Analytics to Make Hybrid Work, Work

by

Andrew J. Tindall
B.S. Electrical Engineering, United States Military Academy (2012)

Submitted to the MIT Sloan School of Management
MIT Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of
Master of Business Administration
Master of Science in Electrical Engineering and Computer Science
in conjunction with the Leaders for Global Operations program
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
May 2022

Author ....
MIT Sloan School of Management
MIT Department of Electrical Engineering and Computer Science
May 6, 2022

Certified by ....
Roy E. Welsch
Professor of Statistics and Data Science
Thesis Supervisor

Certified by ....
Duane S. Boning
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by ....
Maura Herson
Assistant Dean, MBA Program
MIT Sloan School of Management

Accepted by ....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Analytics to Make Hybrid Work, Work

by

Andrew J. Tindall

Submitted to the MIT Sloan School of Management
MIT Department of Electrical Engineering and Computer Science
on May 6, 2022, in partial fulfillment of the
requirements for the degree of
Master of Business Administration
Master of Science in Electrical Engineering and Computer Science
in conjunction with the Leaders for Global Operations program

Abstract

Hybrid work is a coordination problem at heart—how frequently and on which days of the week should hybrid employees come into the office? The COVID-19 pandemic accelerated a remote work revolution and caused the hybrid model—where employees split time between in-office and remote work—to become the norm as employees return to the office in 2022 and beyond. The shift to fully remote work during the pandemic highlighted numerous remote work benefits. The challenge is that remote collaboration is more difficult and time consuming to orchestrate—potentially decreasing innovation.

Acknowledging that remote and in-person work have different, and at many times complementary goals, our study tests whether employee collaboration data can help organizations solve the coordination problem inherent in hybrid work. We find that collaboration data can align work groups to maximize in-person collaboration gains while minimizing the number of days in office per week. We use data to recommend the optimal in-office frequency and find that offices will be 60% under capacity when employees return. Most importantly, we think about offices as networks—the value of being in the office scales non-linearly as users increase. We find that organizations can use collaboration data to model employee networks and appropriately align work communities. Ultimately, we develop a scheduling system that will help stabilize office space demand in 2022 and beyond.

Roy E. Welsch
Title: Professor of Statistics and Data Science
Thesis Supervisor

Duane S. Boning
Title: Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Acknowledgments

This study was truly a team effort—I want to say thank you to the Verizon leaders, Verizon's Advanced Analytics team, and my academic advisors.

A special thanks to my VTeam colleagues—Matt, Tim, Jenn, Prarthna, Tatev, Abdul, and Steve, just to name a few—this study would not have happened without your support! It was a pleasure to work with such a professional team that takes pride in its work (and loves data!). The collaborative culture I experienced during this study is why I am joining Verizon post graduation—excited to keep working together in the near future.

To my advisors, Roy and Duane—it was great to work with such accomplished academics who provide valuable insights at every turn! Thank you for your constant curiosity in our weekly touch points; I always came away with a new insight or a new method to experiment with next.

Finally, to my family—Beth Ann and Rose—thank you for all of your support over the past two years!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Verizon Communications was created on June 30, 2000 by Bell Atlantic Corp. and GTE Corp., in one of the largest mergers in U.S. business history. GTE and Bell Atlantic evolved and grew through decades of mergers, acquisitions and divestitures. Today, Verizon is a global communications technology company delivering the promise of the digital world to millions of customers every day. Verizon's primary product is connectivity—in early 2009 it became the largest wireless service provider in the U.S., as measured by the total number of customers. Verizon is currently focused on investment in and expansion of 5G ultra wide band services. Based on this history of merger and acquisitions, Verizon maintains a large real estate network—the Global Real Estate division maintains a 110 million square foot portfolio with an annual operating budget exceeding $1.7B. In 2019, Verizon initiated a hub strategy focused on consolidation to regional hubs aligned with major metropolitan centers [34].

Verizon shifted to nearly 100% remote work for all corporate functions in early 2020 due to the Global Corona Virus Pandemic (COVID-19). The COVID-19 disruption accelerated trends towards employee flexibility in our era of ubiquitous connectivity. Following this forced remote work experiment, Verizon believes that a hybrid work environment is the workplace of the future.

In early 2021, Verizon launched its WorkForward initiative that designated over 90% of employees as hybrid. A hybrid employee operates with a blend of in-person work days at the work site, and fully remote days working from home. This study is focused on how to leverage data to optimize the hybrid work environment for employees and the company—we anticipate our findings will influence real estate strategy. Our team establishes that remote work and in-person work have differing, but complementary benefits. With this basis set, we present methods for using data to define a hybrid work operational model.

This study has three primary aims and two secondary aims. The primary aims are to:

1. Develop a data backed methodology to recommend how many days per week hybrid employees should come into the office—a "curve" of hybrid work.

2. Design, implement and assess a model that aligns schedules under the current system constraints; assess increase in stakeholder alignment.

3. Design a new hybrid scheduling process that is not constrained by the current system; architect an improved system at scale.

The secondary aims are to:

1. Use Monte Carlo simulations to model campus constraints—design an approach that can be updated using behavioral data in 2022.

2. Recommend experimental structure for testing the improved scheduling system.

While it is important to highlight that this study is conducted at Verizon—a large telecommunications provider with an expansive commercial office network—our goal is to demonstrate that the study's findings generalize across industries. Every organization that provides value via collaboration of knowledge workers has to navigate the new hybrid environment. We provide an approach that any corporation can use to provide structure in the hybrid environment.

1.1 Problem Statement

Following an unprecedented shift to remote work in 2019, countless organizations seek to discover and operationalize the optimal hybrid work strategy. The hybrid work model has become popular because when managed correctly, it allows for a blend of in-office work—which is shown to increase organizational innovation—and remote work—which is shown to increase individual worker productivity. The shift to hybrid work underscored excess capacity in commercial office space portfolios. At the start of this project, the overarching request from leadership was to understand how the shift to hybrid work impacts consolidation strategies and organizational effectiveness—we quickly identified that in order to answer these questions, we first have to understand employee behavior and informal networks. Therefore, we have framed the underlying problem into four distinct components—how can we use data to recommend in-office frequency, model employee networks, align in-person collaboration and forecast the impact of hybrid work on real estate portfolio strategy?

First, organizations need to determine the recommended frequency of hybrid worker in-office time (e.g., 3 day/wk, 1 day/wk, 2 day/mo). The current schedule recommendation practice is a one-size fits all approach based on heuristics and assumptions around job function. How can we use data to provide an in-office frequency recommendation?

Second, leadership needs tools to model employee work groups and properly align in-office collaboration (e.g., a certain team is 1 day/wk, which day would be most collaborative?). In a matrixed, knowledge work environment, the problem of optimizing for employee collaboration can't be managed by senior leaders. How can we model employee networks, gain insights about how hybrid work changes collaboration and determine optimal in-person work schedules?

Finally, real estate strategy and operations need to visualize how hybrid employee behavior affects portfolio consolidation, capacity utilization and space modifications decisions (e.g., should below-capacity sites be shuttered on Fridays; should low-collaboration sites be divested). How can we use data to simulate hybrid work and model campus constraints? Can we use networks to define a real estate metric that provides an objective measure of collaboration at a work campus? Does this metric become a part of the equation for valuing a real estate campus?

1.2 Thesis Organization

When we initially framed this problem, it appeared to build sequentially to the real estate strategy simulations discussed above. As with many real world problems, our solution did not evolve sequentially—ironically, we started by building campus level capacity simulations framed as question four in our problem statement. We quickly identified that capacity concerns need to be addressed up front. As such, the thesis is not organized chronologically by how we solved the problem. Rather, the sequence in which we present our analysis is a recommendation itself. We believe this thesis is a road map to help large corporations implement a structured hybrid work strategy.

We first survey relevant research surrounding remote and in-person work. In **Chapter 2 - Literature Review**, we show that there are different, and many times complementary, benefits to remote versus in-person work. Research illustrates that collaboration is more challenging to orchestrate remotely and maintaining cohesive teams comes at a cost. This drives us to focus on employee collaboration networks. In **Chapter 3 - Current State and Problem Solving Approach**, we present the current scheduling process, stakeholder landscape and data available to solve this problem. We go on to outline our approach which rests on using employee metadata to understand work patterns and model networks.

Starting in **Chapter 4 - A Better Way to Estimate Hybrid Frequency**, we begin to highlight foundational hybrid work decisions that need to be based in data. We present a method to estimate hybrid frequency using historical collaboration data. Then, we design a Monte Carlo approach to simulate a work campus in the hybrid environment and understand system constraints. An organization must understand if capacity is an issue before architecting a complete hybrid work strategy.

Next, in **Chapter 5 - Modeling Employee Networks**, we make the case for using network theory to help align collaboration between knowledge workers. We design a process to model employee collaboration using networks and conduct a proof of concept with data from one business unit. We generalize the power of networks and provide a variety of metrics which are vital to assess the impact of hybrid work.

In **Chapter 6 - Aligning Schedules Under Current System Constraints**, we build on the proof of concept and optimize schedule alignment at a specific work campus.

We remove the current state constraints in **Chapter 7 - A Better Way to Schedule**, and develop a scalable approach to scheduling that takes into account relationships and office capacity. We use unsupervised machine learning to cluster communities and draw on our proof of concept to test and evaluate clustering accuracy. In **Chapter 8 - Testing the New System and Future Work**, we recap the outcomes of the study and how Verizon can test improvements at a larger scale. We also discuss how the study findings generalize beyond Verizon. Finally, this chapter discusses future extensions of this study.

Chapter 2

Literature Review

We began by surveying existing research on hybrid work. We find three prevailing themes—the employer perspective on hybrid work, the influence of proximity on collaboration, and the employee perspective on hybrid work. Ultimately, we interweave these themes to think about how the optimal hybrid work design balances the interests of the employer and employees.

2.1 Remote Work - What's Best for the Company?

The remote work revolution began gaining momentum in the mid 2000s and continued in pockets until the large shift to remote work occurred in early 2020. This section proceeds chronologically and details early experiments in remote work along with analysis of remote work during the COVID-19 pandemic. We present a variety of sources that highlight the associated benefits and costs of having an entirely remote workforce.

Cisco conducted an important large-scale modern era business experiment with remote work practices in 2009. This pilot program observed almost 2,000 Cisco employees and evaluated impact on productivity, quality of life and job satisfaction. The key finding was remote work practices could save $277M per year in costs for Cisco and employees [8, 6]. Beyond cost savings, Cisco used surveys to estimate that 69% of employees saw increased productivity and 75% of employees thought work time-lines improved. Overall, the pilot program recommended moving to a model where employees work remotely two days per week—though details on how this conclusion was reached were not provided [6].

Beyond business oriented pilot programs measured with surveys, academics have been actively researching remote work for over two decades. Many of these studies hone in on specific types of workers whose productivity can be measured against a quota. For example, Bloom et al. of the Stanford Graduate School of Business studied a large Chinese travel agency as it moved call operators to fully remote work [4]. This study estimates a 13% increase in productivity during the trial program. While this study is commonly cited to support a remote work shift, it lacks application outside the domain of an hourly, quota driven employee. Ultimately, the work group studied did not provide value through collaboration or innovation. Furthermore, after nine months 50% of the employees requested to return to the office, despite an average commute time of 40 minutes.

On the opposite coast, Harvard Business School Professor Tsedal Neely has done extensive research on how to effectively lead and manage a remote work force in the knowledge economy. Neely's key message is that leaders have to deliberately plan when teams need to meet and collaborate. Her research shows that properly planned and resourced "launch meetings" are the key to capitalizing on the deep focus gained from working remotely. In short, teams still need to collaborate to understand the leader's vision and build relationships with key stakeholders—but this doesn't mean employees need to be in the office everyday [21]. This research helps our team identify that hybrid schedules require an intentionality; companies can't simply assign random schedules focused on optimizing office space capacity utilization across a work week.

Two key studies using COVID-19 remote work data were published about a year into the pandemic. The first is an academic study from the University of Chicago Becker Friedman Institute for Economics. Gibbs et al. provide a comprehensive analysis of knowledge workers in the information technology/analytics field at a large corporation—they describe the campus of this company as being similar to that of Apple or Amazon. The research team had comprehensive access to human resources data (e.g. goal tracking, performance reviews, and demographic metadata) and software that tracked employee work patterns. The key finding was that working hours rose by 18% and average output declined slightly. Ultimately, fully remote work yielded a productivity **decrease** of 8-19% across the company. This study argues that decreased productivity is driven by increased costs to orchestrate meaningful collaborations in the remote environment.

Gibbs et al. control for pandemic lockdowns and parenting challenges to conclude that **1.4 hours per week, per employee** is expended on excess coordination and check-ins associated with remote work—this is coined as the "Work From Home (WFH) Effect." Furthermore, they found that the WFH effect reduced focused work by 1.4 hours per week, per employee. We use these findings to estimate **net weekly cost of fully remote work at 1.4 hours per employee—or .28 hours of value lost per day of remote work**, in the rest of this thesis. We are confident that the collaborative nature of the company Gibbs et. al studied can translate across knowledge work specialities.

Finally, Microsoft conducted an independent study of its employees during the COVID-19 pandemic [29]. Microsoft's findings largely align with Gibbs et al.—on average during the pandemic, workers worked more and delivered the same productivity. Additional insights are provided on how Microsoft employees collaborated during pandemic remote work. Researchers empirically observed that "stronger ties seemed to endure while weak ties seemed to atrophy." For organizations that deliver value through innovation, this trend is concerning. In sociology, strength of a tie is summarized by interaction time, intensity, intimacy and reciprocity [29, 10]. Strong ties are associated with in-group interactions—work with your manager or direct peers—while weak ties promote interdisciplinary collaboration [21]. Two parties sharing a weak tie generally work in a similar "community" but not directly together or for the same supervisor.

These findings demonstrate that productivity is not the only concern: most corporate knowledge jobs deliver value through interdisciplinary collaboration and innovation. Even Nicholas Bloom who is an economist and staunch advocate of remote work options, admits concern surrounding innovation in the solely remote environment. In mid 2021 he stated, “I fear this collapse of in-office face time will lead to a slump in innovation. The new ideas we are losing today could show up as fewer new products in 2021 and beyond, lowering long-run growth” [32]. This highlights concern around making long term decisions based on short term data—many firms are moving to strictly remote work without truly considering the long term costs.

2.1.1 Does Employee Proximity Influence Collaboration?

Corporate leaders and employees are both questioning the value of being in the office. This section details two studies that analyze how space and employee proximity impact collaboration. The first study was conducted by Thomas Allen at the Massachusetts Institute of Technology [1] . In line with Prof. Bloom’s concern about innovation, Allen hypothesized that employee proximity influences interdisciplinary collaboration. Allen showed that physical proximity has an inverse exponential relationship with the probability of communication between colleagues. Visually, this is represented as “The Allen Curve” displayed in Figure 2-1. Allen et al. argue that simply designing a workplace to create the *potential* for collaboration delivers innovative ideas. This research provides the basis for many open floor office plans familiar today.

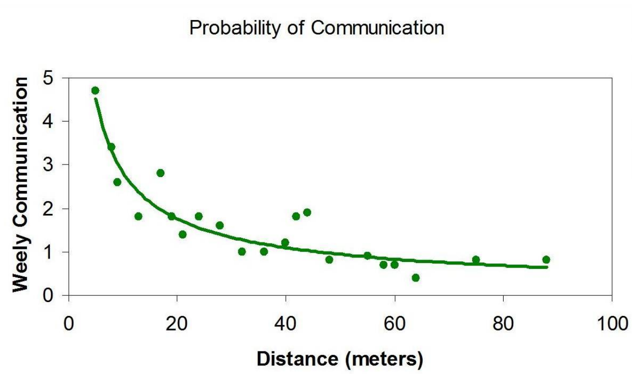


Figure 2-1: The Allen Curve (Adapted from [1])

A more recent study at the Massachusetts Institute of Technology validates Allen et al.’s findings. In 2017, Claudel et al. proved that "The Allen Curve" holds even in our modern era of connectedness [7]. This study uses collaboration data to build network models and shows that academic collaboration—measured by co-authorship—is driven by **spatial** versus hierarchical/institutional ties. Collaboration data is fit to an exponential decay model shown in Figure 2-2—chart A uses co-authored patent data and chart B uses co-authored papers. It is clear that spatial alignment of work communities is a factor in how knowledge workers collaborate.

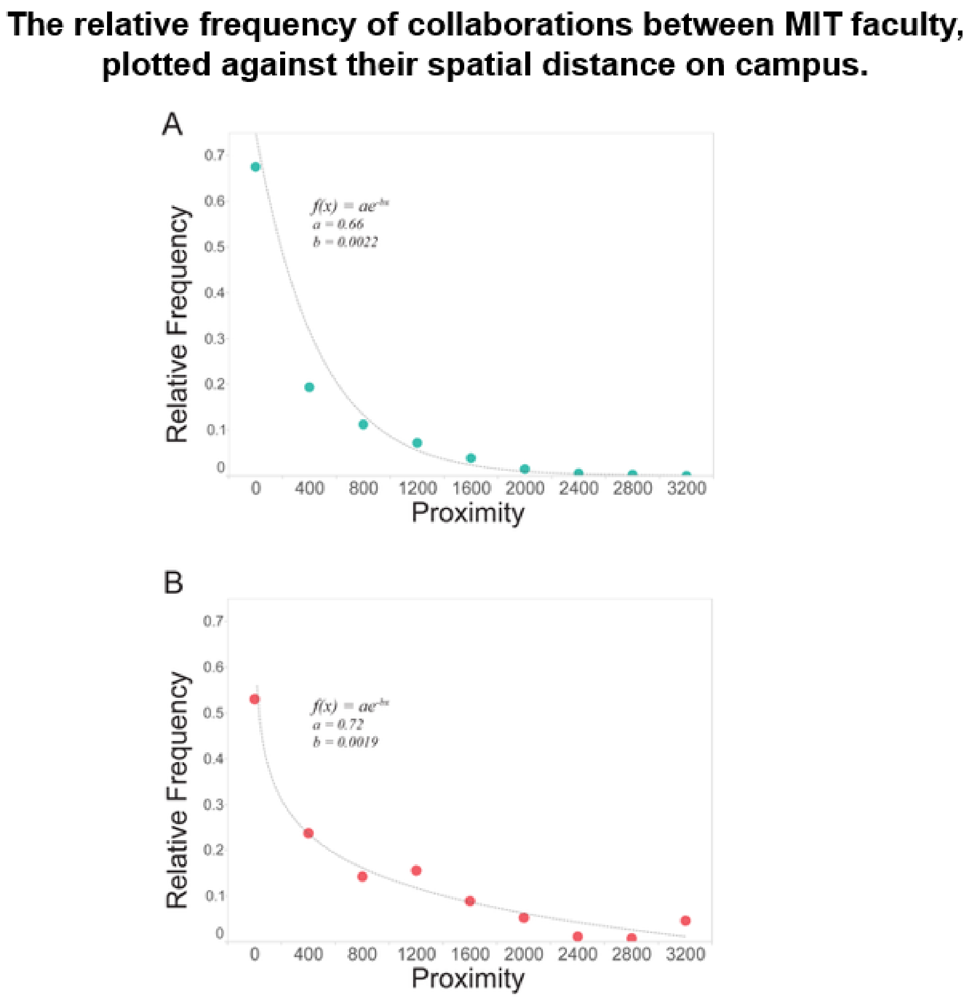


Figure 2-2: Validating the Allen Curve in the Modern Era (Adapted from [7])

From the employer perspective, our conclusions are threefold. First, knowledge work fundamentally delivers value via the exchange of ideas—a change from the industrial era measures of value. Second, research shows that employee proximity influences collaboration—strong ties may continue to flourish in a fully remote environment, but weak ties, which are valuable for innovation, atrophy. And finally, our overarching conclusion is that organizations need to build hybrid work systems that cultivate the right in-person interactions within an employee's community—this is how organizations will unlock the complementary benefits of remote and in-office work.

2.2 What's Best for the Employee - Flexibility?

Up until this section, the research surveyed has been framed from the perspective of the employer—we believe the employee perspective is also a part of the hybrid work equation. Overall, the pandemic changed employee priorities, and this must be taken into account as corporations design hybrid work models.

A number of studies show that flexibility is the "fastest-rising job priority in the U.S." [11, 2]. Adam Grant, an organizational psychologist at The Wharton School, has been a proponent of work-life balance research for years. He recently published an editorial piece that succinctly describes the generational shift our society is witnessing in the post COVID-19 workforce [11]. Grant, along with other scholars, argue that COVID-19 gave employees a taste of the freedom that comes with working from home and "the taste of freedom left [them] hungry for more." By early 2021, a number of large companies began insisting that employees return to the office full-time—this sparked what many are now calling "The Great Resignation." While this moniker may be caused by a variety of factors, employees do appear to be leaving jobs that are not flexible; hybrid work design is about more than just productivity, it is critical to retain top talent.

A separate 2021 study published in Harvard Business Review presents data that illustrates flexibility is critical [25]. Of 5,000 knowledge workers surveyed, 59% reported that "flexibility" is more important that compensation. Furthermore, 77% said they would prefer the ability to work from anywhere versus an amenity strewn corporate campus. This study also highlights an important point about bringing employees back—any mandate will naturally feel like a violation of flexibility and autonomy. In the study, 59% of employees said they would not work for a company that required five in-office days a week. Specific anecdotes include resignations from Apple following a strict in-office requirement of three days in-office per week. Apple employees responded with demands that hybrid work arrangements be an autonomous decision for each team.

Another interesting facet of Grant's argument is that this tension between employers and employees started more than a decade ago—COVID-19 just accelerated change that was already happening [11]. The new generation of workers, more commonly referred to as millennials, care more about autonomy and flexibility than net worth. The shift we are witnessing is forcing employers to rethink the standard industrial era assumption that work is conducted for 40 hours a week from 9 to 5 Monday through Friday. Grant highlights that successful companies should replace back to work mandates with systems that protect employee focus time and create intense interpersonal interactions at a regular interval. This thesis also sees this as the future of hybrid work.

From the employer vantage point, we see the next step as finding a balance—giving employees flexibility but providing enough structure to drive important, in-person interactions which stagnate with fully remote work. As presented in the Apple anecdote, a requirement for employees to come back for no particular reason is bound to fail—organizations must provide structure that connects networks and gives employees a reason to come to the office in-person.

2.3 Literature Review Conclusions

Bringing together these perspectives from relevant literature, we come to three conclusions . First, organizations need to find the right blend of remote and in-person work to unlock the benefits of the hybrid design and remain effective long term. Second, employee proximity affects collaboration—to maintain important weak ties in networks, many of which drive innovative ideas, employees need to have the *potential* for in-person interaction. And finally, organizations need systems that align collaboration and provide employees value when they come into the office—mandates alone are unlikely to work.

Chapter 3

Current State and Problem Solving Approach

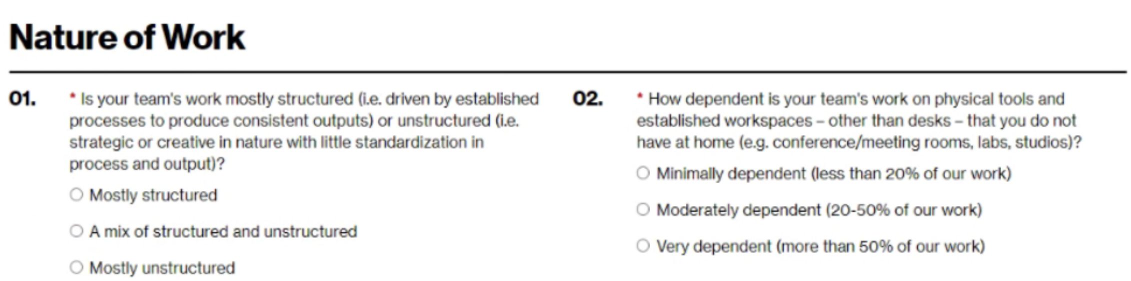
In this chapter, we detail the WorkForward program, the current scheduling process and the stakeholder landscape surrounding WorkForward. The current system is made up of sub-systems, or tools as described by the subject organization, that are designed to help business leaders manage hybrid work—we present each sub-system and discuss the flow of scheduling. WorkForward has a number of stakeholders spread across the human resources and real estate business units. The political considerations of the stakeholder landscape, along with critical feedback from the customers using these tools, shape our overall approach to the problem.

3.1 Current State of Hybrid Scheduling

By mid 2020, corporations started thinking about how the COVID-19 pandemic fundamentally changed the nature of the workplace. Verizon launched the WorkForward program to help employees navigate the new hybrid work environment. As a part of the program, 90% of employees were classified as hybrid—this means following an in-person/remote split set by leadership. WorkForward consists of three main sub-systems—the Human Resources (HR) Recommender Tool, the Human Resources (HR) Scheduler Tool, and Book-A-Space. This section will provide an overview of the current process. The flow of information through the system is displayed in Figrue 3-1—each tool will be discussed in depth.

Figure 3-1: Current State Scheduling Process

The HR Recommender Tool is the first step in the current process. Each director has the opportunity to use—but is not required to use—the HR Recommender Tool prior to submitting schedules for all subordinate employees. For organizational context, the typical director in the subject organization has anywhere from 100 to 200 employees with several associate directors and senior managers managing daily operations. The tool is a simple survey that asks the director questions about the nature of work across his/her team—directors are given three possible responses to each question. A sample of view of the tool is shown is Figure 3-2.



Nature of Work

01. * Is your team's work mostly structured (i.e. driven by established processes to produce consistent outputs) or unstructured (i.e. strategic or creative in nature with little standardization in process and output)?

- ○ Mostly structured
- ○ A mix of structured and unstructured
- ○ Mostly unstructured

02. * How dependent is your team's work on physical tools and established workspaces – other than desks – that you do not have at home (e.g. conference/meeting rooms, labs, studios)?

- ○ Minimally dependent (less than 20% of our work)
- ○ Moderately dependent (20-50% of our work)
- ○ Very dependent (more than 50% of our work)

Figure 3-2: HR Recommender Tool

In addition to several questions on the nature of work, the survey directly asks about frequency of collaboration required on the team. The responses to these questions are limited to A) rarely (monthly), B) intermittently (biweekly), and C) regularly (weekly). The survey responses are scored using a Likert scale algorithm that produces a frequency recommendation. There are three possible rhythm recommendations: weekly, monthly, and biweekly; and three possible day recommendations: one day, two days, and three days. This yields nine potential recommendations such as: weekly—one day every week, monthly—three days one week a month, or biweekly—two days every other week. It is important to highlight that the output from the HR Recommender Tool is a non-binding recommendation—directors can choose to follow or ignore the recommendation. Furthermore, the tool does not provide a recommendation around which day of the week or week of the month. This is left up to the director to determine manually.

