and Bradley R Staats. 2013. Breaking them in or eliciting their best? Reframing socialization around newcomers' authentic self-expression. *Adm. Sci. Q.* 58, 1 (March 2013), 1–36.

[13] Daniel M Cable, Francesca Gino, and Bradley R Staats. 2013. Breaking them in or eliciting their best? Reframing socialization around newcomers' authentic self-expression. *Administrative science quarterly* 58, 1 (2013), 1–36.

[14] Daniel M Cable and Timothy A Judge. 1996. Person–organization fit, job choice decisions, and organizational entry. *Organizational behavior and human decision processes* 67, 3 (1996), 294–311.

[15] Cam Caldwell and Ray Peters. 2018. New employee onboarding–psychological contracts and ethical perspectives. *Journal of Management Development* (2018).

[16] Gerardo Canfora, Massimiliano Di Penta, Rocco Oliveto, and Sebastiano Panichella. 2012. Who is going to mentor newcomers in open source projects?. In *Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering* (Cary, North Carolina) *(FSE '12, Article 44)*. Association for Computing Machinery, New York, NY, USA, 1–11.

[17] Daniel Carmody, Martina Mazzarello, Paolo Santi, Trevor Harris, Sune Lehmann, Timur Abbiasov, Robin Dunbar, and Carlo Ratti. 2022. The effect of co-location of human communication networks. *arXiv preprint arXiv:2201.02230* (2022).

[18] Gilad Chen. 2005. Newcomer Adaptation in Teams: Multilevel Antecedents and Outcomes. *AMJ* 48, 1 (Feb. 2005), 101–116.

[19] Cecily D Cooper and Nancy B Kurland. 2002. Telecommuting, professional isolation, and employee development in public and private organizations. *J. Organ. Behav.* 23, 4 (June 2002), 511–532.

[20] Munmun De Choudhury, Winter A Mason, Jake M Hofman, and Duncan J Watts. 2010. Inferring relevant social networks from interpersonal communication. In *Proceedings of the 19th international conference on World wide web* (Raleigh, North Carolina, USA) *(WWW '10)*. Association for Computing Machinery, New York, NY, USA, 301–310.

[21] Jana Diesner and Kathleen M Carley. 2005. Exploration of communication networks from the Enron email corpus. In *SIAM International Conference on Data Mining: Workshop on Link Analysis, Counterterrorism and Security, Newport Beach, CA.* Citeseer, 3–14.

[22] James Dominic, Charles Ritter, and Paige Rodeghero. 2020. Onboarding Bot for Newcomers to Software Engineering. In *Proceedings of the International Conference on Software and System Processes* (Seoul, Republic of Korea) *(ICSSP '20)*. Association for Computing Machinery, New York, NY, USA, 91–94.

[23] James Dominic, Charles Ritter, and Paige Rodeghero. 2020. Onboarding bot for newcomers to software engineering. In *Proceedings of the International Conference on Software and System Processes*. 91–94.

[24] David Easley and Jon Kleinberg. 2010. *Networks, crowds, and markets: Reasoning about a highly connected world.* Cambridge university press.

[25] Allison M Ellis, Talya N Bauer, and Berrin Erdogan. 2014. New Employee Organizational Socialization. *Handbook of socialization: Theory and research* 301 (2014).

[26] Fabian Fagerholm, Alejandro S Guinea, Jürgen Münch, and Jay Borenstein. 2014. The role of mentoring and project characteristics for onboarding in open source software projects. In *Proceedings of the 8th ACM/IEEE international symposium on empirical software engineering and measurement.* 1–10.

[27] Fabian Fagerholm, Alejandro S Guinea, Jürgen Münch, and Jay Borenstein. 2014. The role of mentoring and project characteristics for onboarding in open source software projects. In *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement* (Torino, Italy) *(ESEM '14, Article 55)*. Association for Computing Machinery, New York, NY, USA, 1–10.

[28] Timothy D Golden, John F Veiga, and Zeki Simsek. 2006. Telecommuting's differential impact on work-family conflict: is there no place like home? *J. Appl. Psychol.* 91, 6 (Nov. 2006), 1340–1350.

[29] M S Granovetter. 1973. The strength of weak ties. *Am. J. Sociol.* (1973).

[30] Douglas T Hall. 1996. Careers of the 21st century. *The Academy of Management Executive* 10, 4 (1996), 8–16.

[31] Helen Ai He and Elaine M Huang. 2014. A qualitative study of workplace intercultural communication tensions in dyadic face-to-face and computer-mediated interactions. In *Proceedings of the 2014 conference on Designing interactive systems.* 415–424.

[32] Pamela Hinds, Sara B Kiesler, and Sara Kiesler. 2002. *Distributed work.* MIT press.

[33] Felichism W Kabo, Natalie Cotton-Nessler, Yongha Hwang, Margaret C Levenstein, and Jason Owen-Smith. 2014. Proximity effects on the dynamics and outcomes of scientific collaborations. *Research Policy* 43, 9 (2014), 1469–1485.

[34] Ralph Katz et al. 1997. *The human side of managing technological innovation: A collection of readings.* Oxford University Press, USA.

[35] Benjamin Koehne, Patrick C Shih, and Judith S Olson. 2012. Remote and alone: coping with being the remote member on the team. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) *(CSCW '12)*. Association for Computing Machinery, New York, NY, USA, 1257–1266.