The second tool in the process is the HR Scheduler Tool. After receiving a frequency recommendation, directors enter the HR Scheduler Tool and assign schedules to each subordinate employee. In this tool the directors are given an additional scheduling option—custom. The custom schedule option allows the director to assign in-office days without a recurring rhythm. This gives the current system four hybrid schedule models—weekly, biweekly, monthly and custom. The director can submit different schedules for each employee: however, given the challenge of doing so across a large business unit, we observed directors submitting one schedule for the entire business unit even if employees are geographically dispersed and serve different purposes.

Once the schedule is submitted in the HR Scheduler Tool, the schedule appears on each employee's profile in the company intranet. The goal is visibility and predictability; published schedules help colleagues coordinate in-person engagements. In the vein of predictability, HR stated that schedules will be set once a year and fixed for the entire calendar year. Finally, there is no capacity data behind the schedule submission or recommendation—i.e. all directors could submit the same schedule at a campus even if that scheduling scheme would put the campus over capacity on certain days of the week.

The third and final sub-system is the space reservation tool "Book-A-Space."

Book-A-Space is the application employees use to reserve desk and meeting room space. When employees enter the app, there is an option to book reservations using the employee's assigned schedule from the HR Scheduler Tool—this creates reservations for 30 days in line with the employee's schedule in the HR Scheduler Tool. For example, a biweekly Tuesday/Thursday employee would see four reservations populate for the next four weeks—two days, every other week. Our team refers to these as "scheduled" bookings.

Employees do not have to use this "book on-schedule" feature and are not restricted from booking outside of the assigned schedule. We refer to these as "flexible" or "flex" bookings—an employee needs to be in the office outside of his/her assigned schedule. Leadership states that an employee's schedule defines the minimum amount of time he/she is expected to be in the office; however, the frequency an employee schedules in-office time does not have an upper bound—i.e., a weekly employee can come in five days a week by flex booking a desk four days a week in addition to one scheduled day per week.

These three sub-systems outline the scheduling process from initial frequency recommendation through reservation booking. The details of each sub-system influenced how we framed the problem and conducted stakeholder interviews. Three points to remember from the current state are; 1) frequency is recommended based on director survey answers, 2) recommendations do not specify days of the week or align work groups, and 3) employees can book "scheduled" or "flexible" space reservations in Book-A-Space. We will return to these points when defining the problem solving approach. But first, it is helpful to expand on the stakeholder landscape surrounding the current state.

3.2 Stakeholder Interviews and Mapping

While mapping the current process our research team also conducted stakeholder interviews across the Human Resources and Real Estate business units. In this analysis we discovered an HR/Real Estate silo and heard directors describe the challenge of manually aligning employee schedules with the proper stakeholders.

At the top level, the design of our subject organization's hybrid work system involves two macro stakeholders—Human Resources (HR) and Real Estate (RE). Within these two silos, there are three categories of tactical stakeholders—tech tools, campus capacity, and portfolio strategy. These tactical stakeholders use data and drive decisions that straddle the divide between Human Resources and Real Estate. A streamlined model of the stakeholder landscape is visually displayed in Figure 3-3.

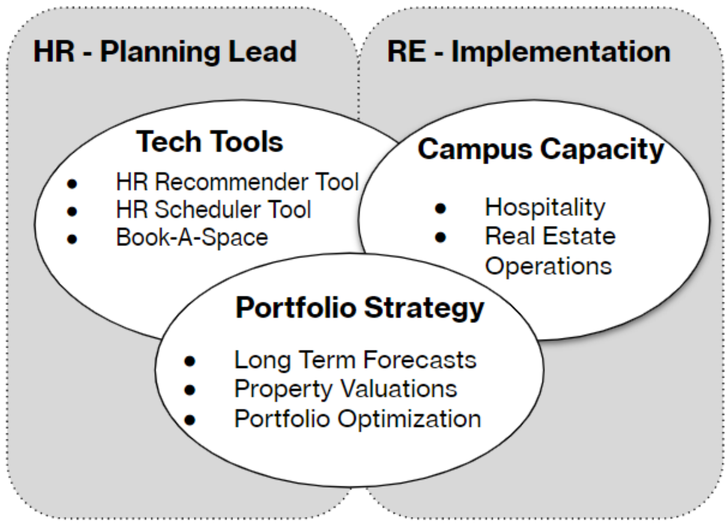


Figure 3-3: Hybrid Work Design Stakeholder Landscape

In theory, company leadership stated that HR owned planning of the hybrid work system and Real Estate would own implementation. Early in the study, we observed that this transition was challenging for the subject organization because of the divide between HR and Real Estate; HR did not want to give up control of the system and iterate on changes recommended by Real Estate.

As with many large scale initiatives, the tactical stakeholders focus on influencing specific outcomes. On the Real Estate side, campus capacity stakeholders want to ensure the scheduling system does not overwhelm the campus on a specific day of the week. In a similar vein, portfolio strategy stakeholders are advocating for a system that stabilizes office space demand quickly and allows elimination of excess capacity in the real estate network. On the HR side, the focus is on providing tools that help business leaders and employees remain productive in the hybrid environment. The ownership of technology tools is split between HR and Real Estate—this creates a challenge for data sharing and integration. HR owns the HR Recommender Tool and the HR Scheduler Tool, while Real Estate owns the Book-A-Space application.

Beyond the stakeholders influencing hybrid system design, our team also engaged with perhaps the most critical stakeholder, the customer—or in this case, the directors and employees. Our interviews with directors using the system helped focus the study. Director feedback was threefold. First, directors voiced concern about the frequency recommendations from the HR Recommender Tool. Second, directors felt the business unit level was too high for them to effectively assign in-office frequency. And finally, directors said it was overwhelming to align employee schedules with those of key stakeholders. Our team saw that these three areas also answered senior leadership questions around if the organization was implementing hybrid work in the most efficient and promising manner.

Identifying the silo between HR and Real Estate is critical to our strategy. We will develop the overall problem solving approach later in this chapter, but for now it is important to note how the organizational dynamics shape our focus. Since this was an HR designed system, we turn to the customer to build a case for improvements prior to gaining HR leadership buy-in. We specifically use a director inside the Real Estate organization as a proof of concept. By building a concrete proof of concept, our team is able to effectively identify where analytics can improve the current system and convince HR leadership to implement a large scale experiment testing analytical methods to set frequency and align schedules.

3.3 Data Available

We use three main data sources to solve this problem—employee Google calendar data, employee details data, and employee schedule data. This section describes the data structure and data fields for each source.

Employee Google calendar data is logged for each employee using the corporate network. These data are filtered to include all events pertinent to workplace activities—personal events are eliminated and only engagements between subject organization employees remain. These data are retrieved using an application programming interface (API) that passes an employee email address and returns a text file of the employee's corporate calendar. These data are merged to create a comma separated value file with calendar events from all employees of interest over a specified time horizon.

Each row in these data is one event from an employees calendar. For each calendar event there are a number of data fields—unique meeting identifier, meeting subject, start time, end time, location, by name list of invitees, and by name list of confirmed attendees. Since the API passes an email, these data do contain duplicate events—if we pass employee A email and employee B email and both attend the same event, this event will show up twice in the initial data. Throughout our analysis we use the unique meeting identifier to ensure that duplicate meetings are not present in the data. All calendar data in our study is structured and filtered in this manner.

The second source of data comes from an employee details database. The data fields include the employee's identification number, job title, email, campus assignment, business unit alignment, where he/she sits in the organizational hierarchy and a contractor/full-time designator. These data are primarily used to understand employee campus assignments and embed the organizational hierarchy into our modeling solutions.

Finally, we use employee schedule data held in an HR database. These data contain one row for every hybrid employee. The data fields include the employee identification number, campus assignment, leadership hierarchy, schedule type (Weekly, Monthly, BiWeekly, Custom), day of the week assigned, and week of the month assigned (1, 2, 3, 4). Since this is organized at the employee identification number level it can be joined with the employee details data to gather more descriptive information about an employee as needed.

Combining these three data sources delivers what we refer to as employee metadata—in plain terms, a comprehensive set of data to describe the how/when/who of hybrid employee interactions. We provide more detail about these data sources when describing our analytical methods in Chapters 4, 5, 6 and 7.

3.4 Hypotheses and Problem Solving Approach

We believe that analytics can solve four problems in the current system—recommending hybrid employee frequency, modeling campus capacity, determining if director level is correct for setting schedules, and aligning employee schedules with stakeholders. Overall, our hypothesis is that network models using employee metadata will improve hybrid work outcomes for the subject organization. Network models and employee metadata naturally quantify a "curve" of hybrid work that profiles how frequent a hybrid employee should physically come into the office. Furthermore, we believe that applying analytics and unsupervised machine learning on these networks will deliver a measurable increase in schedule alignment. Our goal is to test this hypothesis and design a scheduling system that is rooted in analytics. Ultimately, an analytical solution to these problems will enable proactive corporate real estate decisions.

Our approach starts by thinking about frequency and campus capacity—how often should hybrid employees come into the office and is the subject organization at risk of exceeding capacity at hub campuses? In Chapter 4, we analyze a variety of factors including meeting frequency, employee desk/meeting room booking behavior, and average number of days in office by hybrid employees. Our goal is to design an index methodology to recommend hybrid frequency. To understand capacity constraints, we explore using Monte Carlo simulations to model the hybrid environment at hub campuses. Our goal is to design a simulation framework that enables real estate strategy and operations leaders to make proactive portfolio decisions. Proactive decisions are a challenge because hybrid work introduces new variables and unpredictability when modeling office space capacity.

In Chapter 5, we experiment with network models to quantify employee collaboration. Organizational Network Analysis is a modeling technique that traditionally relies on surveys from employees to build networks. Our approach uses employee metadata to understand and visualize collaboration patterns; we believe this data is less biased than survey data and allows for implementation at a much larger scale than traditional methods. Furthermore, we experiment with exponential random graph models—similar to response test models with non-network data—to test if the director level is the right level to set employee schedules.

To make this concrete and test the hypothesis on a manageable sample, we design a business experiment within one business unit—a business intelligence unit made up of 200 geographically dispersed, hybrid employees. The goal of this business experiment is to test two hypotheses: 1) the current scheduling methodology is inefficient, and 2) network modeling provides a measurable increase in stakeholder alignment. We seek to develop a process that models director level networks at a specific campus. Furthermore, we examine how this director level model can be de-aggregrated to the individual employee level. Finally, we ideate on how network metrics can be used by organizations to understand the impact of hybrid work.

Following this descriptive network modeling in the proof of concept, in Chapter 6 we develop a method to improve scheduling in the business intelligence unit. We start by quantifying the baseline schedule alignment with stakeholders and move from descriptive analytics to prescriptive analytics by formulating an optimization to maximize stakeholder schedule alignment. The customer for this prescriptive portion is the business intelligence director and key stakeholders. Our goal is to demonstrate that we can optimize scheduling while subject to the constraint that schedules must be set at the director level——a level we hypothesize is not granular enough.

Following the proof of concept under the current system constraints, in Chapter 7 we present a model that aligns schedules by relying solely on employee metadata patterns. This model removes the current constraint requiring directors to set schedules. Instead, we use unsupervised machine learning to cluster employee communities and recommend schedules based on these cluster assignments. We quantify how these assignments improve scheduling alignment compared to the heuristically set schedules in the current state. The primary customers are company executives determining remote work polices and the goal is to design an analytically grounded employee scheduling pipeline.

Overall, our approach focuses on testing if employee networks can help organizations design schedules that maximize in-person collaboration in work communities. We believe that the sequence of our approach is an insight for organizations structuring hybrid work—understand frequency first, then employee networks. This will help deliver the optimal in-person/remote mix for hybrid employees and keep communities connected in the hybrid environment. Furthermore, we believe a system rooted in data will improve employee satisfaction and stabilize hybrid employee behavior. Quickly stabilizing how frequently hybrid employees come into the office will in turn enable proactive real estate consolidation strategies.

3.5 Future State - How will we know the problem is solved?

The desired future state is a scheduling pipeline that leverages analytics to set hybrid employee frequency, align hybrid work groups at campuses and balance campus capacity. Since the hybrid work system is in the late stages of development, our goal is not to completely overhaul the system in this study. Rather, we focus on building a proof of concept with the new scheduling pipeline—ideally this convinces the subject organization to launch a large scale experiment in 2022 when hybrid work is officially implemented.

The high level goal is to build a system that improves hybrid work outcomes and enables proactive real estate strategy. With this in mind, the problem can be framed sequentially: 1) can analytics help recommend hybrid employee frequency?, 2) who should set schedules?, 3) can networks align schedules/balance capacity?, 4) understand hybrid employee behavior, and finally, 5) determine optimal real estate strategy. This is framework is depicted visually in Figure 3-4. The first three questions are the focus of this study.

Our goal is to use analytics to improve the first three decisions because these affect

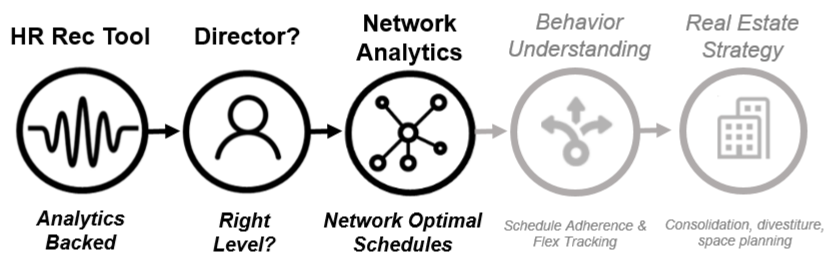


Figure 3-4: Hybrid Work Real Estate Strategy Model

real estate strategy long term. This study will be successful if we test and quantify how analytics can improve scheduling—the ultimate goal is to build a scheduling pipeline the subject organization adopts long term. The secondary aim is provide a plan for testing and implementing the improved system. This study is truly an experiment within an experiment—COVID19 has fundamentally changed the way we live and work—we hope to contribute to the larger experiment by showing how data can help organizations decipher the optimal balance between fully remote and fully in-person work.

Chapter 4

A Better Way to Estimate Hybrid Frequency

In this chapter, we describe an approach to recommending hybrid employee in-office frequency using employee metadata. Literature surrounding in-office frequency reveals that common methods rely on employee/manager perceptions versus objective data. Furthermore, research shows that the pandemic has resulted in over collaboration among knowledge workers—this is costing organizations time and money. We argue that hybrid frequency should be determined using objective collaboration data. This allows a hybrid employee to minimize time in-office while still achieving a similar quantity of in-person collaboration. In turn, this allows real estate strategy to simulate campus capacity and make proactive portfolio consolidation decisions. Hybrid frequency is the first of many hybrid work decisions; it is critical to ground this decision in data because it affects many portions of the system downstream.

4.1 Current Research on In-Office Frequency

Currently, organizations are not anchoring in-office frequency decisions to data. We believe that the frequency decision is critical because it has cascading impact on the hybrid system—in-office hospitality, space configuration and portfolio size are all dependent on how frequently hybrid employees come to the office. This section presents the current frequency research—mainly from well-respected consultancies—and we highlight the subjectivity of these decisions. Then, we present survey data around weekly in-office frequency to help organizations benchmark decisions.

BCG and McKinsey, global consultancies, have published numerous pieces on hybrid work models. Generally speaking, these articles provide recommendations by displaying the spectrum from fully on-site to primarily remote, and aligning why an organization should chose a certain model [33]. Figure 4-1 displays one such model recommender chart—as we discussed in Chapter 2, different work models optimize different objectives.

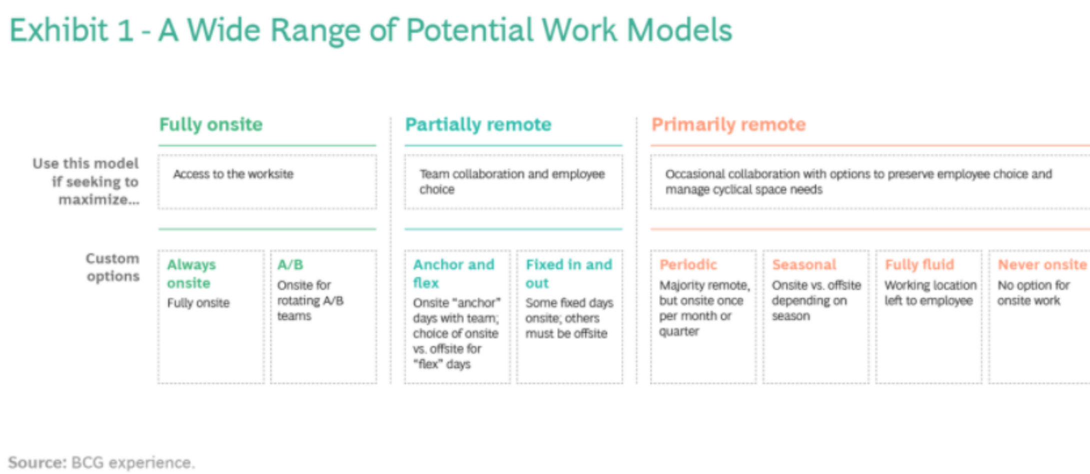


Figure 4-1: BCG Model Selection Criteria [33]

Additional recommendations have been made by researchers in academia. One such study from the Massachusetts Institute of Technology's Sloan School of Management establishes a four step framework where organizations heuristically estimate the number of days needed to optimize for key business metrics [28]. Similar to a Likert scale, this model is displayed below in Figure 4-2.

These types of subjective tools represent the prevailing perspective on how to select in-office frequency. In our subject organization, a Likert scale model similar to Figure 4-2 sits behind the survey based HR Recommender tool that provides a recommendation. Although these tools make a subjective process slightly more

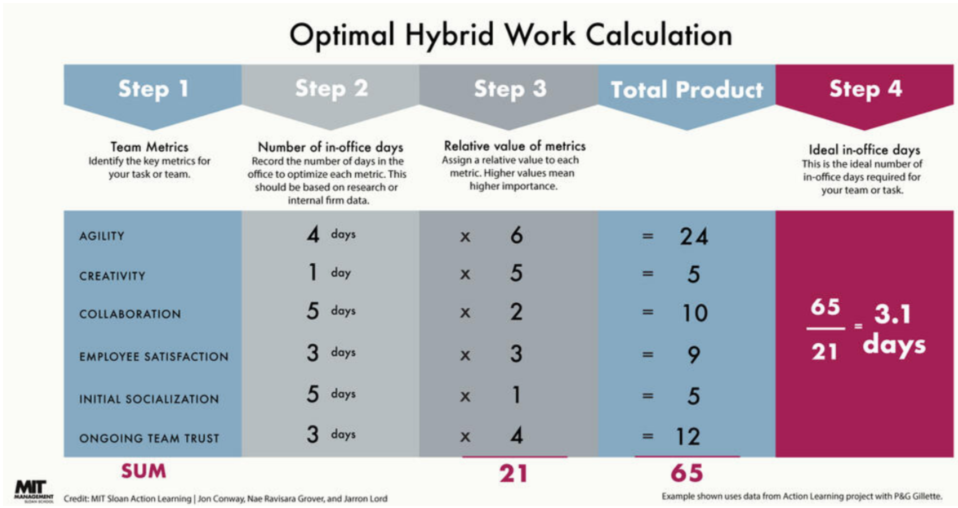


Figure 4-2: MIT Research Model [28]

objective, the process is still biased by human input. Our perspective is that these tools do not leverage enough data and need to be anchored to an unbiased data source.

We also consider data from a study that is compiling hybrid frequency data across a diverse sample of organizations. This benchmark analysis allows our team to assess where our subject organization resides on the frequency spectrum. A team from Stanford and the University of Chicago has been collecting survey data on employee in-office frequency throughout the pandemic [3]. The surveys have asked two main questions: 1) How many days per week does your employer plan for you to work from home once the pandemic concludes? 2) How many days per week would you like to work from home once the pandemic ends? These data are presented in Figure 4-3, which shows employer plans for number of days working from home for hybrid employees, and Figure 4-4, which shows the number of days per week hybrid employees desire to work from home. It is apparent that the longer the fully remote trend lasts, employers are increasing the number of days per week hybrid employees can work from home. In contrast, employee desire to work from home has leveled off at just over two days per week.

Using these data as a benchmark, our subject organization is well below the average in terms of in-office frequency—on balance, hybrid employees are scheduled to work from home four days per week versus the benchmark of one day per week work-

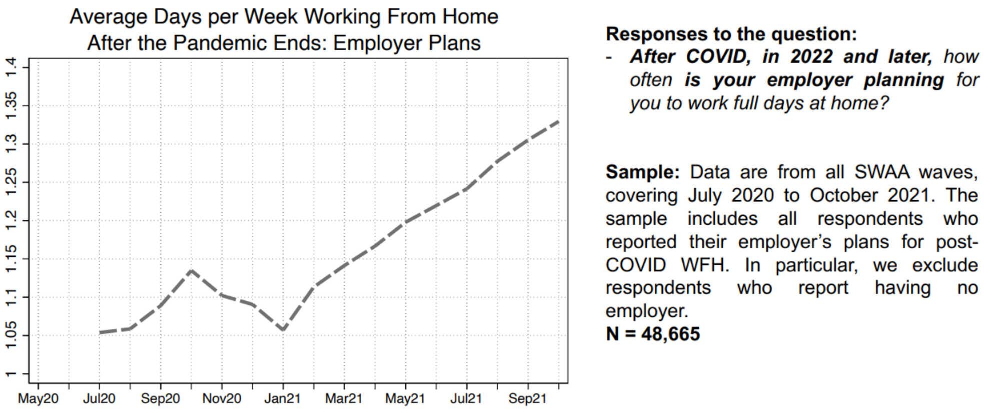


Figure 4-3: Employers Plans for In-Office Frequency [3]

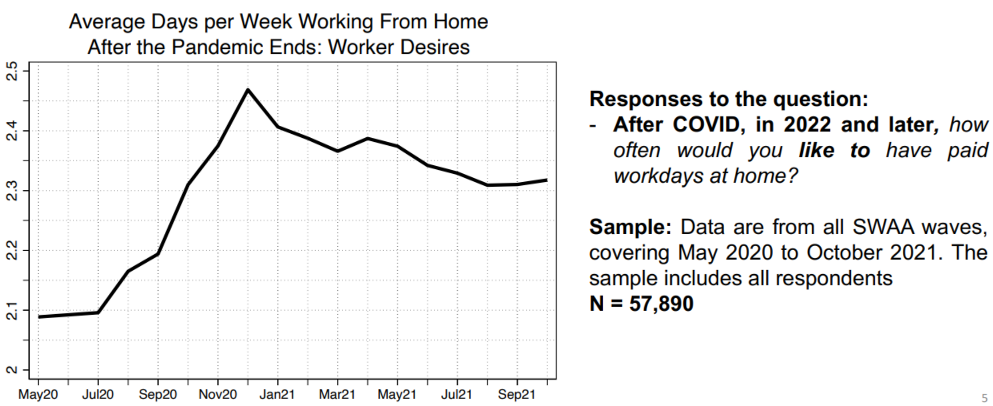


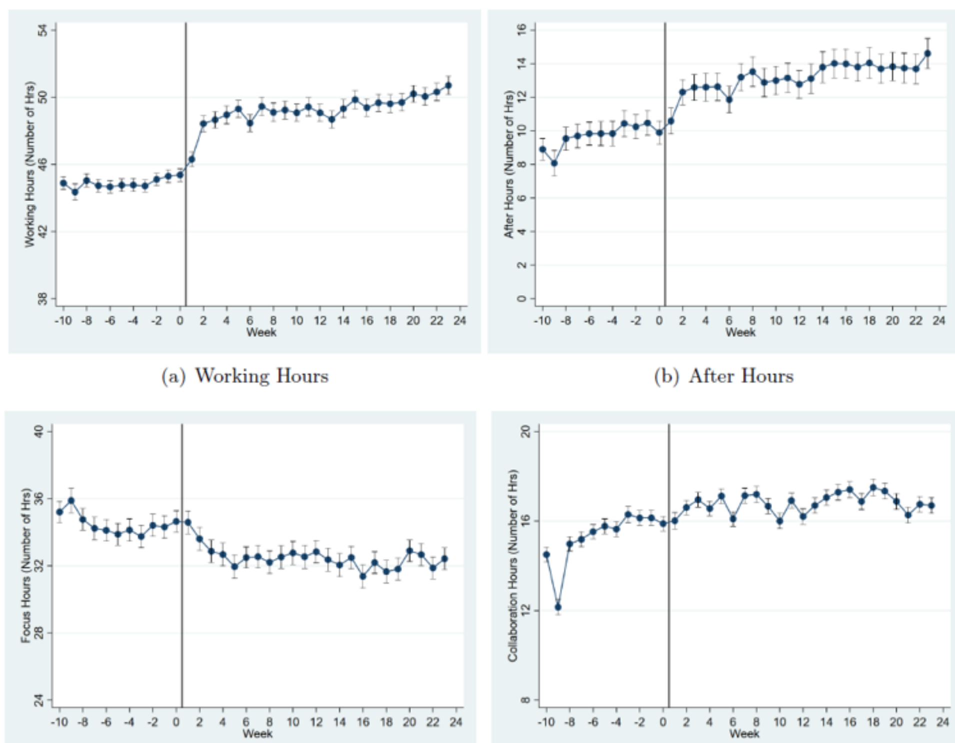
Figure 4-4: Hybrid Employee Work From Home Preferences [3]

ing from home. Being an outlier is not necessarily negative—if employees can gain the requisite amount of collaboration with one day in the office per week this actually could be a competitive advantage. The challenge is calibrating where this balance lies.

4.1.1 Pandemic Impact on Collaboration

To decide how much collaboration is enough, we need to acknowledge how COVID19 has changed collaboration. Overall, COVID19 has created a corporate culture of over collaboration—knowledge workers have an endless stream of Zoom, Microsoft Teams and Bluejeans calls. It is important to acknowledge this finding; organizations can gain a competitive advantage by designing hybrid tools that help employees prioritize collaboration on in-office days and heads down focus work on remote days.

The data tells the story of over collaboration in the post pandemic era—Gibbs et al. use Microsoft teams data to visualize how COVID19 impacted the IT firm in their study [9]. Figure 4-5 displays the data from Gibbs highlighting the increase in collaboration and decrease in focused work. In Figure 4-5, the vertical bar at week 0 represents when the pandemic lock downs first occurred in March of 2020. This study demonstrates that the initial spike in collaboration and decrease in focus hours was sustained throughout remote work. Furthermore, there is a significant upward trend in after hours work, as employees struggle to offset the increase in meetings and calls associated with fully remote work.



(a) Working Hours

(b) After Hours

(c) Focus Hours

(d) Collaboration Hours

Figure 4-5: Pandemic Impact on Collaboration [9]

We test the collaboration hypothesis in our subject organization. We use a sample of 600 employee Google calendars from June to September of 2019, pre COVID19, and from June to September of 2021, during fully remote work due to COVID19.

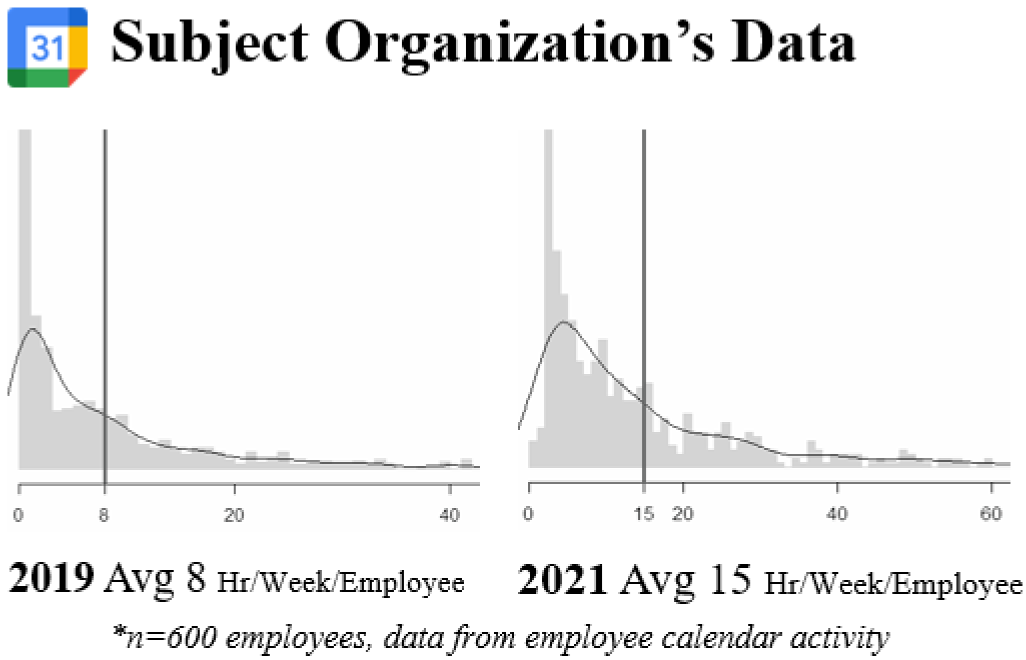


Figure 4-6: Testing Over Collaboration Hypothesis at the Subject Organization

After data cleaning and controlling for pandemic associated events—e.g., monthly meditation or virtual social gatherings—the sample displays an 87.5% increase in weekly time in meetings from 8 hours in 2019 to 15 hours in 2021.

This finding leads to three conclusions and guides the focus of this study. First, fully remote work has caused employees to spend more of the work week in meetings. Second, if we properly align employee networks we can return collaboration to pre-pandemic levels. Finally, by aligning schedules we may be able to minimize employee time in the office while maximizing collaboration. This would in turn achieve the balance we have discussed—employees achieve a sufficient amount of in-person collaboration but the individual focus time gained during fully remote work is protected.

Based on these analyses, we hypothesize that organizations can use calendar data to solve for frequency and optimize collaboration. The next section discusses how organizations can calibrate in-office frequency using collaboration data. Then, in Chapter 5, 6 and 7, the core of our study addresses how organizations can use networks to foster the right collaboration and protect valuable focus hours.

4.2 Calendar Data as a Frequency Methodology

Different job functions require different levels of collaboration. This is a key principle that organizations should strive to operationalize with hybrid work—in-office frequency is naturally a function how much an employee needs to collaborate. Unfortunately, many organizations are taking a "one size fits all" approach to in-office frequency. In this portion of the study, we start by validating that the current scheduling system does not provide a correlation between need for collaboration and in-office frequency. This finding pushes us to design a new system that leverages collaboration data to recommend in-office frequency.

We analyze the same sample of 600 employees to test whether employees with more collaborative roles are assigned a higher in-office frequency. The data presented in Figures 4-7 and 4-8 displays how many hours each employee collaborates per week and the employee's assigned in-office frequency per month. The number of employees in the sample is reduced to 198 after removing contractors and employees that had not published schedules in the HR Scheduler system. These data confirm our hypothesis that the current system does not produce in-office frequency assignments correlated with the collaboration required in a role.

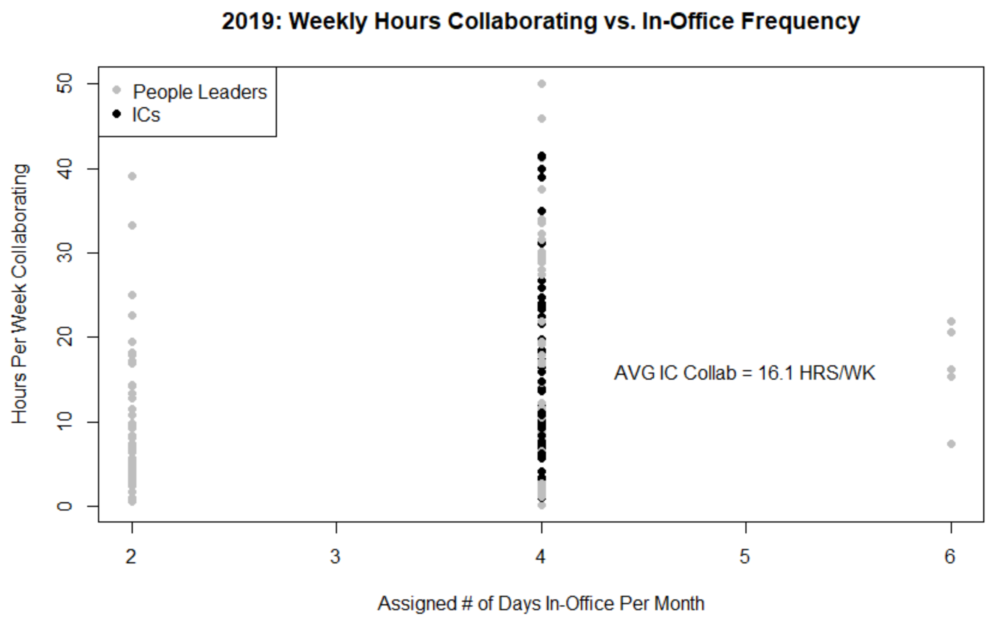


Figure 4-7: 2019 Hours Collaborating vs. Assigned In-Office Frequency

n=198 employees with in-office schedules published

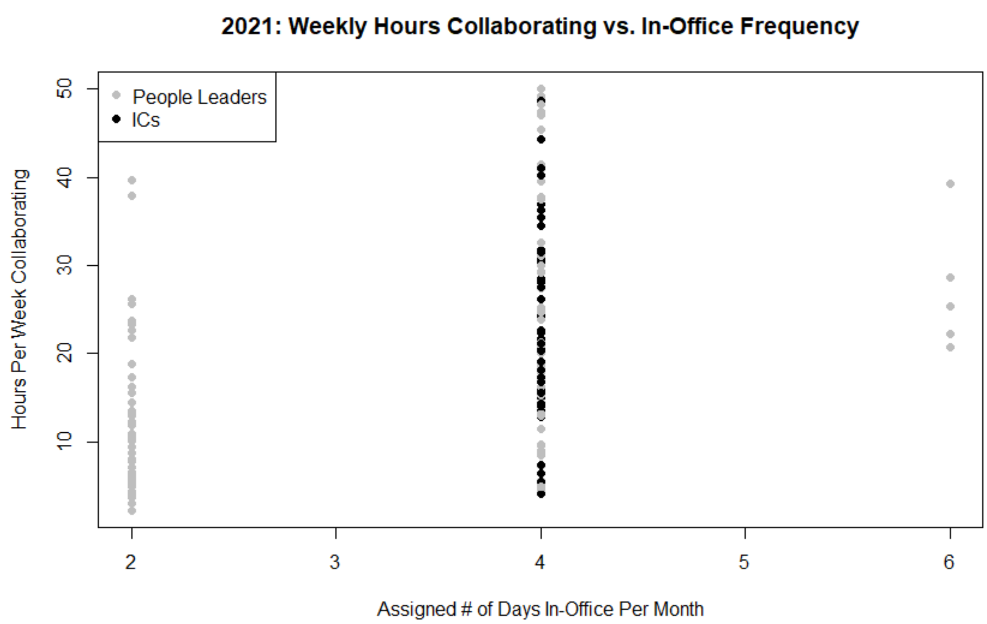


Figure 4-8: 2021 Hours Collaborating vs. Assigned In-Office Frequency

n=198 employees with in-office schedules published

Our solution is to generate frequency recommendations using collaboration data—this achieves the overarching goal of maximizing in-person collaboration while minimizing the number of days per week a hybrid employee comes into the office. We propose a simple "collaboration index." This index represents the potential in-person collaboration at the employee's assigned campus, relative to the total number of hours worked per week.

$$\textit{Employee Collaboration Index} = \frac{\textit{Potential In-Person Collaboration Hrs/Week}}{\textit{Average Total Work Hrs/Week}} \tag{4.1}$$

After calculating the collaboration index, organizations can map the index—which is derived from objective data versus subjective surveys—to in-office frequency recommendations. The sensitivity of this translation can be determined by business leaders—in the translation chart in Table 7.8 for our subject organization, we assume that an employee can realistically spend half of an eight hour day in the office completing formal collaboration. Therefore, if the employee has more than four hours of potential in-person collaboration per week, they should come in two days versus one day per week. This allows for time to move between meetings, engage in informal communication and arrange lunch/coffee engagements.

Collaboration Index (CI)	**Estimated Collaboration**	**Recommended Frequency**
$0.00 < CI < 0.025$	0 - 1 Hours/Week	1 Day/Month
$0.025 < CI < 0.10$	1 - 4 Hours/Week	1 Day/Week
$0.10 < CI < 0.30$	4 - 12 Hours/Week	2 Days/Week
$0.30 < CI < 1.0$	12 - 40 Hours/Week	3 Days/Week

Table 4.1: Collaboration Index Translation

By rooting in-office frequency recommendations in objective data, organizations can begin designing the optimal hybrid environment. Although this collaboration index heuristic may not achieve a global optimum, it is simple enough for leaders to implement and will move the organization towards a culture where employees come into the office commensurate with their collaboration requirements. This will enable organizations to maximize potential in-person collaboration while minimizing time in-office. Furthermore, a data backed index will improve employee satisfaction with the system—frequency recommendations will make sense to employees. And finally, organizations can architect in space for the serendipitous interactions that Allen showed are critical for innovation. For example, our subject organization only maps four hours of a typical work day to formal collaboration, leaving space for those critical informal engagements.

Based on this initial frequency versus collaboration mismatch, we analyze a larger data sample to validate this hypothesis and demonstrate how the collaboration index solution would improve the current state. The next section details these data, validates the mismatch at a global level and simulates in-office frequency using the collaboration index.

4.2.1 Aligning In-Office Frequency through Data

This section applies our frequency methodology on a larger data sample. We present a use case that confirms our hypothesis that frequency is not aligned with collaboration. Furthermore, we illustrate how an objective collaboration index can improve the overall frequency alignment.

The data sample for this use case is a year of calendar data (2021) for all employees under the chief operating officer in the subject organization—3,383 employees designated as hybrid. The chief operations officer's organization is comprised of all operations and legal functions. Since we want in-office frequency to be driven by potential in-person engagements at a campus, we filter these data to the headquarters campus. This produces a sample of 548 employees assigned to the headquarters campus. The number of data points is further pared down by the number of employees who have published schedules in the HR Scheduler system—only 346 had submitted schedules at the time of the study.

We analyze the 2021 calendar data for these 346 employees to estimate the potential "meaningful" in-person interactions per week for each employee. We accomplish this by parsing each calendar event, determining if more than two employees assigned to the headquarters were invited to attend and then calculating an in-person to remote ratio in that specific event.

The in-person to remote ratio allows us to objectively determine if a potential interaction was "meaningful"—if two employees could have attended in person, but it was a large 50 person meeting with 48 remote attendees, it does not seem critical that the two employees come into the office. However, if eight employees out of ten attending the meeting are all assigned at the headquarters, this is a collaborative engagement that may provide value in-person. The in-person to remote ratio objectively captures subjective judgments on the value of being in the office. We set the "meaningful" threshold equal to the median of the remote/in-person ratio. As a result, events with small in-person/remote ratio—below the median—are not counted as meaningful; events with an in-person/remote ratio above the median are deemed meaningful. Although there are flaws in this approach, we believe it achieves a balance—very large, mostly remote events are removed while large all-hands where a preponderance of the team is assigned at the headquarters campus remain in the data.

We apply this approach to our sample of 346 employees assigned to the headquarters campus and estimate that the average "meaningful" in-person collaborations per week in 2021 was **3.3 hours/week**. This discovery provides an answer to one of the subject organizations initial questions—are we getting in-office frequency right? Meaning, are employees coming in too few or too many times per week? In the chief operating officer's business unit, employees were scheduled to be in the office 0.87 days/week on average—just under one day per week. At a high level, this matches with the collaboration index we developed in the previous section—employees with one to four hours of in-person collaboration per week should come in one day per week.

A granular investigation tells a different story. We match each employee's calendar to the frequency submitted in the HR Scheduler system. These data are presented in Figure 4-9.

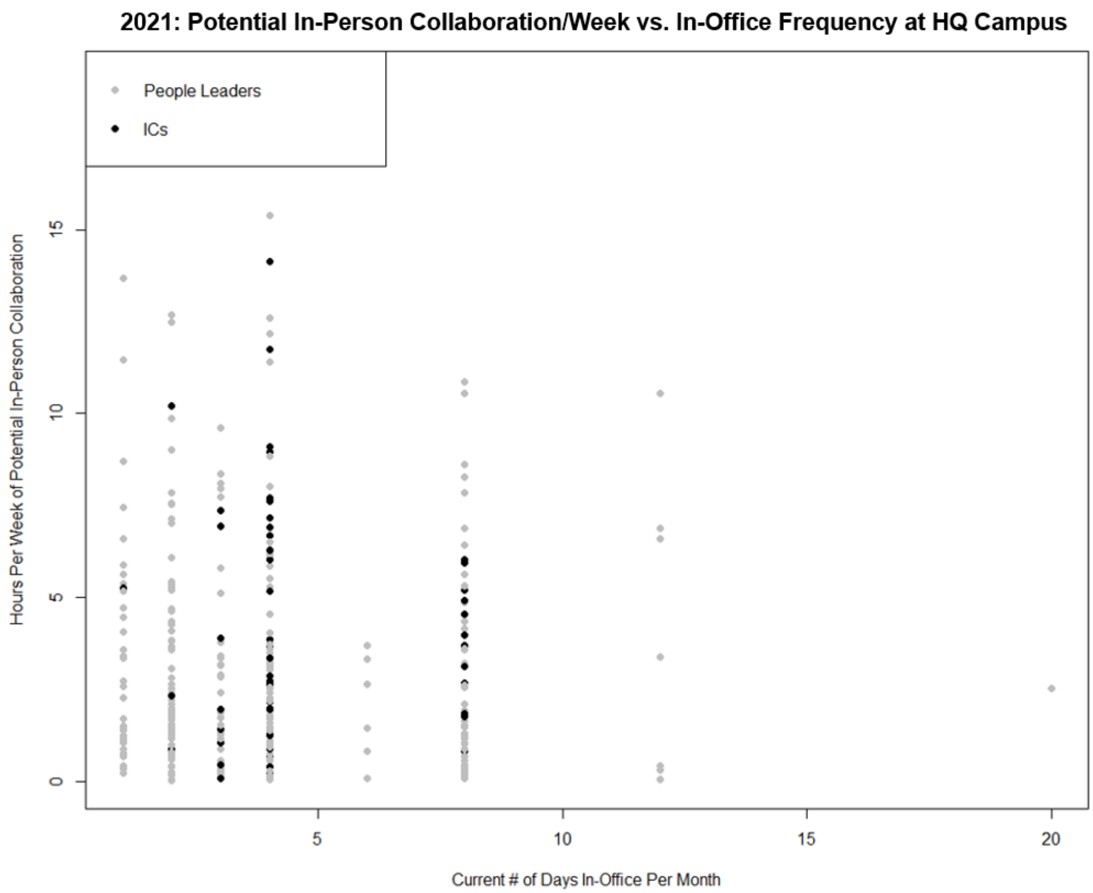


Figure 4-9: 2021 Potential In-Person Collaboration vs. Assigned In-Office Frequency

n=346 Hybrid Employees at HQ Campus

First, we can use these data to confirm that once we filter collaborations to a specific campus, the number of collaborative hours drops significantly. In this sample, the maximum number of collaborative hours is just over 15 hours per week. Previously, in our initial 600 employee sample we saw employees collaborating up to and above 40 hours per week when we did not align by campus—as a global organization teams/business units are geographically distributed. This tells us that the requisite amount of collaboration at an employee's assigned campus is much lower than executives might estimate through intuition.

More importantly, Figure 4-9 tells us that the "one size fits all" approach to in-office frequency is not delivering an optimal mix of in-office/remote work. This was our hypothesis since the HR Schedule Recommender Tool relied on the director's view of how collaborative his/her business unit is—the data confirms this hypothesis. If frequency was being properly calibrated, we would expect to see a meaningful correlation between in-person collaboration and monthly in-office frequency.

There are two primary arguments against bespoke frequency recommendations—it is challenging to scale organizationally and employees may bias the process if each fills out an individual frequency recommender tool. We believe the collaboration index can solve both of these. As for scaling, a simple data pipeline can transform calendar data and produce a collaboration index for each employee. This metric is the only factor that drives the frequency recommendation—yielding recommendations commensurate with an employee's potential in-person collaborations. An important word here is *potential*. This removes the bias; an employee cannot skew the recommendation by never going into the office—the potential interactions will still be counted and drive his/her frequency recommendation.

To illustrate how the collaboration index would change Figure 4-9, we conduct a simple simulation. We calculate the collaboration index for each employee in the sample and assign the employee frequency according to the map displayed earlier in Table 7.8. The results are displayed in Figure 4-10, and we see each employee's frequency scale relative to in-person collaboration.

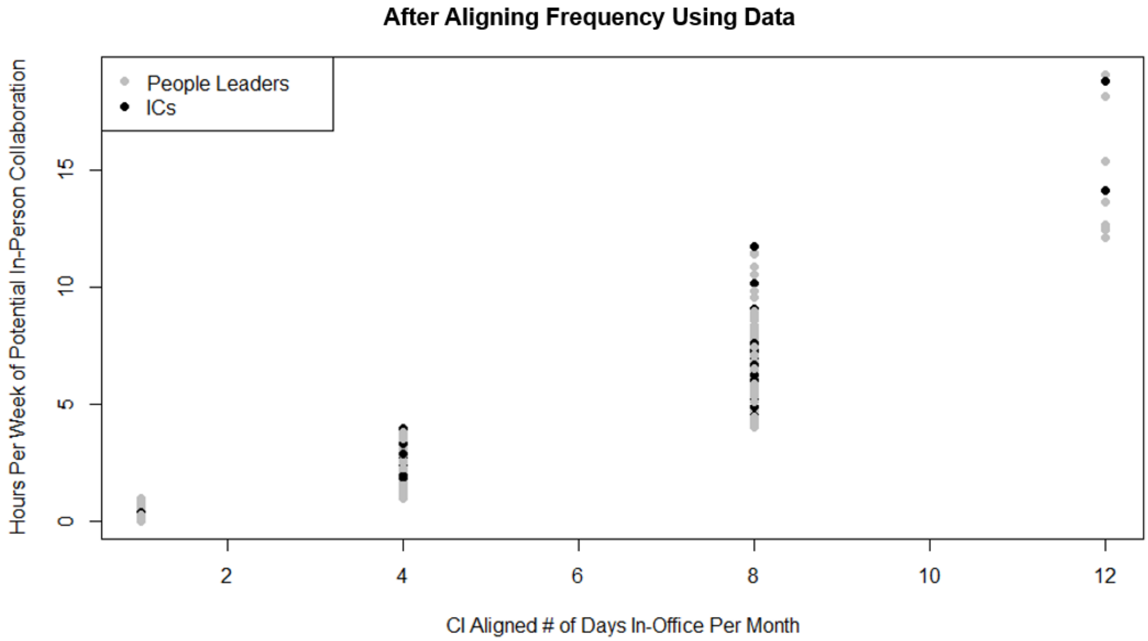


Figure 4-10: Figure 4-9 After Aligning Frequency Using Collaboration Index

n=346 Hybrid Employees at HQ Campus

Ultimately, calendar data produces a frequency recommendation free from the opinion of each individual leader or employee. This is important from a real estate perspective because it allows for more precise long term capacity planning—in short, more OPEX savings because real estate capacity is matched to demand. This concept of matching capacity to demand is not new—it is a core principle in manufacturing. The era of hybrid work brings this principle to light in the commercial real estate domain. Organizations that use data to recommend frequency will be a step ahead of the market because they can shed expensive assets while other organizations wait to decipher frequency via employee behavior—a signal that will likely take a number of years to stabilize post pandemic.

This leads into the next factor in hybrid work design—how can we forecast employee behavior with limited historical data?

4.3 Modeling System Constraints at Hub Sites

As presented, we see that frequency is the first decision impacting real estate network capacity. This section focuses on the second, and arguably harder to decipher factor, employee behavior—how often will hybrid employees actually come to the office post pandemic? We believe that setting hybrid schedules is critical to begin influencing employee behavior; schedules give real estate strategy a baseline occupancy estimate and provide predictability for work groups to connect—we urge organizations to think about implementing a simple scheduling system. By setting schedules, organizations can structure an ambiguous problem and model employee behavior. This section describes a general simulation design which organizations can use to model a variety of hybrid work scenarios. Ultimately, this process will help organizations understand the system constraints before designing a more complex hybrid work scheduling algorithm.

The Monte Carlo simulation rests on considering each employee arrival as an independent Bernoulli trial. Since the subject organization is setting schedules, on any given day of the week, an employee is either scheduled or not scheduled. If an employee is not scheduled, he/she can still get a last minute call from a colleague and need to "flex" into the office. As such, employees are initially bucketed into two categories—scheduled or flex. To model mathematically, we define two Bernoulli random variables, A_s, arrival scheduled and A_f, arrival flex.

$$A_s = \begin{cases} 1 & w.p.\ p_s \\ 0 & w.p.\ (1 - p_s) \end{cases}, \tag{4.2}$$

$$A_f = \begin{cases} 1 & w.p.\ p_f \\ 0 & w.p.\ (1 - p_f) \end{cases}, \tag{4.3}$$

Since these two events are mutually exclusive—i.e., an employee can only fall in the scheduled category or the flex bucket, not both—the total number of arrivals on day d, A_d is represented by:

$$A_d = \sum_{i=1}^{n_s} A_{si} + \sum_{j=1}^{n_f} A_{fj} \quad , \tag{4.4}$$

where, $n = n_s + n_f$, and:

$$n_s = \text{Total Number of Possible Scheduled Employees on Day } d,$$

$$n_f = \text{Total Number of Possible Flex Employees on Day } d,$$

$$n = \text{Total Number of Hybrid Employees assigned to campus},$$

The average arrivals on any given day is calculated by using the properties of the probabilistic model. By definition of a Bernoulli random variable, the expected values for A_s and A_f are:

$$\mathbb{E}[A_s] = \mathbb{P}(A_s = 1) = p_s,$$

$$\mathbb{E}[A_f] = \mathbb{P}(A_f = 1) = p_f,$$

And subsequently, by applying linearity of expectations, the average arrivals on day d is:

$$\mathbb{E}[A_d] = \mathbb{E}[\sum_{i=1}^{n_s} A_{si}] + \mathbb{E}[\sum_{j=1}^{n_f} A_{fj}],$$

$$\mathbb{E}[A_d] = n_s\mathbb{E}[A_s] + n_f\mathbb{E}[A_f],$$

$$\mathbb{E}[A_d] = n_s p_s + n_f p_f.$$

This progression shows the underlying probabilistic model behind the Monte Carlo simulation—on the average, A_d will simply be a function of the number of employees in each category and the probability that each category of employee arrives to the office. Translating this probabilistic model into a simulation helps real estate leaders understand the variability of arrivals throughout the year. Organizations can use behavioral data to estimate p_s and p_f and test the sensitive of a variety of assumptions—we find this to be a powerful approach to address real estate leadership's uncertainty around how frequency and employee behavior impacts capacity of the system.

Since the subject organization's system design further categorizes hybrid employees as weekly, biweekly, monthly and custom, our team expanded the model in Equation 4.3 to the following:

$$A_{ds} = \sum_{i=1}^{n_{sw}} A_{swi} + \sum_{i=1}^{n_{sb}} A_{sbi} + \sum_{i=1}^{n_{sm}} A_{smi} + \sum_{i=1}^{n_{sc}} A_{sci} \quad , \tag{4.5}$$

$$A_{df} = \sum_{j=1}^{n_{fw}} A_{fwj} + \sum_{j=1}^{n_{fb}} A_{fbj} + \sum_{j=1}^{n_{fm}} A_{fmj} + \sum_{j=1}^{n_{fc}} A_{fcj} \quad , \tag{4.6}$$

$$A_d = A_{ds} + A_{df} \quad , \tag{4.7}$$

where A_{sw}, A_{sb}, A_{sm}, A_{sc} and A_{fw}, A_{fb}, A_{fm}, A_{fc} are Bernoulli random variables corresponding to employee categories of scheduled weekly, scheduled biweekly, scheduled monthly, scheduled custom, flex weekly, flex biweekly, flex monthly, and flex custom. Again, the associated probability of arrival, or p_{sw}, can be estimated based on behavioral data, and can be used to simulate the sensitivity of hybrid work scenarios.

Using the model described in Equations 4.4 through 4.6, we recommend organizations simulate scenarios by drawing from a uniform probability distribution to determine whether employee i, arrives at the office campus on day, d. Algorithm 1 details the simulation used to model a year of hybrid work at the subject organization. At a high level, this simulation runs a trial for each employee and uses the probabilities of arrival to simulate if an employee arrives in the office or does not arrive in the office on a certain day.