[36] Gueorgi Kossinets and Duncan J Watts. 2009. Origins of homophily in an evolving social network. *American journal of sociology* 115, 2 (2009), 405–450.

[37] Adriaan Labuschagne and Reid Holmes. 2015. Do onboarding programs work?. In *2015 IEEE/ACM 12th Working Conference on Mining Software Repositories*. IEEE, 381–385.

[38] Paul M Leonardi, Jeffrey W Treem, and Michele H Jackson. 2010. The Connectivity Paradox: Using Technology to Both Decrease and Increase Perceptions of Distance in Distributed Work Arrangements. *J. Appl. Commun. Res.* 38, 1 (Feb. 2010), 85–105.

[39] Jesse S Michel, Michael A Rotch, Jack E Carson, Nathan A Bowling, and Nicole V Shifrin. 2021. Flattening the latent growth curve? Explaining within-person changes in employee well-being during the COVID-19 pandemic. *Occupational Health Science* 5, 3 (2021), 247–275.

[40] Bonnie A Nardi and Steve Whittaker. 2002. The place of face-to-face communication in distributed work. *Distributed work* 83, 112 (2002), 10–7551.

[41] Kaylee L Pike. 2014. New employee onboarding programs and person-organization fit: An examination of socialization tactics. (2014).

[42] Ray Reagans and Bill McEvily. 2003. Network Structure and Knowledge Transfer: The Effects of Cohesion and Range. *Adm. Sci. Q.* 48, 2 (June 2003), 240–267.

[43] Paige Rodeghero, Thomas Zimmermann, Brian Houck, and Denae Ford. 2021. Please Turn Your Cameras on: Remote Onboarding of Software Developers During a Pandemic. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. ieeexplore.ieee.org, 41–50.

[44] D M Romero, B Uzzi, and J Kleinberg. 2016. Social networks under stress. *Proceedings of the 25th International* (2016).

[45] John Edward Ruchinskas. 1983. Communicating in Organizations: The Influence of Context, Job, Task, and Channel. (1983).

[46] Alice Snell. 2006. Researching onboarding best practice: Using research to connect onboarding processes with employee satisfaction. *Strategic HR Review* 5, 6 (Jan. 2006), 32–35.

[47] Igor Steinmacher, Tayana Conte, Marco Aurélio Gerosa, and David Redmiles. 2015. Social barriers faced by newcomers placing their first contribution in open source software projects. In *Proceedings of the 18th ACM conference on Computer supported cooperative work & social computing.* 1379–1392.

[48] Igor Steinmacher, Christoph Treude, and Marco Aurelio Gerosa. 2019. Let Me In: Guidelines for the Successful Onboarding of Newcomers to Open Source Projects. *IEEE Softw.* 36, 4 (July 2019), 41–49.

[49] Jaime Teevan, B Hecht, and S Jaffe. 2020. *The new future of work.* Technical Report. Microsoft internal report.

[50] J Teevan, B Hecht, and S Jaffe. 2020. The new future of work research from Microsoft into the pandemic's impact on work practices. https://www.londonfirst.co.uk/sites/default/files/documents/2021-05/NewFutureOfWorkReport.pdf. Accessed: 2022-10-4.

[51] Thea Turner, Pernilla Qvarfordt, Jacob T Biehl, Gene Golovchinsky, and Maribeth Back. 2010. Exploring the workplace communication ecology. In *Proceedings of the SIGCHI conference on human factors in computing systems.* 841–850.

[52] Johan Ugander, Lars Backstrom, Cameron Marlow, and Jon Kleinberg. 2012. Structural diversity in social contagion. *Proc. Natl. Acad. Sci. U. S. A.* 109, 16 (April 2012), 5962–5966.

[53] Ashwani Kumar Upadhyay and Komal Khandelwal. 2018. Applying artificial intelligence: implications for recruitment. *Strategic HR Review* (2018).

[54] Brian Uzzi and Jarrett Spiro. 2005. Collaboration and creativity: The small world problem. *American journal of sociology* 111, 2 (2005), 447–504.

[55] Andrew H Van de Ven, Andre L Delbecq, and Richard Koenig Jr. 1976. Determinants of coordination modes within organizations. *American sociological review* (1976), 322–338.

[56] Y Jasmine Wu, Brennan Antone, Arshya Srinivas, Leslie DeChurch, and Noshir Contractor. 2021. Teamwork in the time of COVID-19: Creating, dissolving, and reactivating network ties in response to a crisis. *J. Appl. Psychol.* 106, 10 (Oct. 2021), 1483–1492.

[57] Chi-Lan Yang, Naomi Yamashita, Hideaki Kuzuoka, Hao-Chuan Wang, and Eureka Foong. 2022. Distance Matters to Weak Ties: Exploring How Workers Perceive Their Strongly- and Weakly-Connected Collaborators in Remote Workplaces. *Proc. ACM Hum.-Comput. Interact.* 6, GROUP (Jan. 2022), 1–26.

[58] Longqi Yang, David Holtz, Sonia Jaffe, Siddharth Suri, Shilpi Sinha, Jeffrey Weston, Connor Joyce, Neha Shah, Kevin Sherman, Brent Hecht, and Jaime Teevan. 2021. The effects of remote work on collaboration among information workers. *Nat Hum Behav* (Sept. 2021).