Algorithm 1 Simulate a Campus for Year

1: **procedure** SIMULATECAMPUS(n_{sw}, n_{sb}, n_{sm}, n_{sc}, n_w, n_b, n_m, $p_{arrival}$)
2:     n_{sw}, n_{sb}, n_{sm}, n_{sc} ▷ dictionaries with scheduled employee count
3:     p_{sw}, p_{sb}, p_{sm}, p_{sc} ←$\mathrm{p}_{arrival}$ [scheduled] ▷ store arrival probabilities
4:     p_{fw}, p_{fb}, p_{fm}, p_{fc} ←$\mathrm{p}_{arrival}$ [flex]
5:     Initialize $Simulation_{results}$ with columns [Mon,Tue,Wed,Thu,Fri]
6:     Initialize $DaysOfWeek$ to [Mon,Tue,Wed,Thu,Fri]
7:     Intialize Wk_i ▷ Indicator variable to toggle week of the month
8:
9:     **for** weeks 1 to 52 **do**
10:         Wk_i set to week of the month (1/2/3/4)
11:         Initialize $DayCounts$
12:
13:         **for** day in $DaysOfWeek$ **do**
14:             **for all** n_{sw} [Wk_i] [day] **do**
15:                 sample from a uniform distribution
16:                 **if** sample $< p_{sw}$
17:                 **then** increment $DayCounts$[day]
18:
19:             Repeat lines 11 through 14 for each category of scheduled employee
20:
21:             **for all** n_w - n_{fw} [Wk_i] [day] **do** ▷ calculate number of flex possible
22:                 sample from a uniform distribution
23:                 **if** sample $< p_{fw}$
24:                 **then** increment $DayCounts$[day]
25:
26:             Repeat lines 18 through 21 for each category of flex employee
27:
28:         Once week is finished, concatenate $DayCounts$ to $Simulation_{results}$

The scheduling system gives the subject organization more insight into scenarios than the average organization. However, insight on the probabilities of arrival can be gleaned from a combination of historical data and survey data—the organization can track behavior as hybrid employees return, or gather data on expected behavior through surveys. The power of the simulation tool lies in sensitivity analysis—what will it take for our office campus to be over capacity on a given day throughout the year? In the subject organization, we find that the hub campuses will not reach capacity until 100% of scheduled employees and 50% of flex employees come to the office—this is a very high level of attendance considering pre-COVID attendance hovered between 40-60% of assigned employees. The results of one sensitivity study at the headquarters campus is displayed in Figure 4-11.

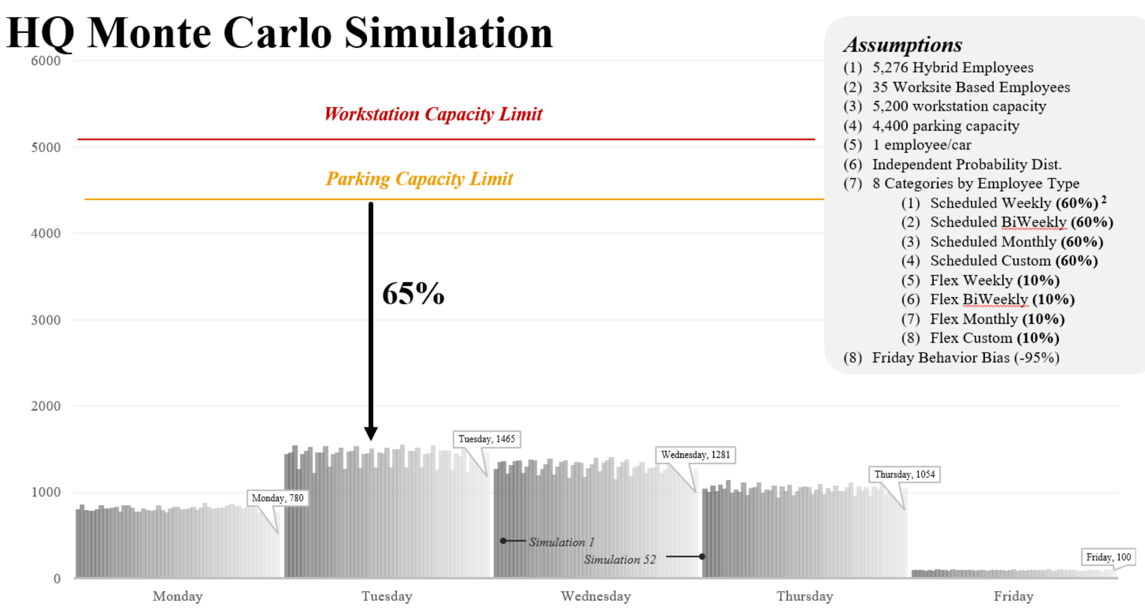


Figure 4-11: HQ Campus Monte Carlo Simulation

Our conclusions from these simulations are threefold. First, the subject organization in our study has, on average, 60% excess capacity at major hub campuses. This is because on average employees are coming in less than one day per week, and schedules are distributed across the week and month. Second, a campus will only approach capacity with a very high schedule adherence (100% scheduled and 50% flex arrivals);

based on pre-pandemic attendance of 40%, this is highly unlikely. This observation is critical as we think about aligning collaboration in this study—our opinion is that a more densely populated office on popular days of the week will deliver greater returns for the employees and the organization. Finally, there may be significant operational savings by closing campuses on low demand days of the week and shifting demand to a more consolidated portion of the work week—i.e., Tuesday, Wednesday, Thursday. In our subject organization, the combination of excess capacity means this is a realistic option to eliminate waste and maximize use of space.

At this point, organizations need to analyze the frequency, behavioral and capacity data discussed to determine the next step in hybrid work design. In our subject organization, we estimate there is significant excess capacity in offices, and the optimal decision is to densely align hybrid employees on three days of the week. For other organizations with capacity concerns, the objective moving forward may be to spread out hybrid demand to ensure that the system is not overwhelmed. In each scenario, the next question is how to align the right employees on the right day to foster the right in-person collaborations—this is the focus of the remainder of this study. Potentially unsurprising at this point, we argue that again, this problem can be solved with employee collaboration data.

Chapter 5

Modeling Employee Networks

In this chapter, we present our process to model the underlying network of knowledge workers. A key finding of this study is that we must understand employee networks before we can properly align schedules. As described in Chapter 3.1 - Current State of Hybrid Scheduling, directors at large corporations are attempting to align hybrid schedules for at times, hundreds of employees—this is an inefficient and ineffective method. The possible permutations of schedules and relationships are too complex for a heuristic solution.

We start this chapter by presenting a process that cleans employee metadata, filters to specific work campus and models a director's network at the work campus. Figure 5-1 visually depicts our process to model director level networks—this chapter discusses our rationale for using networks and then proceeds sequentially through a process to model director level networks.

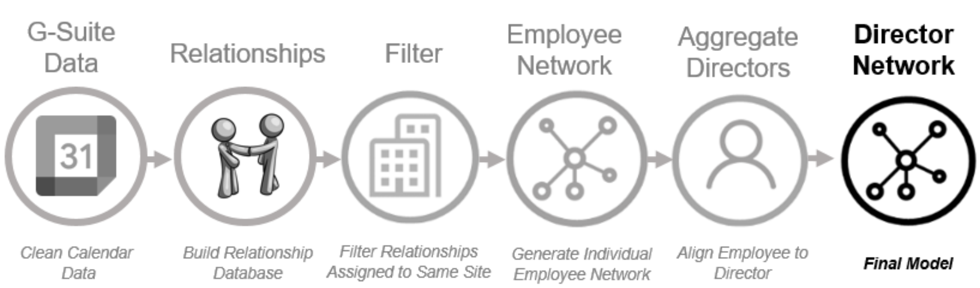


Figure 5-1: Network Modeling Process

The director level network model built in this chapter is the input into an optimization model we describe in Chapter 6. This pipeline improves the current state of hybrid scheduling without a complete process redesign—schedules are still set at the director level but optimized for collaboration.

We continue this chapter by exploring how networks, when applied more broadly, can improve the current scheduling system. We investigate a level beyond the director network model and build an employee level network—where each node in the network is an individual employee. Furthermore, we discuss how business leaders can use network modeling and network metrics to quantify the impact of hybrid work. By the end of this chapter, we demonstrate that network modeling is a powerful tool that captures the abstract notion of employee collaboration. Ultimately, modeling collaboration is the first step to optimizing schedules.

5.1 Network Rationale

In the first two months of research, we conducted numerous small studies measuring schedule adherence across large campuses, using employee behavior as a proxy for schedule effectiveness. Quickly, it became apparent that the highest rate schedule adherence occurred when an employee's schedule was defined as "custom." Even the midst of uncertainty surrounding COVID-19 resurgence in July 2021, employees with custom built schedules—versus the default weekly, biweekly, or monthly recurring option—displayed a rolling adherence of up to 60% compared to less than 20% in the default categories.

In Figure 5-2, we present schedule adherence at the headquarters campus; each line graph represents the schedule adherence for a specific employee schedule type. We primarily analyze data points from the two week period between Tuesday 13 July and Wednesday 21 July. We believe that this data window is valid because it was two weeks after employees began returning to the office and just prior to an increase in COVID19-Delta at the end of July. Leadership soon after announced that the return to office was formally delayed until November. Therefore, this data window is our

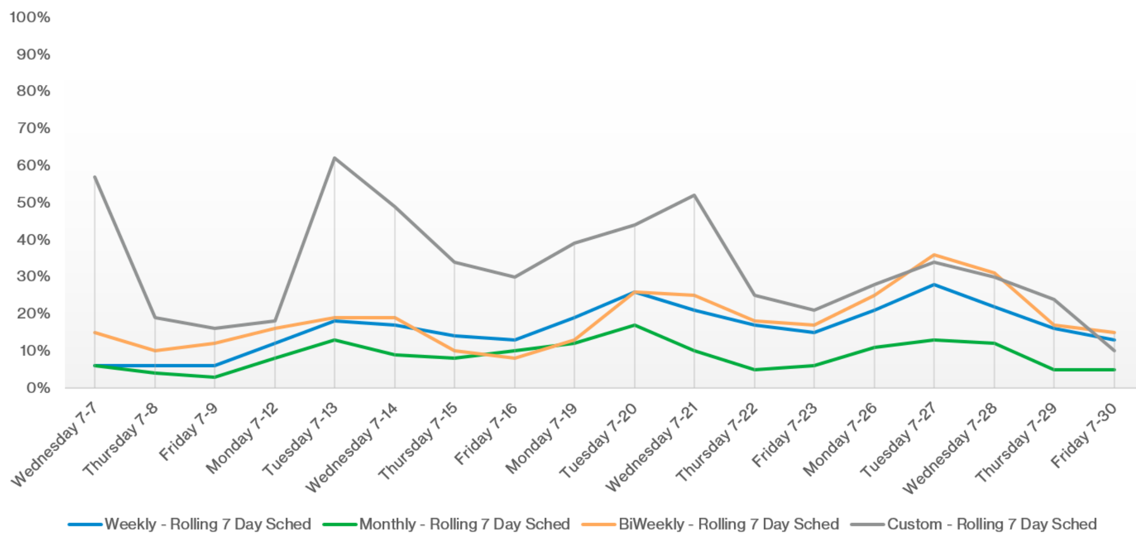


Figure 5-2: Schedule Adherence at HQ Campus
n=2394 Employees Assigned to HQ Campus

only insight into employee hybrid behavior. Based on this data point and consistent with research on remote work by Neely [21], we conclude that intentional schedule setting is critical to maintain employee engagement in the hybrid environment.

Based on this conclusion, we focus on finding a method to properly align schedules using data. This method needs to provide an objective model of employee relationships and be able to scale across a large organization. We leverage the field of organizational network analysis—a technique used in the field of human resource consulting. Historically, organizational network analysis uses employee surveys as the primary data source to model the underlying networks in an organization. This method naturally lends itself to deciphering which employee work groups should be scheduled together.

Cross and Gray published research in June 2021 highlighting the use of organizational network analysis for hybrid scheduling [12]. Their work encountered scheduling practices similar to those in our study—"leaders advocating for hybrid models based on intuition." Cross and Gray worked with a mid-sized biotech organization to model employee interactions and decipher work group “clusters.” These clusters naturally aligned with scheduling recommendations and accounted for 77% of in-person collaboration needed by employees.

We build on this approach by using employee metadata—primarily Google calendar data and employee campus assignments—to fully capture underlying employee networks. Rooting the networks in employee metadata provides a distinct advantage because it is less susceptible to human bias and allows analysis to scale across a large, dynamic organization. Our vision is a scheduling system that is re-calibrated quarterly as work groups naturally change and evolve. Our hypothesis is that models incorporating employee collaboration data will present measurable improvement over heuristic schedules.

5.2 Network Theory Overview

Organizational networks can be mathematically represented in three basic forms: adjacency lists, edge lists and adjacency matrices [19]. Throughout this study, we use the R programming language to convert networks into a visual format for business leaders; however, it is useful to understand the underlying concept of the adjacency matrix. Behind the visualizations, graphs are stored in matrix form as an Nv x Nv adjacency matrix. For a graph $G=(V,E)$, the adjacency matrix A is defined such that:

$$A_{ij} = \begin{cases} 1 & \text{if } i,j \in E \\ 0 & \text{otherwise} \end{cases} \tag{5.1}$$

Adjacency Matrix Representation

Simply put, A is a square matrix with the same number of rows and columns as actors in the network. These matrices can be represented as "symmetric," correlating to a non-directed network, and "asymmetric," correlating to a directed network. Therefore, a non-directed network of four actors is represented by a four by four adjacency matrix with 1 indicating the presence of a connection between two actors [13]. A simple network model is represented below in Figure 5-3.

	Andrew	Beth	Charlie	Dan
Andrew	-	1	0	1
Beth	1	-	1	0
Charlie	0	1	-	0
Dan	1	0	0	-

Figure 5-3: A Simple Adjacency Matrix for Four Node Network

The adjacency matrix can be combined with an edge weight database, which classifies the strength of connection between two nodes, to generate a network model from the data. Behind each node there are also embedded nodal attributes. We will detail the edge and node database structure in depth in the coming sections of this chapter. The power of this modeling technique is that it captures the interdependence of nodes—unlike linear modeling techniques, we cannot assume independence when conducting analysis. Interdependence is powerful when we think about applying networks to help solve hybrid work challenges—knowledge work is not linear nor independent and relationships cannot be fully modeled by an organizational hierarchy.

With this basic definition of a network, we can define descriptive statistics of networks that will help our analysis throughout the chapter. The three basic characteristics of a each node in a network are **degree, strength** and **centrality**. The **degree** d_v of a node v, in a network graph $G = (V, E)$, counts the number of edges in E incident upon v. Simply put, this is the number of connections from a specific node v to other nodes in set V that defines our network. We can graph the degree distribution of a network to understand global tendencies. [19] The **strength** s_v of a node v, in a network graph $G = (V, E)$, is the sum of the weights of edges connected to vertex v. This provides a measure of how strong a node is—similar to degree, we can graph a network's strength distribution. [19] The **centrality** is a mathematical measure of importance for a node v, in a network graph $G = (V, E)$. There are a variety of centrality measures that can be used: our analysis uses eigenvector centrality because it is shown to capture the notion of status or prestige. Eigenvector centrality was originally proposed based on the work of Bonacich [5, 19] and Katz

[18, 19] following the form:

$$c_{Ei}(v) = \alpha \sum_{(u,v) \epsilon E} c_{Ei}(u) \qquad , \tag{5.2}$$

This provides an absolute value between 0 and 1 representing the centrality of a node [19]. When visualizing networks, it is typical to scale the size of a node based on the node's eigenvector centrality—all visualizations in this study follow this pattern. Scaling node size provides us with a quantitative and visually intuitive measure of importance in the network. In Figure 5-4, nodes with larger eigen centrality measures appear with a larger diameter and in the center of the network.

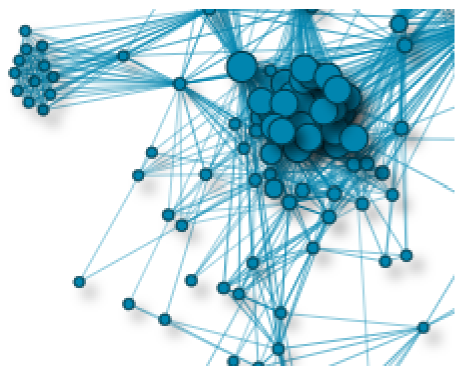


Figure 5-4: Example Eigen Centrality Scaling

This introduction to network theory lays the ground work for our analysis in the coming chapters. We will build upon these foundational ideas to represent and analyze employee networks, and demonstrate how organizations can use network analysis to help design, implement and measure hybrid work.

5.3 Use Case Overview

We start by designing a proof of concept with one business unit in the subject organization. Our initial focus is testing whether network models can optimize hybrid schedules within the constraints of the current system—mainly that directors set schedules for their entire teams. This subsection details our sample population and contextualizes the proof of concept.

The sample population consists of all employees working in a specific business intelligence unit—when the analysis was conducted this unit consisted of 201 employees with 1 director, 4 associate directors, 17 senior managers and 184 data engineering and data science individual contributors. Employees are dispersed geographically across three regional hubs in the United States—the HQ campus in New Jersey, a southeast campus in Atlanta and a southern campus in Texas. These business intelligence teams provide analytical support to the supply chain and real estate business units—these stakeholders are also dispersed geographically across United States hub campuses. The sample population has recurring relationships with a variety of stakeholders across the larger organization. These relationships contribute to the selection of the sample population—we are able to heuristically evaluate model recommendations based on domain knowledge about this portion of the organization. This construct helps us assess the viability and impact of scaling network models beyond the sample population.

The data set consists of 74,000 distinct Google calendar events from June to September 2021. For the proof of concept, the calendar data is largely unstructured and required significant cleaning to create the final data set consisting of a unique meeting identifier and a list of all confirmed meeting attendees. We only had access to calendar data for employees in the business intelligence unit. As such, we place a variety of constraints on our scheduling solution, and believe actual implementation with the full data set will deliver an even larger gain in stakeholder collaboration.

5.3.1 Network Database Design Methodology

As discussed in Chapter 5.2, networks consist of two components—nodes and edges. Behind each component is a database that contains attributes of the node or edge. In our network models, each node represents a specific employee, or a specific director when we aggregate employees to associated directors. The database behind nodes contains 87 attributes for each node. These data points include common attributes such as employee first and last Name, work title, email, business unit, and work location. The node database requires little manipulation—we can use any standard employee detail file. The edge database—or relationship database as we define it—requires more manipulation. The remainder of this section focuses on how to create a meaningful edge database using employee collaboration data.

The edge database stores the strength of connection between two nodes in the network. We start building this database by pre-processing raw calendar data—first we have to determine which employees accepted and declined each invite. This eliminates the noise we find on employee calendars in the remote work environment—e.g., managers who do not attend meetings or working sessions but keep the data on their calendars to understand subordinate focus areas. Then, we model relationships between two nodes using the number of interactions between nodes over a certain time horizon. One meaningful interaction between nodes is a confirmed attendance at the same calendar event. The raw meeting attendee data is transformed to an edge database that contains pairs of nodes and the associated count of connections between these pairs. Technically, we accomplish this by selecting each calendar event, generating the unique pairwise combinations, and concatenating pairs over all calendar events. Figure 5-5 visually depicts the process for each calendar event in our data set. Furthermore, Algorithm 2 provides the detailed process to build the database.

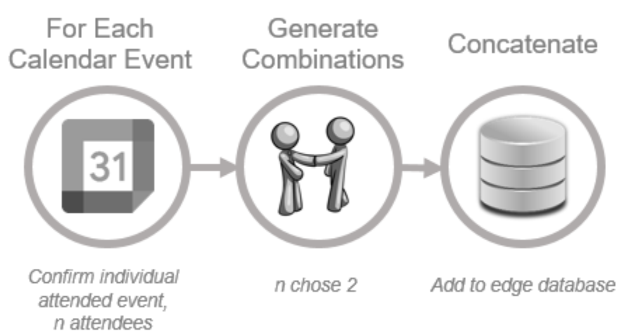


Figure 5-5: Relationship Building Process

When creating the pairs, we are not concerned with order of the connection. We use a function in the R programming language, combn(), to build all pairs of event attendees and then filter out duplicates. Mathematically, we confirm this number

Algorithm 2 Building Relationship Database

```
procedure GenerateRelationshipDatabase(rawCalData)
    Instantiate relationshipDatabase
    Filter rawCalData by unique calendar ID          ▷ Prevents duplicate events
    Clean organizer email, remove @                  ▷ Simplifies matching
    ProcessedCalEvents ← rawCalData
    for len(ProcessedCalEvents) do
        Define actualAttendees
        Pop next event
        attendees ← event[attendees]
        for len(attendees) do
            if attendee[confirmed] = TRUE then
                actualAttendees ← actualAttendees + attendee
        Generate pairwise combos of actualAttendees
        Store each combo as new entry in relationshipDatabase
```

using Equation 5.3 which calculates the number of unique pairs of meeting attendees.

$$\frac{n!}{2!(n-2)!} \tag{5.3}$$

This process results in a database with all the pairwise interactions between nodes in our network—we refer to this as our relationship database as it captures connectivity between nodes. Finally, we aggregate our relationship database by unique pair in order to add a count of "connection occurrences" between two nodes. The final relationship database that creates edges in our network is visually depicted in Table 5.1. We use email addresses as the unique employee identifier, and link the relationship database to node database using each employee's email address.

Node A	Node B	Relationship Strength
Unique Email	Unique Email	Integer $\mathbb{R}+$
Unique Email	Unique Email	Integer $\mathbb{R}+$
Unique Email	Unique Email	Integer $\mathbb{R}+$
Unique Email	Unique Email	Integer $\mathbb{R}+$
Unique Email	Unique Email	Integer $\mathbb{R}+$

Table 5.1: Final Relationship Database Structure

The meaning of this relationship database is rooted in frequency of collaboration between two nodes. We acknowledge that a frequency based metric fails to capture the intensity of collaboration between nodes. However, at a high level, frequency does signal how each employee prioritizes his/her time for collaboration—more collaboration between two nodes means that work is interdependent and these employees can benefit from aligning in-person schedules. The simplicity of the frequency metric also allows us to analyze the distribution of relationship strength and determine a meaningful collaboration threshold based on our sample population.

It is important to note that the strength field of our relationship database could be derived from a variety of underlying data sources. We only had access to Google calendar data, but a more diverse set of underlying data could more accurately portray employee collaboration and produce less biased models. For example, an organization could compile calendar events, emails, and slack communications between employees to build a robust representation of connection strength between two nodes. We chose to call this attribute "strength," versus simply frequency, in an effort to highlight this delineation.

As a recap, a network has two databases that reside behind the nodes and edges. The nodes in our networks are linked to a basic employee database containing numerous attributes such as last name, email, title, business unit, and work location. Our network edges are linked to a relationship database based on employee calendar data. Finally, we associate the node and edge database via employee email address which is a unique identifier contained in each database. Below in Figure 5-6 is the final database design.

5.3.2 Modeling Director Networks at a Campus

The database structure provides the base of a network model. Next, we detail our process that transforms the databases into a director level network model. The process begins by mapping each employee in the relationship database to his/her assigned work location and schedule assigner—this is generally the employee's director or executive director. We analyze the distribution of relationship strength between di-

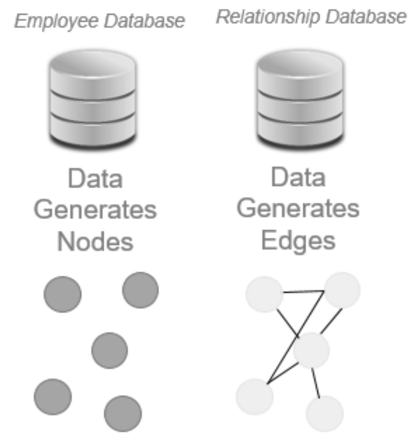


Figure 5-6: Final Database Design for Network Models

rectors and determine the meaningful relationship threshold between two directors—relationships that do not meet this threshold are filtered out of the model. Finally, we use the visNetwork package in R to visualize and analyze this director network model. The descriptive outputs of this analysis become inputs to an optimization model to align schedules in Chapter 6.

The first step in building an aggregated director level network is aligning assigned work locations to the relationship database. We match each employee email in the relationship database to his/her work location. This additional data point enables us to filter to a specific campus of interest—the overall goal is to align in-person schedules at a campus. As such, the relationship database and employee database are filtered to only include employee relationships where each employee is assigned to the same campus. Our proof of concept analysis targets the subject organization's headquarters campus located in New Jersey.

Next, we create a map that takes the emails in the relationship database and matches each to his/her schedule assigner. This should be a simple data pipeline, but we did not have access to a data set aligning each employee to his/her schedule assigner. The majority of employee schedules are assigned by a director; however based on the organizational chart approximately 15% of employee schedules are set by executive director or senior managers.

To solve this problem, we embed the organizational chart into our employee database and assume that if an employee reports to a director, then this director sets the employee's schedule—if an employee reports to an executive director, then this executive director sets the employee's schedule. We tested this process on a sample population of 2,394 employees in wave one. In this early data sample, we had access to the schedule assigner data field, allowing us to validate the interpolation process—we accurately matched 95% of schedule assigners. This interpolation and aggregation results in a database depicted below Table 5.2. The director employee identification (EID) is the transformed unique identifier in the relationship database. This director relationship database represents the aggregated strength of relationships between each director's employees. Ultimately, this database captures the organizational connections that should be driving scheduling decisions.

Node A	Node B	Relationship Strength
Director EID 1	Director EID 1	Integer $\mathbb{R}+$
Director EID 2	Director EID 1	Integer $\mathbb{R}+$
Director EID 3	Director EID 3	Integer $\mathbb{R}+$
Director EID 4	Director EID 1	Integer $\mathbb{R}+$
Director EID 1	Director EID 4	Integer $\mathbb{R}+$

Table 5.2: Director Relationship Database at HQ Campus

Finally, it is important to note that internal relationships between director subordinates are filtered out depending on the modeling goal. The initial network model for our proof of concept excludes internal director relationships—i.e., Table 5.2 "Director EID 1" and "Director EID 1" as Node A and Node B—because we focus on improving stakeholder alignment. The internal/external split does contain valuable information. After presenting our director level model to solve the proof of concept, we generalize how these internal network metrics can help leaders in a hybrid environment.

With this director relationship database established, we analyze the relationship strength distribution to determine the threshold for a meaningful connection between directors. The distribution of relationship strength in our director relationship database is detailed in Table 5.3.

Min.	1st. Qu	Median	Mean	3rd Qu.	Max
1.0	1.0	4.0	20.26	20.00	2115.0

Table 5.3: Distribution of Strength Between Director Organizations

We find that strength between director organizations is a long tailed distribution with the meaningful relationships occurring in the top 10%. Based on this analysis, we choose to select a heuristic threshold and later conduct a sensitivity analysis. The sensitivity analysis of the whole pipeline will be detailed in section 5.7. We set our initial threshold at a value of 100 which captures the top 2.4% of our director level relationships. The distribution is displayed in Figure 5-7—the bottom graph transforms to a log scale using natural log and displays our threshold.

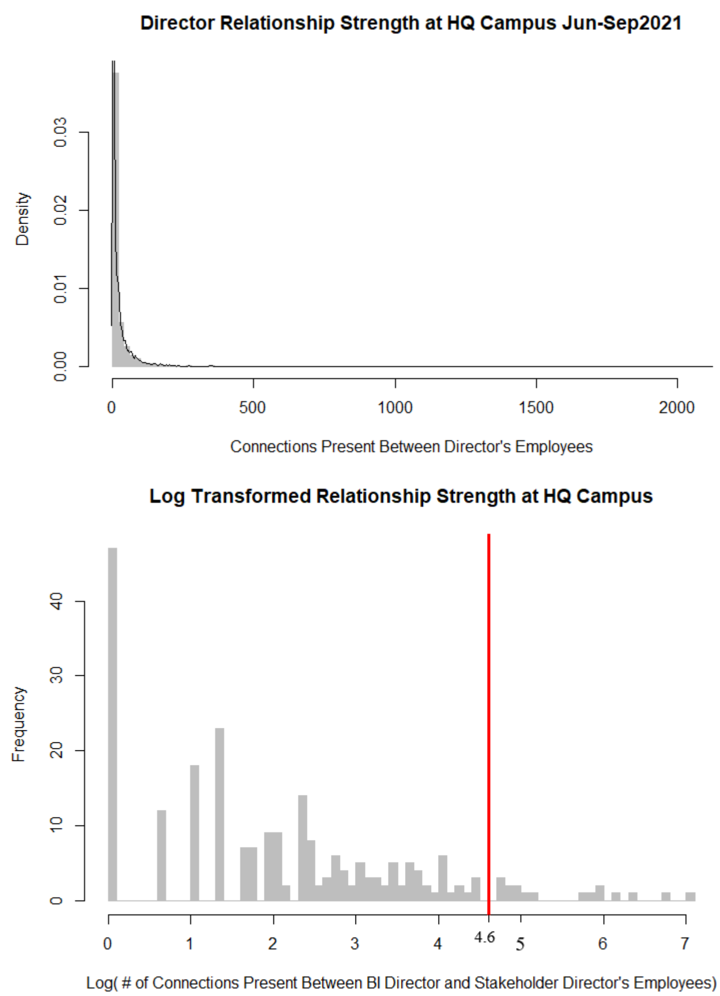


Figure 5-7: Director Relationship Strength at HQ Campus

Based on this analysis, we constrain our director relationship database to only include pairs with a relationship strength greater than 100. This produces a truncated director relationship database with 157 distinct relationships and the distribution displayed in Table 5.4 and Figure 5-8.

In Figure 5-8 we identify six key high strength relationships—this confirms our hypothesis that relationships between certain director's organizations are more robust and vibrant than others. We believe that this finding will hold across any large organization of knowledge workers attempting to orchestrate in-person collaboration in the hybrid environment.

Min.	1st. Qu	Median	Mean	3rd Qu.	Max
100.0	118.0	149.0	186.2	191.0	2115.0

Table 5.4: Truncated Distribution of Strength Between Director Organizations

Next, we actually build the director network model using this truncated director relationship database. The truncation is important because it removes unnecessary noise from the model and enables analysis. In this study, we use two R packages to generate networks—iGraph and visNetwork. We use the suite of iGraph tools to build and analyze networks, and use the visNetwork package for visualizations. VisNetwork allows us to customize visualizations and help leaders understand their organizations key network interactions. As discussed in section 5.2, the size of the node is proportional to a director's eigenvector centrality—this captures the notion of importance of the network and provides a value metric to globally align schedules later in our analysis. The width of an edge is proportional to the strength of the connection between two director's organizations.

Figure 5-9 visualizes our initial network model for the Business Intelligence unit assigned to work at the main headquarters campus.

The network in Figure 5-9 presents a view of how business intelligence employees interact with the Business Intelligence key stakeholders. It is important to highlight that this does not fully capture the stakeholders' extended network; this is because the data set only contains Business Intelligence calendar data and no data on the global

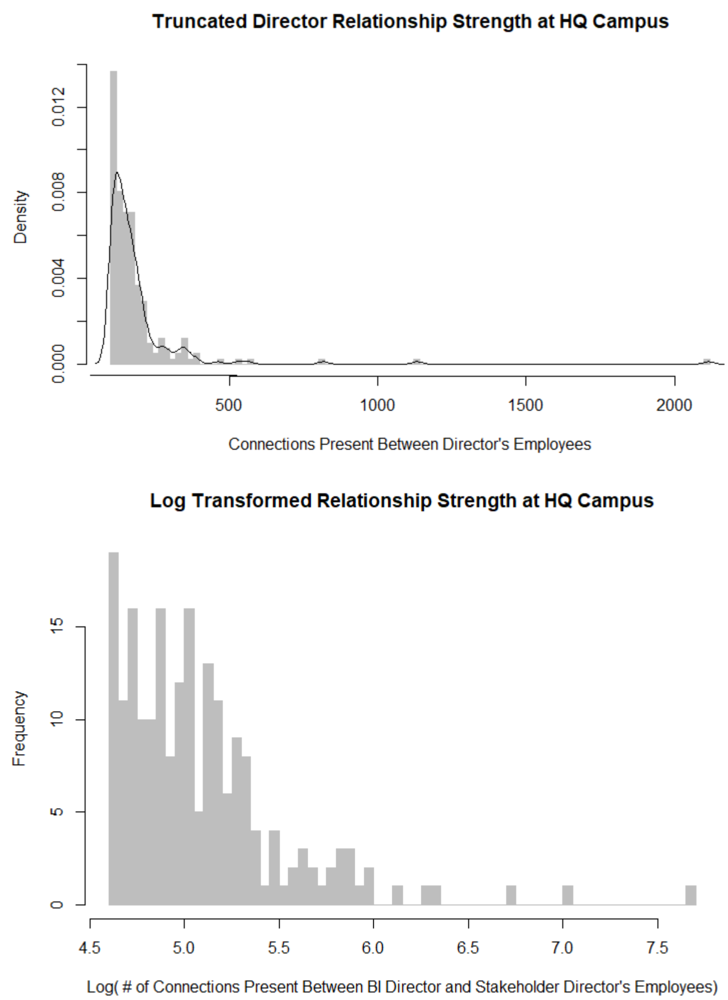


Figure 5-8: Truncated Director Relationship Strength at HQ Campus

calendar activity of stakeholders. Stakeholders only appear in this network when they appear at an event on a Business Intelligence employee's calendar. As we progress to optimizing schedules this limits the increase in alignment we can achieve—for the initial proof of concept we only adjust the Business Intelligence schedule because stakeholders may be coupled to other portions of the global network. With full data access, this modeling pipeline can produce the entire network at a campus site and enable a global analysis.

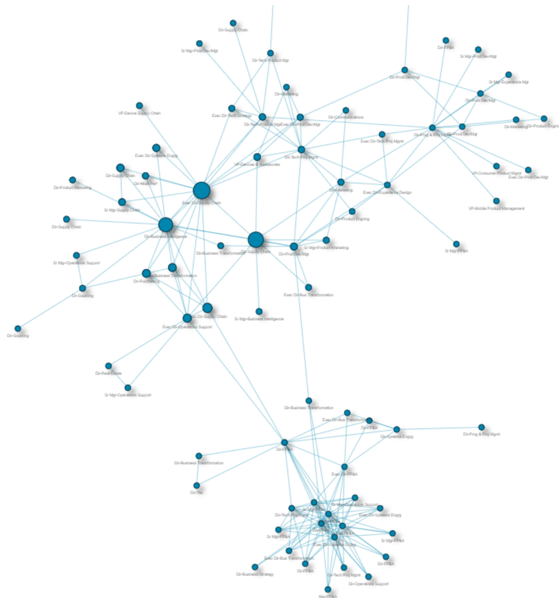


Figure 5-9: Business Intelligence Network (Director Level) at HQ Campus

5.3.3 Business Intelligence Network Analysis

For now, we focus on analyzing the Business Intelligence unit's network. In Figure 5-10, we zoom into the meaningful connections between the Business Intelligence director and her key stakeholders. Each node is labeled with its position—e.g., Director, Business Intelligence—to maintain anonymity throughout the analysis. For ease of reference, we refer to the Business Intelligence director as Director BI and encode her stakeholders with an alphabetic matching A through Q. Each node contains 83 attributes in the node database that were used throughout the analysis to understand relationships. Director BI's key stakeholders are highlighted in red. We can further analyze this data to understand the percentage of relationships we are capturing in this model.

Min.	1st. Qu	Median	Mean	3rd Qu.	Max
1.0	2.5	7.0	36.1	22.0	1134.0

Table 5.5: Distribution of Strength of Business Intelligence Relationships

We observe a few strong relationships which represent the key stakeholders for the business intelligence organization. After truncating the distribution at a rela-

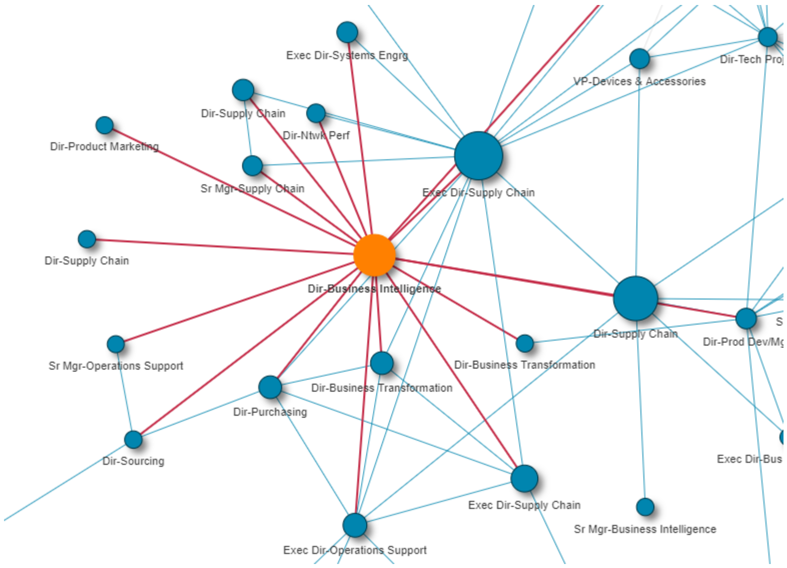


Figure 5-10: Business Intelligence Network (Director Level) at HQ Campus - Zoom-In

tionship strength of 100, we find 17 key stakeholders—the stakeholder statistics are detailed below in Table 5.6. The distribution after truncation is detailed in Table 5.5. These 17 relationships capture 65.9% of all business intelligence interactions at the campus from June to September of 2021. The remaining 34.1% of interactions were spread out across 218 different director organizations. In Table 5.6, the "Rel. % of Engagements" column uses all possible relationships in the denominator—so this column sums to 65.9% or the percent of engagements captured by the top 17 stakeholders. Looking at Director I to Director Q, we observe that the relative strength of the relationships plateaus—these are less frequent than relationships with Director A through H. Therefore, we conclude that the threshold of 100 captures the recurring, meaningful relationships that can benefit from orchestrated in-person collaboration.

The final point to highlight is that the Business Intelligence unit has meaningful relationships with 17 other departments at the campus—in the current state, each director listed in Table 5.6 must communicate with each other and attempt to align schedules. Large organizations arranging hybrid work need to understand the complexity of this problem, especially if they are mandating set schedules. While setting schedules provides structure and can increase meaningful in-person interaction, doing so without analytics presents a high probability for misalignment among work groups.

Stakeholder	Strength	Rel. % of Engagements	Engage/Day	Eigen Centrality
Director A	1134	13.4%	18.9	1
Director B	817	9.6%	13.6	0.879
Director C	563	6.6%	9.4	0.335
Director D	461	5.4%	7.7	0.207
Director E	386	4.6%	6.4	0.241
Director F	366	4.3%	6.1	0.144
Director G	348	4.1%	5.8	0.208
Director H	322	3.8%	5.4	0.166
Director I	166	2.0%	2.8	0.117
Director J	153	1.8%	2.6	0.036
Director K	148	1.7%	2.5	0.0422
Director L	138	1.6%	2.3	0.0322
Director M	128	1.5%	2.1	0.139
Director N	122	1.4%	2.0	0.0953
Director O	120	1.4%	2.0	0.0755
Director P	114	1.3%	1.9	0.0353
Director Q	112	1.3%	1.9	0.0276

Table 5.6: Business Intelligence Stakeholder Metrics

We also analyze the sensitivity of the meaningful relationship threshold. Since the distribution of relationship strength is long tailed, the important relationships fall in the top 10 relationships—below Director I in Table 5.6, the stakeholders represents a small fraction of the overall connections. Our goal is to align the most important interactions; many of the stakeholders on the tail of the distribution are noise. To demonstrate this conclusion, a sensitivity analysis for the meaningful relationship threshold is shown in Table 5.7. The threshold we use to define a meaningful relationship is 100, in boldface in Table 5.7.

Table 5.7 displays how the quantity and relative percentage of relationships retained changes when applying different thresholds for relationship strength. In our networks, we find that selecting a threshold which captures the top 2.5% of relationships is a good rule of thumb. Setting a lower threshold has diminishing returns, as each additional relationship captured is not strong enough to impact stakeholder alignment. This confirms our observation that business units have recurring relationships which represent the bulk of connections in the underlying network.

Threshold	Retained	Relationship %	Added %/Relationship	Threshold Quantile
50	31	77.0%	0.7%	Top 8.0%
60	27	74.4%	0.7%	Top 6.1%
70	22	70.8%	0.9%	Top 4.5%
80	21	69.9%	1.0%	Top 3.5%
90	17	65.9%	0.0%	Top 2.7%
100	**17**	**65.9%**	**0.0%**	**Top 2.4%**
110	17	65.9%	1.3%	Top 1.9%
120	15	63.3%	1.5%	Top 1.3%
130	12	58.9%	1.6%	Top 1.0%
140	11	57.3%	N/A	Top 0.8%

Table 5.7: Sensitivity Analysis—Relationship Strength Threshold

The director network model in Figure 5-9 and subsequent stakeholder metrics presented in Table 5.6 are the key descriptive inputs to our optimization model in Chapter 6. To stay focused on networks, the next section shifts to thinking more generally about how networks can help answer many hybrid work design questions. We return to the director level network model in **Chapter 6** when we optimize the Business Intelligence hybrid schedules.

5.4 Generalized Employee Network Modeling

Since this chapter is focused on network models, we refrain from moving straight into the optimal proof of concept solution. Instead, we use the business intelligence data to generalize the value of network models in the hybrid work environment. In this section, we introduce a class of statistical models to analyze networks, present a process to model the individual employee level network and develop network metrics hybrid organizations can track. This general application of networks is not critical to solving the proof of concept, but is powerful nonetheless. Our goal is to emphasize the complexity of hybrid work and illustrate the power of network analytics.

5.4.1 An Introduction to Exponential Random Graph Models

Network scientists initially conducted statistical analysis of networks using standard regression methods [14]. As we have already shown, network relationships are rooted in dependence and thus violate key independence assumptions that enable traditional regression modeling [14]. The exponential random graph model (ERGM) was developed to fill this analytical gap—originally referred to as p^* models in the early 1970s, these evolved to modern ERGMs. Currently, ERGMs are used by social scientists to understand the local forces that shape global network structure [16]. We introduce the basics of ERGMs and demonstrate how this modeling technique can provide hybrid work insights.

EGRMs are analogous to logistic regression—they predict the probability of an event occurring given a set of exogenous variables [17]. In the network context, we are concerned with the probability that a pair of nodes develop a tie based on sharing certain attributes. As an example, are two employees more likely to have a meaningful relationship if they work for the same manager or align under the same business unit? The underlying premise of EGRMs is simulating randomly generated exponential graphs and comparing these with our real world network to determine significance of certain attributes.

The general formulation of an ERGM considers *G=(V,E)* as a random graph. Let $Y_{ij} = Y_{ji}$ be a binary random variable representing the presence of an edge between nodes—this produces a matrix $\mathbf{Y}$ that is the random adjacency matrix for graph G. Therefore, an exponential random graph model is formulated to estimate the parameters of the joint distribution of $\mathbf{Y}$ [19]:

$$P_\theta(Y = y) = \left(\frac{1}{\kappa}\right) \exp\left(\sum_H \theta_H g_H(y)\right) \tag{5.4}$$

where,

1. each H is a set of possible edges among the vertices in G

2. $g_H(y) = \prod_H Y_{ij}$, meaning it is 1 if H is in Y or zero otherwise.

3. a non-zero value for θ_H means that the Y_{ij} are dependent for all pairs of vertices i, j in H, conditional upon the rest of the graph

4. $\kappa = \kappa(\theta)$ is a normalization constant,

$$\kappa(\theta) = \sum_y \exp\left(\sum_H \theta_H g_H(y)\right)$$

This formulation creates a dependency among elements of $\mathbf{Y}$ and allows simulations, specifically Markov Chain Monte Carlo simulations to assess significance of attributes in the network under study.

5.4.2 Building ERGMs for Our Director Network

We ran experiments fitting ERGMs to the director level network built earlier in this chapter. We also use ERGMs on the more granular employee level network that we build later in this chapter. Our conclusion is that ERGMs are a powerful class of models that organizations can use to understand what drives connection in their networks. By understanding which attributes drive relationships we can start aligning hybrid work based on meaningful factors. In our director level network, we confirm our hypothesis that directors aligned within the same business unit are more likely to share a meaningful relationship. A more telling insight is that when we match director schedule preferences in the network, we do not find evidence that similar schedules increased the probability of a meaningful relationship. This confirms the hypothesis that schedules are not aligned with key stakeholders.

Building on the formulation presented in Section 5.4.1, R provides a practical package to build and test a variety of ERGMs on networks. An procedure to fit an ERGMs is detailed below in Algorithm 3. When the ERGM is built in line 8 of Algorithm 3, we can specify co-variates using the nodecov, nodefactor or nodematch command. Similar to regression modeling, nodecov applies to continuous variables and nodefactor applies to categorical variables. Nodematch assesses the concept of homophily in a network based on a categorical attribute. Homophily is formally defined as "love of sameness;" for our purposes it simply shows whether two nodes are more likely to share a relationship if they also share a particular attribute. This gives us the flexibility to test a variety of hypothesises about our network.

Algorithm 3 Fitting an ERGM

procedure GenerateERGM($NetworkGraph$)
    $AdjacencyMatrix \leftarrow as.adjacency.matrix(NetworkGraph)$
    $NodeAttributes \leftarrow as.data.dataframe(NetworkGraph)$ ▷ Store Attributes
    Transform categorical $NodeAttributes$ into factors
    $NetworkModel \leftarrow as.network(as.matrix(AdjacencyMatrix))$
    for Attribute in NodeAttributes **do**
        $set.vertex.attribute(NetworkModel, AttributeName, list(Attribute))$
    $ERGM \leftarrow ergm(NetworkModel, edges + covariates)$

We gain our first insight by simply matching nodes on an attribute in our node database called PB ORGANIZATION—this is a slightly more granular attribute than business unit; it contains classifications such as Supply Chain and Real Estate, Consumer Marketing Ops, and Corporate Finance. There are actually five different variations of "business unit" connected to each employee in our node database and it is not clear which is the most meaningful for work alignment—we imagine this noise is not unique to our use case.

ERGMs cut through the noise and decipher which attributes should drive hybrid work alignment. When we match on PB ORGANIZATION, we find that director's sharing the same PB ORGANIZATION code have an 80% chance of sharing a meaningful relationship. This finding is supported by a high degree of statistical significance, a p-value of .0001. Results from fitting via Monte Carlo maximum likelihood estimation are detailed below in Table 5.8. Our key result is a simple scheduling heuristic: match director schedules based on PB ORGANIZATION code in the human resources database. In our sample, when compared to randomly assigning schedules, this would provide 4 to 1 odds of successful alignment—a 3x increase from the baseline.

Following this insight, we build models to test schedule alignment in our director network. We incorporate the day of the week that each node (director) planned

Nodal Match/Covariate	Coeff. Est.	Std. Error	Z-Value	P-Value
Edges	-3.0096	0.1024	-29.381	.0001***
NodeMatch(PB ORGANIZATION)	1.4292	0.1874	7.628	.0001***
Null Deviance: 3348				
Residual Deviance: 1069				
AIC: 1075 BIC: 1085				

Table 5.8: ERGM 1 - Matching Nodes on HR Organization Designation

for employees to come into the office. If schedules are aligned, we expect to see a significant relationship similar to our observation in ERGM 1 matching on business unit. Our ERGM 2 matched on business unit incorporates the following co-variate attributes from nodes: days scheduled, monthly in-office freqency, eigen centrality, and betweenness centrality.

Table 5.9 displays our results for the more robust model. ERGM2 matches nodes by granular business unit alignment and an indicator variable designating a node's schedule. For ease of viewing, we only show results when matching Wednesdays—the conclusions hold when we fit this model matching Tuesdays and Thursdays. The data contained a limited number of observations with Monday and Friday schedules. We observe that nodes sharing the same business unit and schedule are not more likely to share a meaningful relationship—when we couple this with the significance of business unit alignment, it is another data point supporting the hypothesis that schedules are not aligned. Furthermore, when analyzing the coefficients in ERGM 2, we see that "Scheduled Wednesday's" actually has a negative coefficient. Although not statistically significant in the model, it is still concerning when our goal is to have work groups in the office with collaborators. Unsurprisingly, nodes with the highest monthly in-office frequency (12 times per month) have a statistically significant increase in the number of meaningful relationships.

Overall, our key finding from this modeling exercise is that organizations are not placing enough analytics behind the alignment of hybrid schedules. We confirm our hypothesis that a certain business unit identifier—in our case the PB ORGANIZATION classifier—can provide a simple heuristic to align schedules. ERGMs are a

Nodal Match/Covariate	Coeff. Est.	Std. Error	Z-Value	P-Value
Edges	-4.046	0.721	-5.609	.0001***
NodeMatch(PB ORGANIZATION)	1.661	0.314	5.284	.0001***
NodeFactor(Month.Frequency = 4)	0.385	0.402	0.958	0.338
NodeFactor(Month.Frequency = 2)	0.307	0.542	0.576	0.571
NodeFactor(Month.Frequency = 3)	0.421	0.313	1.344	0.178
NodeFactor(Month.Frequency = 8)	0.419	0.824	0.509	0.6106
NodeFactor(Month.Frequency = 12)	0.553	0.298	1.854	0.064*
NodeFactor(Month.Frequency = 6)	0.347	0.425	0.817	0.4142
NodeCov(Eigen Centrality)	0.459	0.376	1.219	.2229
NodeCov(Betweenness Centrality)	0.002	0.0003	6.96	.0001***
NodeMatch(Scheduled Wednesday)	-0.907	.683	-1.328	0.1842
Null Deviance: 1698.2				
Residual Deviance: 439.8				
AIC: 461.8 BIC: 518				

Table 5.9: ERGM 2 - Matching Nodes to Organizations and Schedules

tool that organizations can use to understand what drives work groups and align schedules based on those attributes. As we saw with our proof of concept data, we can experiment with different granularity of attributes to find a meaningful level for employee scheduling. ERGMs are not key to finding the optimal solution for our use case, but they help our team confirm many hypothesises surrounding scheduling. Organizations designing hybrid work can use this modeling technique to decipher the right echelon at which to set schedules and test schedule alignment before it impacts organization productivity.

5.4.3 Creating an Employee Level Network

Next, we focus on understanding the larger network contained within the aggregated director level network. We start this section by discussing a method to build the employee level network—where each node is an individual employee versus a director—and we progress to show how organizations can analyze employee level networks using metrics and tools introduced in Sections 5.2 and 5.4.1. Aggregating by director was useful for our proof of concept because this is how scheduling currently takes place—

our conclusion is that for large organizations this aggregation obscures the richest insights of network analysis. The true value of network analysis lies in transforming data to model relationships that are not apparent on an organization chart.

The modeling process for an employee level network is similar to the director level process presented earlier in this chapter. We filter our data to a particular work campus—in this case the organization headquarters in New Jersey—and design a node database where each entry is an individual employee, and an edge database that embeds individual employee relationships. Since we presented the algorithm in detail previously, we move directly to the model and analyze the distribution of relationships. The models presented in this section use the same data as our proof of concept and focus on one campus location. The only nuance is that each node now represents a single employee versus a director's organization.

We begin by analyzing our relationships and estimating a meaningful relationship threshold. The distribution of employee relationship strength is summarized below in Table 5.10.

Min.	1st. Qu	Median	Mean	3rd Qu.	Max
1.0	1.0	2.0	7.5	9.0	156.0

Table 5.10: Distribution of Individual BI Employee Relationship Strength

In Figure 5-11, we display the distribution and log transformation to understand the distribution of relationship strength, and to visualize our meaningful relationship threshold.

Again, we set a threshold that represents a meaningful relationship between two employees, and truncate our data set prior to building the employee network model. We set our meaningful relationship threshold at a strength of seven—this captures all relationships above the third quartile and truncates the dataset from 19,866 pairs to 5,444 meaningful employee relationships. In Table 5.11 and Figure 5-12 below, we display summary statistics of the truncated distribution.

Framing this data in terms of weekly employee engagements helps with comprehension. In this light, on the minimum end of this relationship spectrum, employees

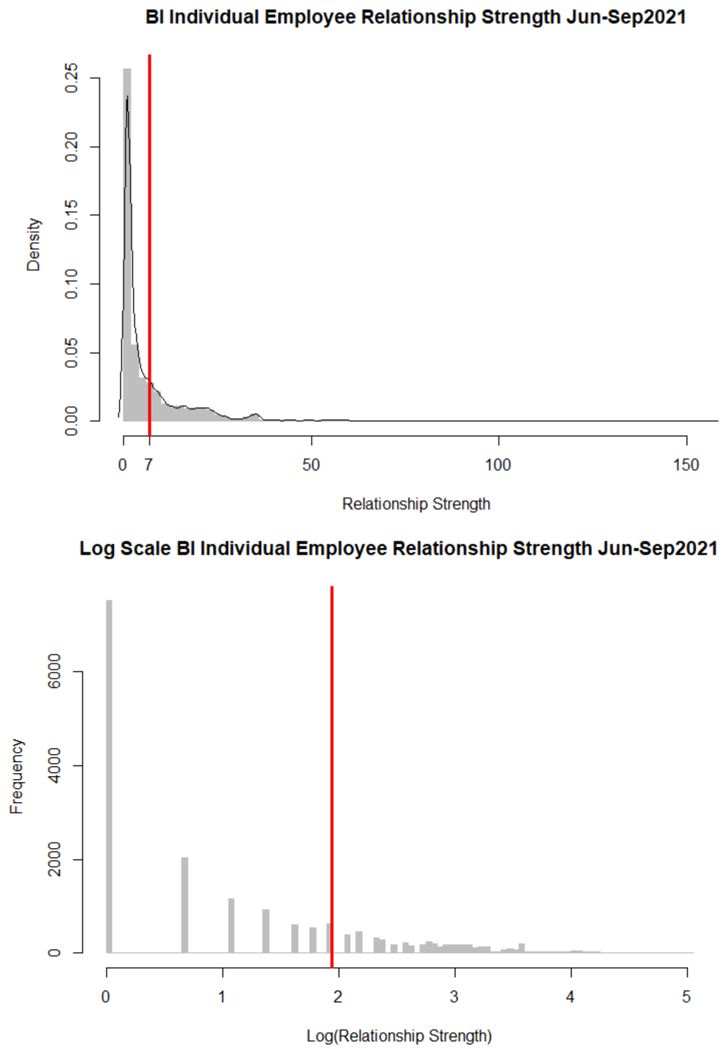


Figure 5-11: Individual BI Employee Relationship Strength at HQ Campus with Threshold in Red

encountered each other eight times over a 12 week period—meaning three weeks out of the month they had an interaction. On average, employees met 21 times during the 12 week period, or 1.7 times per week. Based on this frequency transformation, we believe this is an accurate representation of a meaningful, recurring relationship between two employees—on balance, it seems reasonable that employees meeting about two times per week have a recurring relationship.

Min.	1st. Qu	Median	Mean	3rd Qu.	Max
8.0	11.0	17.0	21.3	25.0	156.0

Table 5.11: Truncated Distribution of Individual BI Employee Relationship Strength

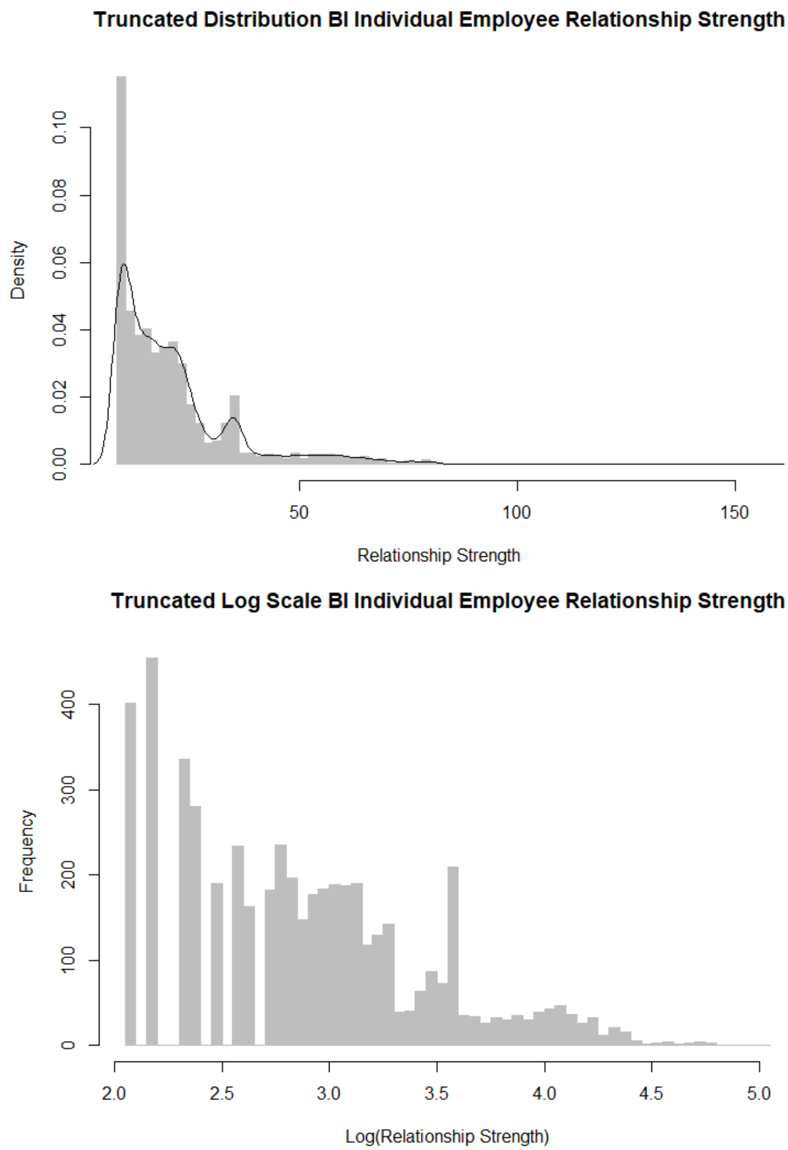


Figure 5-12: Truncated Individual BI Employee Relationship Strength at HQ Campus

We use this employee level data to generate the global BI network, and present the network using VisNetwork in Figure 5-13. In this network each node is an individual employee that is either a member of the Business Intelligence unit or has a meaningful relationship with Business Intelligence employees. Again, with the limited data set the extended global network of all employees in not captured in this network—data is only drawn from Business Intelligence calendars. As hypothesized, this network is complex and there are natural work groups that form. To understand where the Business Intelligence employees lie in the network, in Figure 5-14 we highlight and observe that they manifest across the network and work with a variety of stakeholders.

Figure 5-13: BI Employee Level Network at HQ Campus

Figure 5-14: BI Employees Highlighted in Network

This visual confirms our hypothesis that BI employees collaborate in a variety of different work groups. We believe this is not unique to the specific use case—the nature of knowledge work is matrixed. The employee level network illustrates that scheduling simply by a business unit identifier is not granular enough to orchestrate meaningful in-person collaboration at scale.

5.4.4 ERGMs using the Employee Network

ERGMs can help organizations quantify visual observations in an employee level network model. This section follows the same progression as the previous ERGM analysis and demonstrates that the level of network granularity provides different insights. We find that at the employee level, aligning on the business unit identifier is not significant to orchestrate collaboration. Furthermore, we confirm empirically that schedules are misaligned between key stakeholders in the network.

Our first employee level ERGM tests if aligning on a business unit code can provide meaningful schedule alignment for hybrid employees. The results of ERGM 3 are presented below in Table 5.12. The business unit code is not a factor that drives relationships between employees. This confirms the observation made in Figure 5-14—business intelligence employees are spread throughout the network and engaging with stakeholders from a variety of business units.

Nodal Match/Covariate	Coeff. Est.	Std. Error	Z-Value	P-Value
Edges	-3.397	0.021	-161.739	.0001***
NodeMatch(PB ORGANIZATION)	-0.0458	0.058	-0.781	.435
Null Deviance: 115,667				
Residual Deviance: 23,742				
AIC: 23,746 BIC: 23,765				

Table 5.12: ERGM 3 - Employee Network Matching HR Organization Designation

Next, organizations can investigate if a certain echelon in the leadership hierarchy drives network connection. In our current state scheduling system, the director level sets the schedule, a level we hypothesize is not granular enough. Table 5.13 displays ERGM 4 which tests if sharing a director influences the probability of sharing a meaningful relationship between employees.

Nodal Match/Covariate	Coeff. Est.	Std. Error	Z-Value	P-Value
Edges	-3.399	0.019	-172.508	.0001***
NodeMatch(Director Name)	-0.361	0.203	-1.78	.075*
Null Deviance: 115,667				
Residual Deviance: 23,739				
AIC: 23,743 BIC: 23,762				

Table 5.13: ERGM 4 - Employee Network Matching Employee Director

Nodal Match/Covariate	Coeff. Est.	Std. Error	Z-Value	P-Value
Edges	-3.397	0.019	-171.959	.0001***
NodeMatch(Tier 5 Leader)	-0.339	0.163	-2.081	.0374*
Null Deviance: 115,667				
Residual Deviance: 23,738				
AIC: 23,742 BIC: 23,761				

Table 5.14: ERGM 5 - Employee Network Matching on Tier 5 Leader, a proxy for Director

ERGM 5 simply validates our process to match employees to director names in our database. The coefficients of the models can be interpreted using log-odds—with statistical significance, employees that share the same director are actually 33% less likely to share a meaningful relationship. Framed differently, this means that in our current system, randomly assigning employee schedules would provide a 33% increase in the odds of an employee being in the office on the same day as his/her key collaborators. While this finding is only supported by a p-value of 0.075, we believe it highlights a flaw in the current scheduling system design. More generally, organizations architecting hybrid work can use this modeling technique to decipher the right scheduling echelon. This level will likely be different in every organization. In our proof of concept data, we test a variety of echelons in the organizational hierarchy and find the levels below director—senior manager and manager—are also not statistically significant. This confirms our hypothesis that we need a new method of aligning in-person collaboration.

The final ERGM using the employee network confirms that the subject organization has a schedule alignment challenge. This model matches the human resources business unit designation—e.g., Supply Chain and Real Estate—and matches days scheduled. ERGM 5 summarized in Table **??** only matches on Wednesday, but we also test Tuesday through Thursday to see if this was significant. If schedules were aligned to meaningful relationships, we would expect our model to find "Scheduled Wednesday" as increasing the odds of a meaningful relationship between nodes—Table 5.15 illustrates that this is not the case.

Nodal Match/Covariate	Coeff. Est.	Std. Error	Z-Value	P-Value
Edges	-3.773	0.099	-38.245	.0001***
NodeMatch(PB ORGANIZATION)	-.028	0.059	-0.469	.6389
NodeFactor(Month.Frequency = 2)	0.049	0.056	0.877	0.380
NodeFactor(Month.Frequency = 3)	-0.124	0.103	-1.203	0.229
NodeFactor(Month.Frequency = 4)	0.011	0.043	0.254	0.799
NodeFactor(Month.Frequency = 6)	-0.346	0.359	-0.962	0.336
NodeFactor(Month.Frequency = 8)	0.0007	0.047	0.015	0.988
NodeFactor(Month.Frequency = 12)	-1.336	0.581	-2.299	0.0215*
NodeCov(Eigen Centrality)	1.382	0.047	28.948	.0001***
NodeMatch(Scheduled Wednesday)	0.010	0.038	0.269	0.7881
Null Deviance: 115,667				
Residual Deviance: 23,025				
AIC: 23,045 BIC: 23,138				

Table 5.15: ERGM 5 - Matching Nodes to Organizations and Schedules

In addition to confirming that schedules are not aligned with key stakeholders, this model presents two other interesting insights. First, we observe that employees with a monthly in-office frequency of 12 times a month—or three times a week—are actually 70% less likely to have meaningful connections with nodes in the same business unit that share a Wednesday schedule. It is hard to draw an absolute conclusion; however, this leads us to believe 12 times per month in the office may not be necessary and rather a reactionary solution in an effort to match a variety of stakeholder schedules. This point will be re-engaged as we heuristically map schedules in Chapter 6. Secondly, eigen centrality is clearly an influential predictor to gauge the probability of whether two nodes share a meaningful relationship—this insight also allows us to leverage eigen centrality as we frame an optimization in Chapter 6.

5.4.5 Network Metrics

We continue by discussing key network metrics that can be gleaned from the BI employee level network model. These network metrics can be tracked quarterly—similar to any other business key performance indicator—giving business leaders an understanding of how an organization is evolving in the era of hybrid work. For example, the director of a business development organization may want the organization to have more external connections than internal. Network analytics provides a method to measure if the organization is achieving a goal and a way to visualize how the network has evolved. Furthermore, as hybrid work arrangements unfold every organization has the chance to architect small business experiments to test how certain work designs impact networks—network analytics captures the baseline and enables comparison across time and employee work type.

This subsection presents the four primary descriptive statistics used to describe a network—**node degree, node strength, centrality, and betweenness**. These are analogous to describing the central tendency and measures of variability in a continuous distribution. Next, we present four more sophisticated network metrics—think of these as common business intelligence metrics that quantify business unit performance—**density, transitivity, assortativity, and external-internal index**. In the hybrid work environment, these metrics can be used as benchmarks and quantify the impact of different hybrid work models on organizational cohesion.

Degree

As presented in Section 5.2, the degree of a node describes the number of connections each node shares with other nodes. It can be thought of as a proxy for connectivity of a network. Organizations can use collaboration data from prior to COVID-19 to develop a baseline network degree distribution. Networks can be analyzed in pockets—e.g., at the director level—to understand how the shift to fully remote work has changed the degree distribution of a network. Our hypothesis is that organizations will find that fully remote work can induce collaboration overload—similar to this study's finding in Chapter 4 when analyzing pre-pandemic/post-pandemic collaboration data.

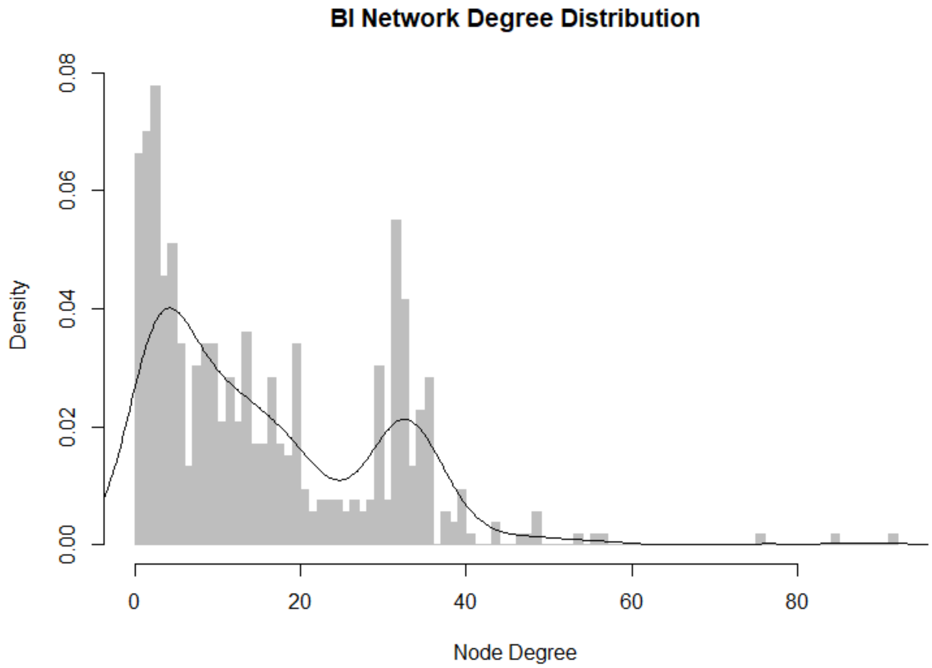


Figure 5-15: BI Network Degree Distribution Q3 2021

Min.	1st. Qu	Median	Mean	3rd Qu.	Max
0.0	4.0	12.0	16.08	29.0	92.0

Table 5.16: BI Network Node Degree Distribution - Fully Remote Environment

We can translate the numbers presented into insights a manager might find useful. On balance, employees in the BI network have 16 meaningful, recurring relationships. At the high end of the spectrum, some BI employees have over 29 meaningful, recurring relationships. As we highlighted in Chapter 4.2, many organizations are making frequency decisions in absence of data—the network node degree distribution can be coupled with meeting analysis to further refine the right frequency recommendation.

A network analytics dashboard could provide directional guidance for leaders by comparing a specific network's degree distribution to the average network profile across the organization. In our example, the director of the BI network in our proof of concept would receive feedback such as "2% Increase in Stakeholder Connectivity since Q2 2021," or "Similar Network Connectivity as other BI Organizations." As employees return, organizations can use in-person collaboration data to measure if schedule and in-office rhythm are helping or hindering collaboration. While this is not a perfect measure, it provides an objective method to assess hybrid work impact on connectivity. As with many business metrics, we believe network metrics are powerful when coupled together, and enable a leader to quantitatively represent a process that is biased towards qualitative decision making.

In addition to tracking network metrics over time, each of the metrics presented in this section allow organizations to execute a variety of business experiments as hybrid work unfolds in 2022 and beyond. An organization can baseline networks pre-COVID, model network metrics throughout COVID with fully remote work, and use in-person collaboration data as employees return in a hybrid schedule to understand the impact. Pre-COVID and during COVID models can be created using the method we presented in our proof of concept. The post COVID in-person networks can be developed using booking data, meeting room data and schedule/adherence data.

Strength

The nodal strength distribution is similar to degree but factors in the weight of relationships between nodes. We believe organizations can use the strength metric alone, or coupled with the degree distribution of the network to produce an average relationship strength metric—average degree of network nodes divided by average strength of network nodes. For a cross-functional organization like the BI organization in our proof of concept, this could be a proxy for global stakeholder connection. In a more nuanced manner, an organization could build the network between two organizations that they desire to work closely together and actually measure how hybrid work is impacting collaboration.

Using our proof of concept data, the strength distribution of the BI network is presented below in Figure 5-16 and Table 5.17. As mentioned, a meaningful metric design for this network might be average strength/relationship—a proxy that helps the BI director understand if hybrid frequency is diluting the quality of stakeholder engagement.

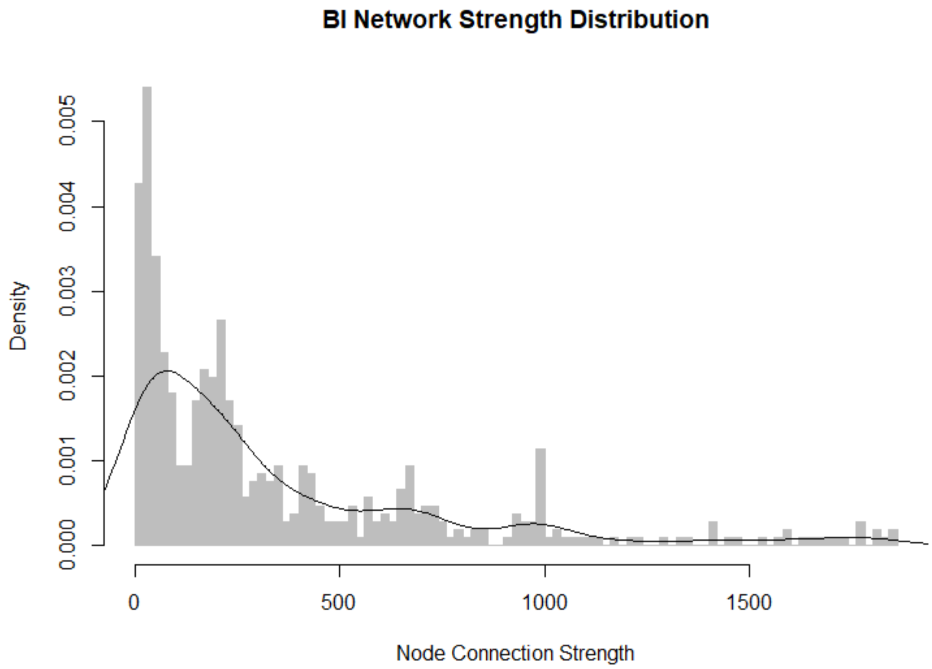


Figure 5-16: BI Network Strength Distribution Q3 2021

Min.	1st. Qu	Median	Mean	3rd Qu.	Max
0.0	53.5	202.0	347.0	468.2	1851.0

Table 5.17: BI Network Node Strength Distribution - Fully Remote Environment

$$Avg.\ Stakeholder\ Engagement = \frac{Average\ Node\ Strength}{Average\ Node\ Degree} \tag{5.5}$$

$$BI\ Avg.\ Stakeholder\ Engagement\ Q3\ 2021 = \frac{347.0}{16.08} = 21.6 \tag{5.6}$$

There are two options for how to use this metric—tracking over time or as a benchmark for leaders. To track over time, we normalize this in a range from 1 to 10 and directors can visualize how a network changes throughout the calendar year. The metric could be applied to compare similar business units coming into the office four times a month versus 12 times a month—does this have significance on in-person collaboration with stakeholders? Our hypothesis is that at a certain frequency threshold—which may be different across organizations—the value of in-office time reaches a point of diminishing returns. In order to benchmark for senior leaders, we can simply model all business intelligence units across the company, normalize the metric to 0 and provide directors a directional gauge compared to the average business intelligence unit.

Eigen Centrality

Centrality provides a mathematical representation of importance in the network. In this study we use eigen vector centrality as detailed in Section 5.2 because it captures the notion of importance or prestige in the network. The eigen centrality ranges from 0 to 1 and can highlight key aspects of the network we are analyzing. The most important node in a network has an eigen centrality of 1. In Figure 5-17 we display the BI Network eigen centrality from Q3 of 2021.

As observed in our BI network visualization, there are clearly important nodes driving much of the connectivity in a network—the nodes with a value above 0.2. As such, the eigen centrality of a node can be factored into the frequency calculation equation. It could be a simple additive structure where nodes displaying above average eigen centrality—e.g., connectors or stakeholder facing roles—are allocated one additional in-office day per week. Again, this would help frequency be linked to data and prevent real estate strategy from maintaining too much or too little office space square footage. While there is not an easy corollary to a leadership metric with eigen centrality, the average eigen centrality of a directors network could be tracked quarterly, and used as a lagging sign of misalignment with stakeholders or lacking in-person connection inside the business unit.

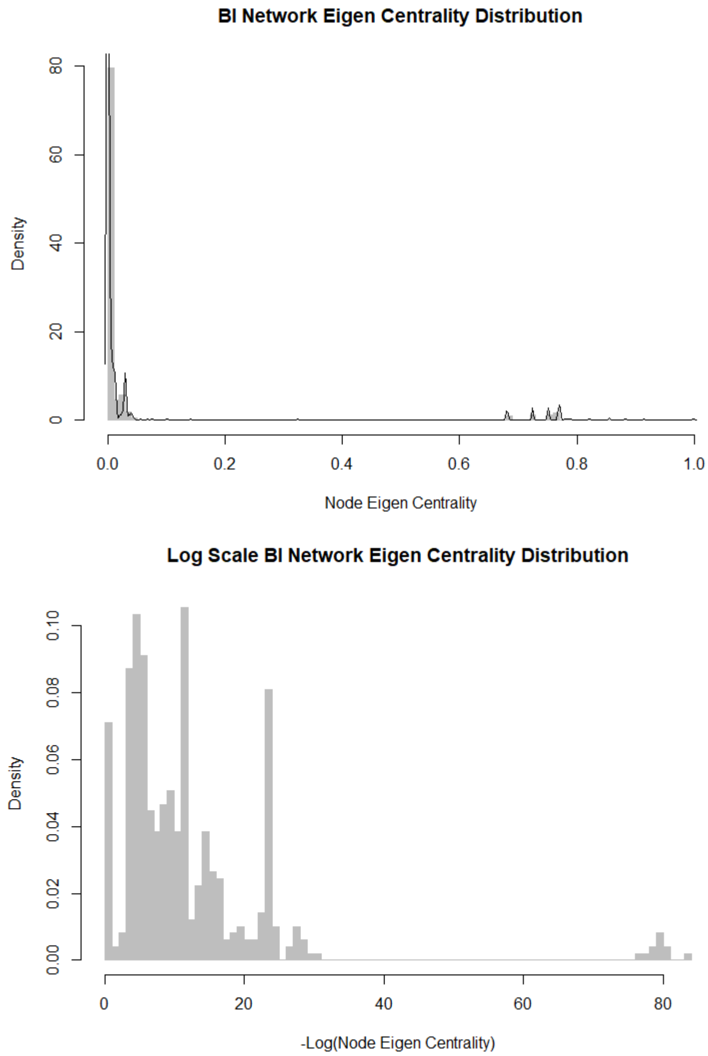


Figure 5-17: BI Eigen Centrality Distribution Q3 2021

Betweenness Centrality

The final network metric is betweenness centrality, or the measure of how nodes lie between each other in relation to the shortest path between nodes. A high betweenness centrality means a node is a power broker—lots of information flows through that specific node. If a network has below average betweenness then it can be interpreted as "flat;" there are not large connector nodes that bottleneck the flow of information.

We believe betweenness is a key metric to monitor in hybrid work—do employees who are in the office more frequently generate a higher betweenness metric? Testing this hypothesis as employees return is critical. If this is true and statistically significant, organizations need to decipher why this is occurring. For a director or senior manager, this information is also useful. A betweenness metric can answer hybrid work questions such as—are my employees who come in more frequently becoming bottlenecks of information? Is my organization remaining as "flat" as during fully remote work?

In Figure 5-18 the BI network betweenness is displayed. We conduct a logarithmic transformation for visualization purposes and see that that average betweenness measures 6.63 on a logarithmic scale. The descriptive statistics are listed in Table 5.18.

Min.	1st. Qu	Median	Mean	3rd Qu.	Max
0.0	0.0	0.667	760.15	444.745	25253.19

Table 5.18: BI Network Betweenness Distribution - Fully Remote Environment

We visualize a business metric titled "organizational hierarchy index" that is scaled from 1 to 10 and uses the logarithmic transform of betweenness to help leaders understand how work design is impacting organizational dynamics. This could easily be interpreted—closer to 10, a business unit has numerous hierarchical bottlenecks that prevent information from moving between key parties; closer to 0, a business unit is more decentralized. Again, this would be useful as a comparison over time and/or used as a benchmark against similar business units across the organization.

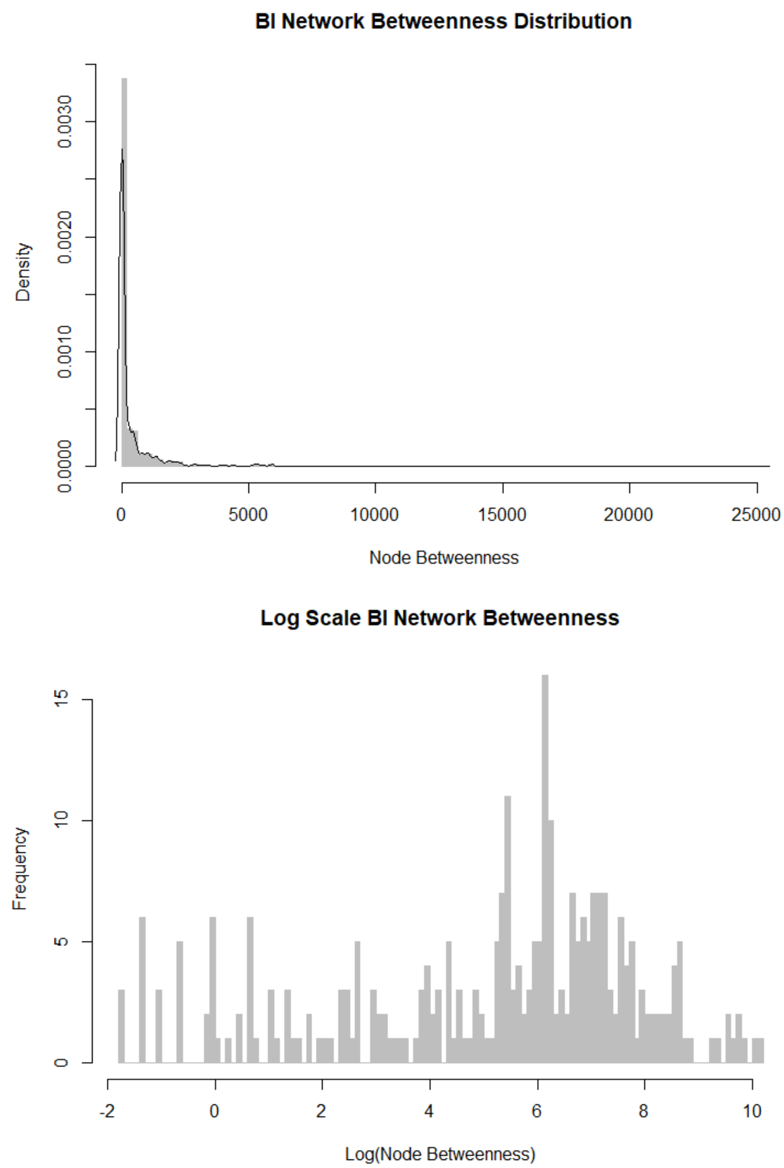


Figure 5-18: BI Betweenness Distribution Q3 2021

Density

Next, we move into descriptive network metrics that are used across network science disciplines to derive insights. The density of a network is simply the number of edges that exist relative to all the potential edges [19]. Mathematically, we can expressed density for a network G as:

$$Density(G) = \frac{E_G}{|V_G|(|V_G| - 1)/2}, \tag{5.7}$$

where V represents nodes in the network G and E represents edges present in network G [19]. The density of a network lies between 0 and 1 and provides one manner of assess cohesion in a network.

Real world networks tend to have low densities: every node is not connected to every other node. For example, in our BI network the density is 0.03. We believe that organizations would have to be deliberate to interpret and use this metric. Looking globally, organizations could benchmark network density during fully remote work and measure changes as hybrid work evolves. This approach could also involve building business unit profiles using data across the organization—analytics teams could then convert the density into a directional measure telling leaders if they fall above or below the average. This could be used to understand how well an internal network is staying connected by only modeling employees inside the business unit—or looking externally, it could be used to triangulate business unit cohesion with stakeholders.

Transitivity

The density of a network can be combined with transitivity to help leaders understand the nature of connection—a low density and high transitivity means work groups are well defined and can be effectively clustered. Transitivity is defined as the probability of two nodes connecting with a third—commonly referred to as the clustering coefficient of a network [19]. Mathematically, transitivity is represented by:

$$Transitivity(G) = \frac{3_\tau \triangle(G)}{\tau_3(G)}, \tag{5.8}$$

where the numerator represents the number of triangles—three nodes connected—found in graph G, and the denominator is the number of potential triangles where three nodes only have two edges. As such, transitivity measures the relative frequency of triangles in the network.

We see transitivity as another metric business leaders can use to assess the impact of hybrid work design. An analytics team can benchmark transitivity before COVID-19 and measure whether hybrid designs are having significant impact on an organization's network transitivity. Again, transitivity can be renamed such that business leaders understand the meaning—"organizational cohesion"—and the measure can be normalized or converted into a percent for ease of interpretation.

Assortativity

Assortativity is the extent to which nodes collaborate with nodes of similar types versus nodes of different types. This can be thought of as a correlation statistic—commonly referred to as the assortativity coefficient [27]. The assortativity coefficient can be calculated on any nodal attribute and defined mathematically as:

$$A = \frac{\sum_i f_{ii} - \sum_i f_{i+} f_{+i}}{1 - \sum_i f_{i+} f_{+i}}, \tag{5.9}$$

where f is the fraction of edges in network G that join a node in the ith category with a node in the jth category, and $f+$ denotes the ith marginal row and column sums of the resulting mixing matrix [19]. More succintly, for all the nodes in our network, we calculate the percentage of nodes that share ties and share the attribute of interest. This metric is malleable as it allows an analytics team to assess how a network "mixes." We believe organizations constructing hybrid environments can leverage assortativity to understand what drives connection. Furthermore, assortativity can be used to test work group schedule assignments—we would want a high assortativity, close to 1, when we calculate assortativity for a work group. In Chapter 7, we build on this idea by assessing clustering algorithms using assortativity.

External-Internal Index

Finally, a powerful network metric that builds on the concept of assortativity is the external-internal (EI) index. Krackhardt and Stern first proposed the EI Index in 1988 to quantify interaction tendencies in social netowrks [13]. Similar to assortativity, this index can be applied on any nodal attribute in the network—in our proof of concept we can apply to director name to understand what percentage of relationships are internal versus external. The index is represented mathematically in Equation 5.10 [13]

$$\textit{E-I Index} = \frac{\textit{External Group Ties - Internal Group Ties}}{\textit{Total Network Ties}} \, . \tag{5.10}$$

The EI index ranges from -1 to 1. An EI index of -1 means that all ties are internal, and an index of +1 means all ties are external. This is a powerful metric because it can be calculated for any nodal attribute. Naturally, the first application on our proof of concept network is calculating the EI index aggregating on director name. This once again confirms our hypothesis that director is not the right level to set schedules—the **EI index = .98** for our BI network when aggregating employees by director name. All else equal, this means that nearly 98% of meaningful, recurring employee relationships occur between employees that do not share a director.

Next, we can analyze just the BI director by using an indicator variable in our node database. The BI Director's EI index = -0.778. This also provides the BI director an indication of how her business unit is behaving—it is slightly more introverted than we might expect for a stakeholder oriented business unit. As organizations transition to hybrid work this metric can help gauge how the organization is evolving based on hybrid work—it would be concerning to see our BI network display a more introverted stance as we track in-person collaboration data going into 2022. Again, this EI index can be tracked over time or benchmarked against similar business unit profiles.

Interestingly, we also find that the EI index can be used to check for schedule alignment in a network. In our network, we encode scheduled days of the week with an indicator in order to allow us to aggregate on this attribute. The EI indices by scheduled day are displayed below in Table 5.19.

EI Tuesday	EI Wednesday	EI Thursday
-0.088	-0.043	-0.003

Table 5.19: EI Indices Aggregating on Schedules

This means that the network is doing a little better than random chance at aligning schedules—an EI of 0 means equal external/internal group ties. Each of the EI indices by day of week are slightly negative, meaning employees sharing a scheduled day are slightly more likely than random chance to also have a meaningful relationship. In a network with correctly assigned schedules, we expect a negatively skewed EI index—employees have relationships with colleagues scheduled on the same day.

5.5 Chapter Summary and Extensions

Network models are an extremely expressive modeling tool that organizations can use to drive hybrid work decisions. Although we initially focused on using networks to solve a problem specific to our proof of concept, network models can help answer a multitude of hybrid work questions. As discussed, many business leaders are concerned that hybrid work will stifle innovation because employees lose serendipitous workplace interactions. Our solution is network representation and analysis methods. Networks naturally provide a model to align work groups and prevent collaboration from stagnating due to misaligned schedules. Furthermore, networks contain a variety of metrics geared toward quantifying how different hybrid schedules impact organizational connectivity.

Chapter 6

Aligning Schedules Under Current System Constraints

Now that we are familiar with network models, our focus returns to finding the optimal solution for the proof of concept. This chapter formulates the goal of stakeholder alignment as a linear program. The director network model from Chapter 5.5 is the primary input to the optimization. In the coming sections, we discuss our proof of concept in depth, translate these details into constraints, formulate an objective and provide the results.

6.1 Proof of Concept Overview and Constraints

As discussed in Chapter 3, many organizations are requiring senior leaders to set schedules for subordinate employees. The goal of the proof of concept is to present a workflow that improves the process without changing the current state. As such, we create a relationship database from the perspective of one such director who leads a business intelligence organization. The nature of work is cross-functional and involves numerous stakeholders.

We pick up from the director level network model developed in Section 5.5—more specifically, the model presented visually in Figure 5-9. This model allows us to identify the top ten stakeholders and understand each organization's eigen centrality, or importance in the network. The first constraint we place on the proof of concept is optimizing for in-person scheduling alignment with only the top ten stakeholders—these top ten account for 84% of the meaningful stakeholder relationships. This simplifies the proof of concept to help executives understand the power of applying optimization. A sensitivity analysis at the end of this chapter presents results factoring in all seventeen business intelligence stakeholders.

Next, all stakeholder schedules and frequencies are fixed. These inputs were submitted by each director and the goal is to realign within the constraints of the current process. The decision to fix each stakeholder's schedule centers around the limited insight gained with our director network model—it was built with only collaboration data from the Business Intelligence employees. This perspective does not capture the larger network. We can not decipher if each Business Intelligence stakeholder is requesting certain days to align with another business unit in the global network. While a limitation in our proof of concept data, this highlights the potential value of a global optimization at the campus level—a network model can capture all of the coupled business units.

From the perspective of the Business Intelligence director, the simplest formulation of stakeholder value is a relative frequency interpretation. The value of stakeholder i is proportional to stakeholder i's relationship strength relative to the top ten stakeholders. Later, to expand this formulation globally we use eigen centrality to leverage the network concept of importance.

The final pieces are modeling the typical month and identifying our objective. While this varies from month to month, we assume the typical month has 20 working days. The objective is to find the four days of the month that the Business Intelligence director should schedule to maximize in-person alignment with stakeholders at the HQ campus.

6.2 Quantifying the Heuristic Solution

We also need to model the current state. The heuristic schedule set by the BI director is a Tuesday/Thursday bi-weekly schedule on the second and fourth weeks of the month. Table 6.1 is a visual mapping of the current state schedules at the HQ campus.

	Monday	Tuesday	Wednesday	Thursday	Friday
Week 1	Director A Director B	Director E	Director A Director C Director D Director E Director F Director G Director I Director J	Director B Director D Director E Director G	
Week 2	Director A Director B	BI Director Director E	Director A Director C Director D Director E Director F Director I Director J	BI Director Director B Director D Director E	
Week 3	Director A Director B	Director E	Director A Director C Director D Director E Director F Director G Director H Director I Director J	Director B Director D Director E Director G	
Week 4	Director A Director B	BI Director Director E	Director A Director C Director D Director E Director F Director G Director I Director J	BI Director Director B Director D Director E	

Table 6.1: Heuristic Schedule Mapping—BI Director Bi-Weekly Tue/Thu

We observe that the BI director is not aligned with the majority of her stakeholders, and our goal is to quantify the improvement analytics can capture. As such, we represent this current state using a binary indicator variable—BI Director equals 1 if scheduled to be in-office, 0 otherwise. This means that our objective of stakeholder alignment, or "alignment value," is a positive, non-zero value if and only if the BI Director is scheduled to be in-office. For a specific day for which the BI director is scheduled, the alignment value equals the sum of the strength of all the stakeholders scheduled on that day. Table 6.2 provides a visual of this transformation. As an example, on the second Tuesday of the month the BI director is scheduled—entry [row = 1, column = 10] equals 1—and the sum of the stakeholders also scheduled on the second Tuesday equals .082. Therefore, the alignment value from the second Tuesday equals .082. The i column helps frame the optimization and allows stakeholders to be referenced intuitively using i from 1 to 10.

Dir	i	Wght	M	T	W	Th	F	M	T	W	Th	F	M	T	W	Th	F	M	T	W	Th	F
BI	N/A	N/A	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	1	0
A	1	.240	1	0	1	0	0	1	0	1	0	0	1	0	1	0	0	1	0	1	0	0
B	2	.173	1	0	0	1	0	1	0	0	1	0	1	0	0	1	0	1	0	0	1	0
C	3	.119	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0
D	4	.098	0	0	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	1	1	0
E	5	.082	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0
F	6	.078	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0
G	7	.074	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
H	8	.068	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
I	9	.035	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0
F	10	.032	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0

Table 6.2: Schedules Transformed to Indicators

Since we are measuring alignment with stakeholders, the baseline value of a biweekly Tuesday/Thursday schedule is 0.87 out of a total possible value of 6.49. This value is calculated by summing the stakeholder weight in each Tuesday/Thursday column when the BI director is scheduled—or where the first row in Table 6.2 equals 1. For ease of interpretation, we convert alignment into a percentage and establish our baseline stakeholder alignment at 13% aligned.

Another point to highlight when mapping schedules is the number of schedule variations that appear in this small sample. Our finding is that the number of schedule options over complicates the decision for directors and causes more alignment issues than a simple, streamlined offering. In Chapter 7, we formally lay out a recommendation for streamlining options. For now, the key observation is that in our sample of 10 directors, there are seven different schedule combinations—bi-weekly Tue/Thu 2nd and 4th week, bi-weekly Wed/Thu 1st and 3rd week, weekly Mon/Thu, weekly Wed, weekly Wed/Thu, weekly Tue/Wed/Thu, and monthly 3rd Wed. Although the weekly schedules are slightly misaligned, much larger misalignment is rooted in the variation created by bi-weekly selections. Part of the misalignment is because the number of weeks in a month is dynamic throughout the year—four weeks versus five weeks—making the rhythm challenging to decipher. In the entire company, 13% were assigned a bi-weekly schedule. One quick step to eliminate schedule misalignment is removing the bi-weekly schedule option.

Table 6.2 is a helpful visual when formulating the linear optimization. Although it may seem obvious in this example that the BI director should shift to a weekly schedule on Wednesdays, our goal is to highlight the power of optimization to executives with this simple use case. The power of this framework is amplified once we gain access to the full data set of collaboration at a campus.

6.3 Optimization Formulation - Simple Model

This section translates the proof of concept overview and baseline data into an optimization formulation. Below, we define the sets and parameters used in the formulation, identify the BI schedule as our decision variable and frame our objective to maximize alignment with stakeholders. We constrain the number of days in-office to match director preferences in constraints 6.5 through 6.15, and force the stakeholder schedules to remain constant using constraints 6.16 through 6.25.

Sets

$$i : 1, ...10 \qquad \text{directors A through J}$$

$$j : 1, ...20 \qquad \text{days}$$

Parameters

$$X_{ij} = \begin{cases} 1 & \text{director } i \text{ assigned on day } j \\ 0 & \end{cases} \tag{6.1}$$

$$V_i \propto \text{director } i\text{'s relationship strength} \tag{6.2}$$

Decision Variables

$$a_i = \begin{cases} 1 & \text{Business Intelligence aligned w/ director } i \\ 0 & \end{cases} \tag{6.3}$$

Formulation

$$\text{maximize} \sum_{i=1}^{10} \sum_{j=1}^{20} a_j \cdot V_i \cdot X_{ij} \tag{6.4}$$

$$\text{s.t.} \sum_{j=1}^{20} a_j = 4 \tag{6.5}$$

$$\sum_{j=1}^{20} X_{1j} = 8 \tag{6.6}$$

$$\sum_{j=1}^{20} X_{2j} = 8 \tag{6.7}$$

$$\sum_{j=1}^{20} X_{3j} = 4 \tag{6.8}$$

$$\sum_{j=1}^{20} X_{4j} = 8 \tag{6.9}$$

$$\sum_{j=1}^{20} X_{5j} = 12 \tag{6.10}$$

$$\sum_{j=1}^{20} X_{6j} = 4 \tag{6.11}$$

$$\sum_{j=1}^{20} X_{7j} = 4 \tag{6.12}$$

$$\sum_{j=1}^{20} X_{8j} = 1 \tag{6.13}$$

$$\sum_{j=1}^{20} X_{9j} = 4 \tag{6.14}$$

$$\sum_{j=1}^{20} X_{10j} = 4 \tag{6.15}$$

$$X_{i=1,j=1}, X_{1,3}, X_{1,6}, X_{1,8}, X_{1,11}, X_{1,13}, X_{1,16}, X_{1,18} = 1 \tag{6.16}$$

$$X_{2,1}, X_{2,4}, X_{2,6}, X_{2,9}, X_{2,11}, X_{2,14}, X_{2,16}, X_{2,19} = 1 \tag{6.17}$$

$$X_{3,3}, X_{3,8}, X_{3,13}, X_{3,18} = 1 \tag{6.18}$$

$$X_{4,3}, X_{4,4}, X_{4,8}, X_{4,9}, X_{4,13}, X_{4,14}, X_{4,18}, X_{4,19} = 1 \tag{6.19}$$

$$X_{5,2}, X_{5,3}, X_{5,4}, X_{5,7}, X_{5,8}, X_{5,9}, X_{5,12}, X_{5,13}, X_{5,14}, X_{5,17}, X_{5,18}, X_{5,19} = 1 \tag{6.20}$$

$$X_{6,3}, X_{6,8}, X_{6,13}, X_{6,18} = 1 \tag{6.21}$$

$$X_{7,3}, X_{7,4}, X_{7,12}, X_{7,13} = 1 \tag{6.22}$$

$$X_{8,13} = 1 \tag{6.23}$$

$$X_{9,3}, X_{9,8}, X_{9,13}, X_{9,18} = 1 \tag{6.24}$$

$$X_{10,3}, X_{10,8}, X_{10,13}, X_{10,18} = 1 \tag{6.25}$$

6.4 Results

Although this formulation is over constrained and only optimizes the BI director's schedule, it still produces measurable improvement in alignment. The resulting maximum alignment value is 2.95 out of 6.49—an increase of 157% from our baseline. Unsurprisingly, the recommended BI director schedule is a weekly cadence on Wednesday.

This Wednesday recommendation is concerning to real estate business leaders—leaders are concerned campuses may exceed capacity on popular days of the week. This highlights the importance of modeling constraints in the system. Our finding is that offices clearly operate with a network effect—the value of the resource increases non-linearly as the right work groups co-locate. As such, if an organization has no concerns around exceeding capacity, the goal should be finding a solution that clusters the correct work groups and densely populates the office on popular days of the week (unsurprisingly Tuesdays, Wednesdays and Thursdays). As we move to a solution at scale in Chapter 7, this is an important finding to carry forward from the proof of concept.

In addition to the initial results, we are able to conduct a sensitivity analysis for the BI director because it is framed as an optimization. We acknowledge that frequency will ultimately be determined by business leaders, but network analytics provides a quantitative measure of what a leader is gaining by increasing the frequency of in-office schedules. Table 6.4 presents the data from a sensitivity analysis varying the number of days the BI employees come into the office per month.

	Baseline	2-Days	4-Days	6-Days	8-Days	10-Days
Days/Month	4	2	4	6	8	10
Alignment Value	0.87	1.59	2.95	3.81	4.63	5.46
% Aligned	13%	24%	45%	59%	71%	84%
% Increase	0%	82%	239%	338%	432%	528%
Rate of Change	N/A	1.82	1.86	1.29	1.22	1.18
% Align/Day	3%	12%	11%	10%	9%	8%
Schedule Rec.	N/A	Wk1:Wed Wk2:Wed	Wk1:Wed Wk2:Wed Wk3:Wed Wk4:Wed	Wk1:Wed/Thu Wk2:Wed Wk3:Wed/Thu Wk4: Wed	Wk1:Wed/Thu Wk2:Mon/Wed Wk3:Wed/Thu Wk4:Mon/Wed	Wk1:Mon/Wed/Thu Wk2:Mon/Wed Wk3:Mon/Wed/Thu Wk4:Mon/Wed

Table 6.3: Results with Sensitivity - BI Director Scheduling

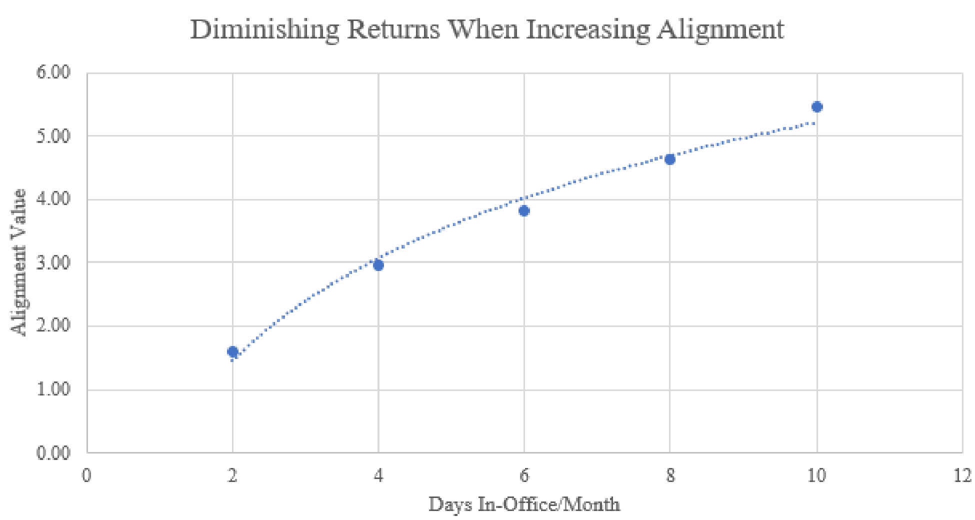


Figure 6-1: Stakeholder Alignment Shows a Point of Diminishing Returns

Quantifying alignment is powerful for directors. The sensitivity analysis helps a leader understand what his/her organization gains by increasing frequency from four days per month to six days per month—in our proof of concept stakeholder alignment increases by 14%. As presented in Chapter 4, human resources is empowering frequency decisions to business leaders. We agree with this approach—leaders should own the operational rhythm of their business unit. Our key point is that by applying network modeling and optimization we can provide business leaders key data to make an informed decision. Finally, and as mentioned previously, the global increase in alignment would be significantly larger if we had access to all data at the headquarters campus.

6.5 Localized Business Value

The business case for realignment is intuitive—better schedules increase in-person collaboration with stakeholders and remove the detrimental aspects of collaborating remotely—a win-win that provides value to the organization and employees. We believe that there is both quantitative and qualitative value to realigning schedules via network analytics.

Starting with the quantitative value, we estimate realignments by anchoring to Gibbs' study "Work from Home Productivity" [9]. In Chapter 2, we outlined the merits of Gibbs' study and highlighted that fully remote work costs 1.4 hours per employee per week. This means, when working remotely, employees work 1.4 hours longer per week to accomplish the same quantity of work as working fully in-person. Gibbs attributes this net negative productivity effect to the challenge of orchestrating collaboration in the remote environment—an effect coined as the work from home (WFH) Effect [9]. In Chapter 4, we confirmed the hypothesis that the subject organization is experiencing the collaboration overload referenced in Gibbs work. As such, we anchor to Gibbs' findings to estimate the business value of realigning the Business Intelligence schedules.

To start, let us outline our assumptions. First, the average number of days in office per week in our subject organization is one day per week. We take a conservative approach and assume this means realignment only captures one-fifth of the extra 1.4 hours worked each week. The logic behind this assumption is that if schedules are perfectly aligned one day a week for in-office work, employees still experience the WFH effect on the remaining four work days. Of course, the embedded assumption is that the 0.28 hours gained by alignment are used to provide value to the business. Next, we estimate the value of time using the average employee salary, benefits, and overhead throughout the typical working year. Finally, we factor in the increase in alignment from 13% to 45%.

$$Value/Day = \frac{1.4\ hours}{week} \cdot \frac{1\ day}{week} = \frac{0.28\ hours}{week} \tag{6.26}$$

$$BI\ Hours/Week = \frac{0.28\ hours}{week} \cdot\ 200\ employees\ \cdot .33\ improvement = 18.48 \tag{6.27}$$

Equation 6.27 converts the .28 hours into a metric for the whole Business Intelligence unit, 200 employees, and scales by the alignment improvement our optimization provided, 33%. Next, we simply apply our estimates for employee value per hour. We estimate that employee average salary, benefits and overhead sum to approximately $150,000 per year per employee.

$$BI\ \$\ Value/Year = \frac{18.48\ hours}{week} \cdot \frac{\$150,000}{2080\ hours/year} = \$1,332\ /week \tag{6.28}$$

Converting this value estimate to a yearly figure yields $88K—meaning, throughout the calendar year, realigning the Business Intelligence organization schedules is worth approximately $88K in work contributions that would have otherwise been spent orchestrating remote collaborations.

We scale the proof of concept data above to estimate the value of realigning schedules across the 51K administrative employees classified as hybrid. Following the same logic as above, and assuming an average alignment improvement of 33%, the enterprise value is **$16.3M** per year. We acknowledge that the various assumptions made in this calculation yield a wide confidence interval—our goal is not to provide a precise valuation, but rather highlight the importance of a phenomena that is being overlooked in many hybrid work designs. Based on our analysis, aligning schedules is a critical piece of a well designed system that enables employees to move seamlessly between remote and in-person work offerings.

In addition to the quantitative value, there is intangible value to improving alignment. Employees are demanding flexibility and the return to office strategy is critical to retain top talent. It is our opinion that misaligned schedules will contribute to worker attrition in the current employment market. As an employee, we can all see the frustration of being told to come to the office on a day when your stakeholders are not also present in the office. If large organizations are going to take the scheduling decision out of the employees hands, they need an analytical method to understand collaboration and align schedules accordingly.

6.6 Sensitivity Analysis and Centrality Optimization

This section present a sensitivity analysis and builds a slightly more sophisticated optimization around the centrality of stakeholders versus relative frequency of interaction.

To demonstrate the flexibility of the optimization tool, we move on to include all 16 BI stakeholders—one stakeholder had not published a schedule and is not included, reducing our meaningful BI stakeholders from 17 to 16. We add constraints for our six additional directors and calculate that the potential total alignment value with all stakeholders is 6.87. Potential total alignment increases from 6.49, or 5.8%, demonstrating that the additional stakeholders represent a marginal amount of BI stakeholder interactions. Nonetheless, our formulation yields slightly different schedule recommendations in order to gain alignment with all BI stakeholders, as seen in Table 6.4.

The full stakeholder optimization provides marginal improvements for a 2 day, 4 day and 8 day frequency schedule. We still see that the 4 day, weekly, Wednesday schedule is optimal based on stakeholder selections and the BI directors preference for 4 days in-office per month. Our key take away is that the stakeholder optimization model is not significantly sensitive—we produce similar results accounting for the top 10 stakeholders versus all top 16.

	Baseline	2-Days	4-Days	6-Days	8-Days	10-Days
Days/Month	4	2	4	6	8	10
Alignment Value	0.92	1.7	3.15	3.81	4.93	5.80
% Aligned	13%	25%	46%	55%	72%	84%
% Increase	0%	90%	242%	314%	436%	530%
Rate of Change	N/A	1.85	1.85	1.21	1.29	1.18
% Align/Day	3%	12%	11%	9%	9%	8%
Schedule Rec.	N/A	Wk1:Wed Wk2:Wed	Wk1:Wed Wk2:Wed Wk3:Wed Wk4:Wed	Wk1:Wed Wk2:Mon/Wed Wk3:Wed/Thu Wk4: Wed	Wk1:Mon/Wed Wk2:Mon/Wed Wk3:Wed/Thu Wk4:Mon/Wed	Wk1:Mon/Wed/Thu Wk2:Mon/Wed Wk3:Mon/Wed/Thu Wk4:Mon/Wed
Relative to Top 10 Model	N/A	+1%	+1%	-4%	+1%	0%

Table 6.4: Sensitivity Analysis—BI Director Scheduling, All 16 Stakeholders, Optimizing on Relative Frequency

It is helpful to visualize the increase in alignment and the rate of change in alignment. Leaders can use visuals in Figure 6-2 and Figure 6-3 to analyze the benefit of adding in-office days. Furthermore, we can pinpoint the point of diminishing returns for adding additional in-office work days.

In the proof of concept data presented in Figure 6-3, the rate of change in alignment levels off at approximately 3-4 days per month. This finding confirms our hypothesis that business units in-office schedules reach a point of diminishing returns—adding in-office days is not likely worth sacrificing focused remote work. For our proof of concept data, we estimate that the point of diminishing returns is 3.6 days per month. The data in Figure 6-3, coupled with frequency data presented in Chapter 4, drives our recommendation that the BI Director should target a frequency of four days in office per month.

The point of diminishing returns will naturally converge with reference to the stakeholders modeled—if the average days in-office per month for our stakeholders in the BI network model was 8 versus 6.1, then we would observe the alignment rate of change start decreasing at approximately five or six days per month. This connects to our recommendation of setting in-office frequency using data. Our perspective is

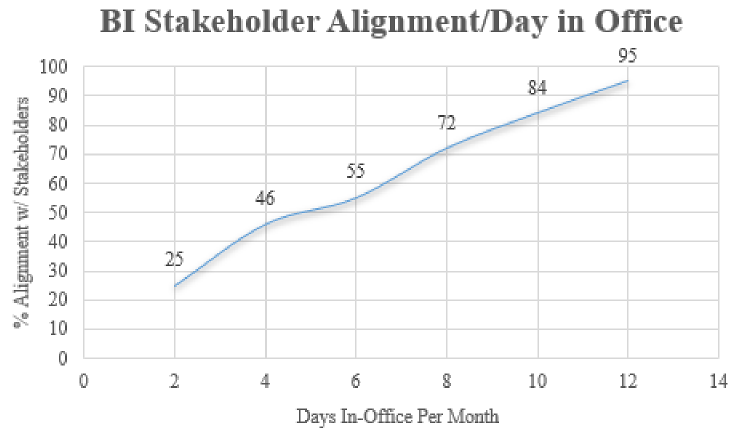


Figure 6-2: BI Stakeholder Alignment Per Day in Office

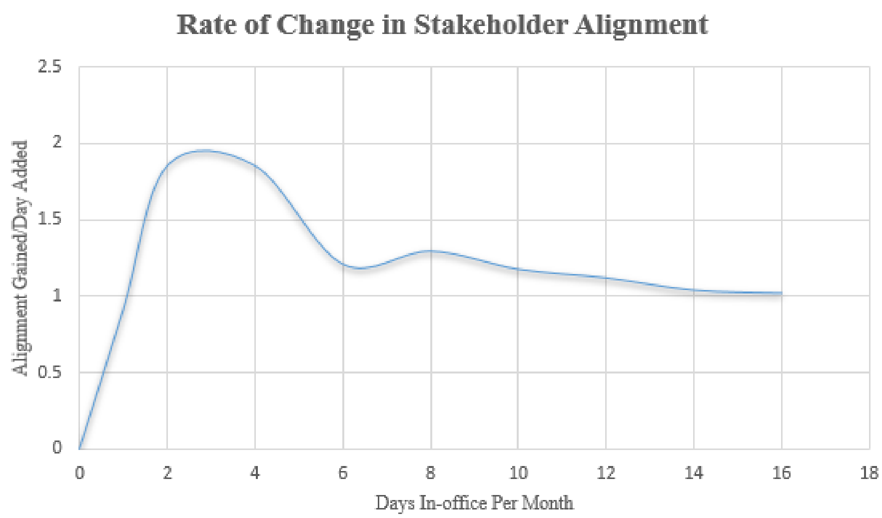


Figure 6-3: Rate of Change - BI Stakeholder Alignment Per Day Added

that organizations need to make top level frequency recommendations using data or employees will simply start coming into the office more and more frequently and/or sporadically to engage stakeholders. This is not the ideal solution for real estate strategy planners.

Recommending frequency using data is the proactive solution. If we embed frequency at the start, leaders can visualize the trade-off displayed in Figure 6-2 and Figure 6-3. Ultimately, a pipeline would provide these visual tools and enable each director to make a business decision around in-office frequency.

6.6.1 Optimizing on Eigen Centrality Versus Relative Frequency

Since we built a network model to provide descriptive inputs to our optimization, we can optimize for any specific network metric. As discussed, calculating the eigen centrality of a node quantifies the notion of importance. We test our optimization pipeline using eigen centrality as V in Equation 6.2. It is interesting to begin by visualizing the stakeholders ordered according to relative frequency versus eigen centrality.

Director	Freq.	Director	Centrality
Director A	24.0%	Director A	1.00
Director B	17.3%	Director B	0.879
Director C	11.9%	Director C	0.335
Director D	9.7%	Director E	0.241
Director E	8.2%	Director G	0.208
Director F	7.8%	Director D	0.207
Director G	7.4%	Director H	0.166
Director H	6.8%	Director F	0.144
Director I	3.5%	Director M	0.139
Director J	3.2%	Director I	0.117
Director K	2.6%	Director N	0.0953
Director L	2.5%	Director O	0.0755
Director M	2.3%	Director K	0.0422
Director N	2.2%	Director J	0.035
Director O	2.1%	Director P	0.0353
Director P	2.0%	Director L	0.0322

Table 6.5: Relative Frequency vs. Eigen Centrality

In Table 6.5, eigen centrality provides a new vantage point on stakeholder prioritization. This is the power of the network—the BI director can align to the most important, not simply most frequent, stakeholders in her network. We believe it is best to prioritize stakeholder alignment according to eigen centrality—a hypothesis we validated using ERGMs in **Chapter 5**. Employees are more likely to share relationships with nodes that have a large eigen centrality. For our proof of concept, we started with a relative frequency interpretation of relationship strength because business leaders naturally understand and connect with this metric. As organizations look to move beyond a proof of concept and put a network modeling tool into production, eigen centrality is the metric to optimize.

	Baseline	2-Days	4-Days	6-Days	8-Days	10-Days
Days/Month	4	2	4	6	8	10
Alignment Value	3.31	4.94	9.51	13.45	17.46	20.47
% Aligned	14%	20%	39%	55%	72%	84%
% Increase	0%	187%	713%	1166%	1626%	1972%
Rate of Change	N/A	1.49	1.93	1.41	1.30	1.17
% Align/Day	3%	10%	10%	9%	9%	8%
Schedule Rec.	N/A	Wk1:Wed Wk2:Wed	Wk1:Wed Wk2:Wed Wk3:Wed Wk4:Wed	Wk1:Mon/Wed Wk2:Mon/Wed Wk3:Wed Wk4: Wed	Wk1:Mon/Wed Wk2:Mon/Wed Wk3:Mon/Wed Wk4:Mon/Wed	Wk1:Mon/Wed/Thu Wk2:Mon/Wed Wk3:Mon/Wed/Thu Wk4:Mon/Wed

Table 6.6: Sensitivity Analysis—BI Director Scheduling, All 16 Stakeholders, Optimizing on Eigen Centrality

The optimization results are presented in Table 6.6. The overarching change to schedule recommendations is a bias towards a Monday/Wednesday recommendation—a reaction to the shift in priorities driven by eigen centrality prioritization.

Similar to the relative frequency sensitivity analysis, we can visualize how the rate of change in alignment changes as we add in-office days. We observe an interesting phenomena here. Shifting to an eigen centrality optimization simulates what would happen as stakeholders increase in-office frequency—we see that the optimal point on Figure 6-4 increases above four days per month. This is because Director E and Director O have a higher prioritization according to eigen centrality and have an above average in-office frequency.

Figure 6-4 supports two key points mentioned throughout the chapter—employee networks are highly coupled systems, and frequency decisions without data may undermine the purpose of hybrid work. Our hypothesis based on this observation is, if important nodes in the network are biased toward higher in-office frequency, the behavior of the entire network will gradually shift to a higher in-office frequency. We tested this hypothesis on the proof of concept data and did not find a meaningful correlation between importance and frequency. Organizations orchestrating hybrid work selections need to test this hypothesis globally. A positive or negative correlation between importance and in-office frequency is concerning—on one end, important ex-

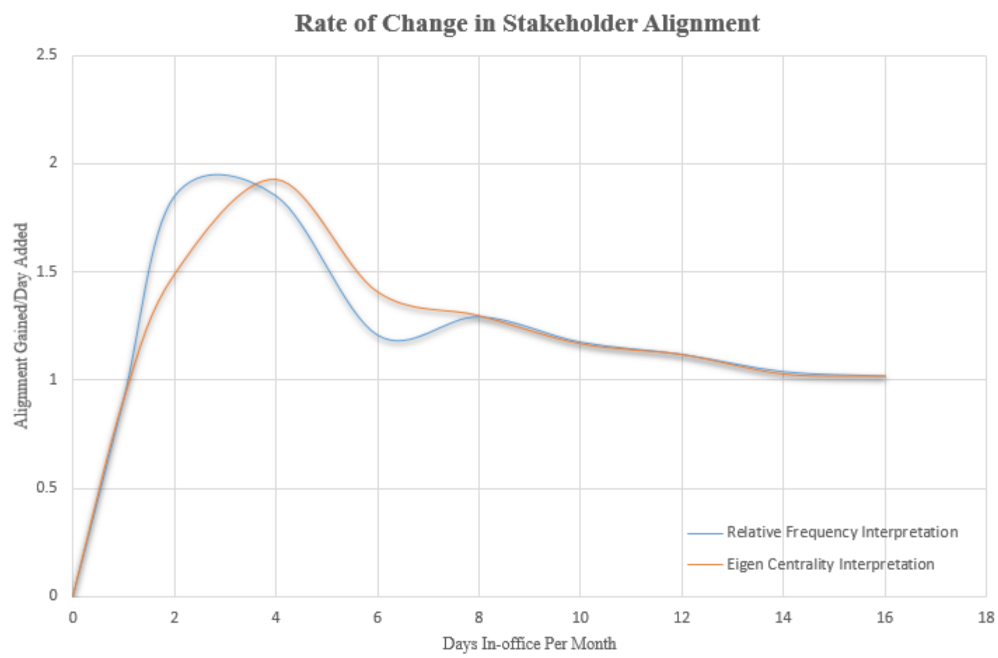


Figure 6-4: Rate of Change - BI Stakeholder Alignment Per Day Added, Eigen Centrality versus Relative Frequency

ecutives are not embodying the hybrid designation, on the other, important nodes are less likely to be in the office helping co-workers solve problems. Our simple proposal to avoid this situation, and help the hybrid system stabilize quickly, is data driven frequency recommendations.

In short, this supports our argument that the ideal hybrid pipeline starts by grounding frequency estimates in data. A data pipeline would then model networks, optimize on eigen centrality and provide visual tools to business leaders. These tools would be used just like other business analytics models to inform decisions around adding/subtracting in-office frequency. And finally, as presented in Chapter 5, organizations can use a variety of network metrics to monitor the evolution and impact of hybrid work.

6.7 Key Discoveries and Recommendations

When our team presented these optimized schedules to the BI director, she made a comment that truly highlighted the need for an analytical solution. She was not surprised by the top 10 stakeholders, as these findings matched her expectations—rather, she was surprised by the fact these stakeholders had changed schedules since the initial submission deadline. The BI director had done a heuristic optimization by talking to her peers and using her time to map schedules. However, this is an interdependent system, and BI stakeholders connected with their collaborators and changed schedules—likely, based on recency basis versus data. This highlights the need for a global optimization.

Furthermore, from a real estate perspective, leadership can not make proactive strategy decisions before schedules are aligned. We have demonstrated in Chapter 4, 5, and 6 that data can drive hybrid frequency, networks and alignment. Organizations need to design a hybrid system that uses this data proactively versus simply waiting to monitor employee behavior. The coupling effect discussed with the BI director illustrates that employee behavior will be very noisy as we transition into a hybrid work environment—leading real estate strategy to delay valuable consolidation decisions.

Chapter 7

A Better Way to Schedule

In this chapter, we design a new system that leverages network properties to cluster work communities at a campus and align schedules. Our key operational recommendations for organizations orchestrating hybrid work are: 1) simplify the number of schedule types, and 2) align in-office days using network relationships versus hierarchical relationships. The new system design starts by grounding frequency recommendations in data and moves into network modeling. An analytics pipeline, as illustrated in Figure 7-1, generates the network at a campus, builds optimal communities using unsupervised clustering, tests significance of these clusters and aggregates similar clusters into three work groups. These clustering takes into account employee in-office frequency to build frequency proportional work communities. The three work groups can be aligned to Tuesday, Wednesday and Thursday; this matches employee bias against Monday/Friday in-office assignments. Ultimately, this pipeline optimizes network effects in the office while balancing capacity across the work week.

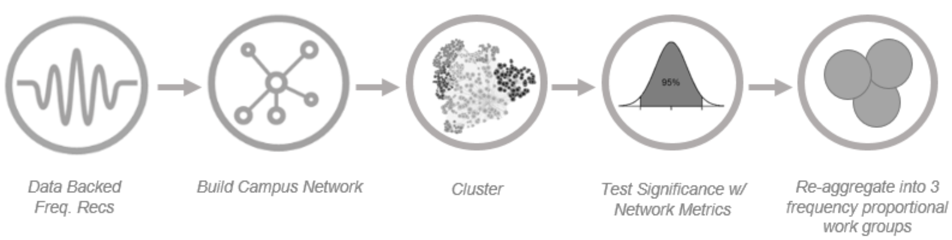


Figure 7-1: Proposed System Design

This pipeline is a simple, scalable method that solves the scheduling challenge for directors. A business leader receives a data backed frequency recommendation for each employee, and each employee is placed into one of three work groups based on network models. One counterpoint to this approach is a director or senior manager may not fall in the same work group as his/her direct reports. Our data in demonstrates Chapter 5 that meaningful work relationships are not necessarily aligned with an organizational chart. Furthermore, if there is a hierarchical relationship, then it will be present in the data—direct leaders of a team will interact and be aligned in the same work group. In sum, this system is rooted in data and will align employees based on work versus strict organizational ties.

This chapter strengthens the argument for modifying the current system in our subject organization. We apply the new system to our proof of concept data and assess the performance. One challenge of unsupervised methods is assessing model performance. Our approach is to first test and quantify on our director network model—this grounds the approach in real data. Finally, we demonstrate how network metrics can be used to validate clustering results.

7.1 Why would we want to do this?

The dichotomy evolving in the post COVID-19 era is the employer's desire to maintain structure versus the employee's demand for flexibility. Although structure and flexibility are naturally antonyms, we believe a data backed scheduling system can deliver structure without violating the employee demands for flexibility. Our hypothesis is that an archaic, hierarchically grounded scheduling system will violate the employee demand for flexibility. And most importantly, hierarchically grounded scheduling will fail to deliver the organizational benefits and employee satisfaction that comes with effective workplace interactions. We encourage all organizations to test this hypothesis via employee behavior in the years to come. Lacking omniscience about what will happen, we use our data from the proof of concept to support this hypothesis and make the case for changing—or at least conducting an A/B experiment on—the current scheduling system.

A powerful data point for why we should change the scheduling system is taking our employee network from the proof of concept and simply color coding each employee by his/her director. This visually corroborates the lack of statistical evidence that employees develop meaningful work relationships via organizational chart ties. Figure 7-2 illustrates the disparity of work groups versus director alignments. In the network each node is an individual employee and the colors correspond to the 17 different directors present in our proof of concept data. As we discussed in Chapter 5, there is no data that shows employees under the same director are more likely to share relationships.

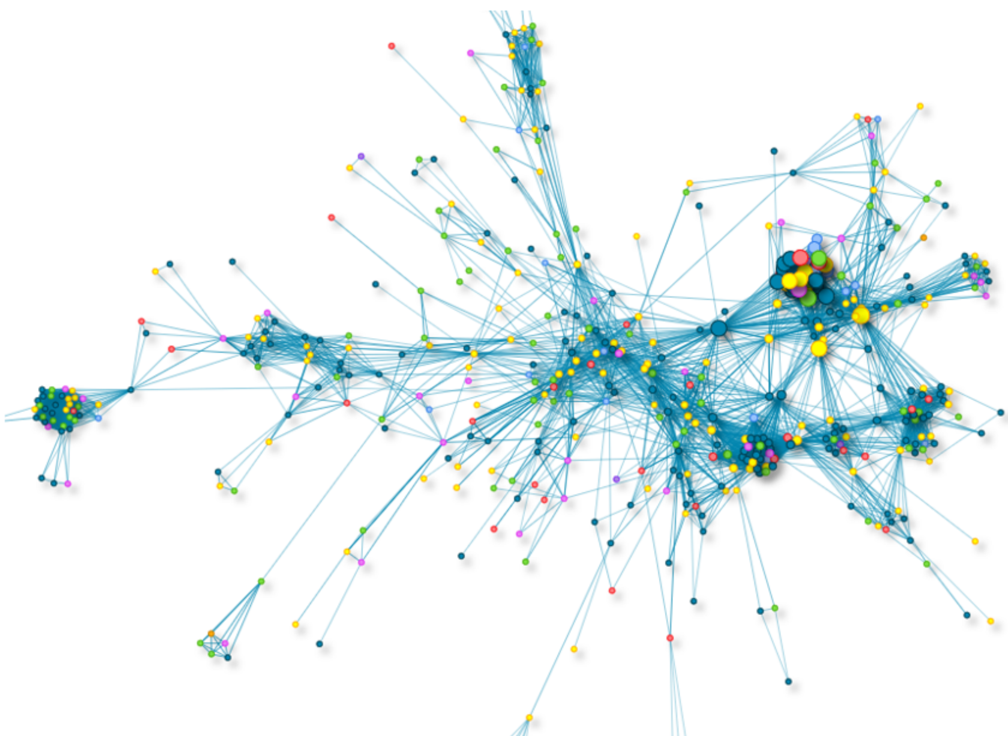


Figure 7-2: BI Proof of Concept—Employee Level Network, Colored by Director

We also analyze schedules by day of the week and create network visualizations to show how many business leaders seem to be defaulting to Tuesday schedules in our sample. In Figure 7-3, we observe that 58% of employees, or 216 out of 369 in the sample, are scheduled on Tuesdays. Beyond our proof of concept sample, our study observes a large scheduling bias towards Tuesday and Wednesday across the subject organization. Figure 7-3 can be contrasted against Figure 7-4 showing employees scheduled on Wednesday—in our proof of concept 58% of employees have a Tuesday schedule versus 28% on Wednesdays. In each of these figures, we observe employees that work closely together but are scheduled on two different days of the week.

Figure 7-3: BI Proof of Concept—Employee Level Network, Scheduled Tuesday Highlighted

Figure 7-4: BI Proof of Concept—Employee Level Network, Scheduled Wednesday Highlighted

As demonstrated in Chapter 4, scheduling is globally biased towards Tuesday/Wednesday. This could be the result of a number of cognitive biases in the decision making process, but our goal is not to pinpoint why this Tuesday/Wednesday overload is occurring. We believe the "MidWeek" effect is due to some combination of being the easiest scheduling solution and a network effect. The point being, if we can accurately group into three work groups, we can balance demand across Tuesday/Wednesday/Thursday and ensure that the right employees are in the office on corresponding days. The potential gains from changing the system are proper work group alignment and balanced system capacity on Tuesday/Wednesday/Thursday, the high demand work days. Furthermore, the organization and employees do not lose structure or flexibility—while some employees may prefer certain days of the week, our data show that the difference between coming in on a Tuesday/Wednesday/Thursday is negligible; the true preference is against Monday/Friday in-office scheduling.

Based on these observations, we develop a scalable solution that uses data to align schedules and balance capacity at a campus. Next, we delve into the technical foundations of clustering communities in our network. With this groundwork set, we test and assess performance based on the proof of concept data used throughout the study.

7.2 Introduction to Community Clustering Algorithms

Intuitively, a community in a network is a cluster of closely connected nodes. More formally, a community is "a subgroup of nodes that has many edges between its members, and a few edges connecting to nonmembers [15]." Detecting work communities is an optimization problem at heart. In short, nodes are assigned to communities based on data in the network, and community quality is iteratively assessed using a performance metric. This section introduces two classes of algorithms—modularity optimization algorithms and spectral algorithms. We discuss the difference between approaches and detail two modularity optimization alogrithms—*FastGreedy* and *Spinglass*—and one spectral alogrithm—*Leading Eigenvector*. This sets a basis for testing community clustering alogrithms on our proof of concept networks.

The first class of community detection methods are the modularity optimization alogrithms. Modularity is a metric used to measure the quality of an assigned community in the network [15]. First introduced by Newman and Girvan in 2004, modularity is a score that relies on randomly generated networks to baseline community structure in the network under study. Modularity is best described in steps. The first step is calculating the proportion of edges that link two nodes in the community, this can be thought of as, p_{within}, or the proportion of edges that fall in the community [15]. The next step is generating random networks of the same degree as the network under study. This quantifies the fraction of connections expected within the assigned communities due to random chance. This can be described as, p_{random}, or the proportion of nodes that randomly fall in the same community. Finally, modularity is equal to the difference between these two proportions, $modularity = p_{within} - p_{random}$. Modularity quantifies the quality of community clusters on a scale between -1 and 1. A modularity of positive 1 means nodes only interact with nodes in the same community; a modularity of 0 means the community structure is not significantly different than a randomly generated network. Broadly speaking, a modularity of above 0.3 means that the community structure in the network is significant and contains valuable insights into how nodes form relationships.

The simplest modularity based optimization is the *FastGreedy* algorithm. *FastGreedy* is an agglomerative—or merge driven—method. Each node starts in its own community and *FastGreedy* optimizes modularity by merging pairs of communities that produce the greatest increase in modularity.

Algorithm 4 Fast Greedy Community Clustering

```
procedure FAST GREEDY CLUSTER(NetworkGraph)
    Begin with every node in its own community
    Calculate the change in modularity for each pair of communities
    Merge the two communities with the greatest increase in modularity
    Repeat steps 3 and 4 until all nodes are in one large community
    Select community assignments with highest modularity
```

The *FastGreedy* algorithm was first introduced in 2004 by Newman and follows a hierarchical approach [22]. As shown in Algorithm 5, the modularity is calculated at each merge building a hierarchy. Each level in this hierarchy has a modularity—the level with the highest modularity is selected as the optimal community assignment. A number of studies have assessed algorithm performance. On balance, *FastGreedy* underestimates the number of communities and performs worse as the size of the network increases, but it delivers optimal clustering speed [35].

The next modularity based algorithm is the *Spinglass community* algorithm. *Spinglass* was first introduced in 2006 by Reichardt and Bornholdt, and translates a modeling technique from the field of statistical mechanics [15]. In short, the model is a system of spins with the goal of optimizing an energy function. In community detection, the spins assign group labels to nodes, and the energy of a spin is used to assess quality of community assignments. The energy function is formulated as a minimization—rewarding edges within communities and penalizing edges between communities. A simplified interpretation of the objective function is,

$$-\sum \textit{internal edges} + \sum \textit{internal nonedges} + \sum \textit{external edges} - \sum \textit{external nonedges}$$

This objective is a simplified form of the objective Hoffman et al. presented in their 2017 publication [15]. The negative terms act as a reward since this is a minimization—we want more edges internal to the community assignment and less edges external to the community. Simultaneously, the positive terms penalize edges that do not appear in the community and external edges that connect to other communities [31]. Broadly speaking, the *Spinglass community* tends to overestimate the number of actual communities but performs well on large networks up to fifty thousand nodes [35]. Practically, a variety of packages have a *Spinglass community* function that executes this optimization formulation.

The *Leading Eigenvector* algorithm is the third clustering approach we test in this study. *Leading Eigenvector* uses the concept of modularity but incorporates traditional spectral clustering concepts [15]. Spectral methods naturally align with networks because they rely on the concept of affinity—or how points are interrelated—versus the absolute location of a point to determine cluster alignment [20]. For community clustering, the modularity matrix is used in place of the Laplacian matrix traditionally used in spectral clustering [15]. Newman introduced *Leading Eigenvector* in 2006 by reframing modularity optimization in terms of the eigen values of the modularity matrix [23].

In short, the method works by calculating the modularity matrix, computing the eigenvector for the largest positive eigenvalue of this modularity matrix, and separating nodes based on the sign of the eigenvector [24]. This process is divisive because it continues to divide each of the communities until there is no improvement in overall modularity [15]. Again, for practical application on networks, R contains a robust leading eigenvector package that implements the algorithm on a network object. *Leading Eigenvector* clustering generally produces a marginally higher modularity than other methods but is computationally expensive [35].

An important step of any clustering procedures is assessing the results. In addition to cluster modularity—the measure our routines are optimizing—we can use a network metric to assess cluster goodness of fit. When working with the director level network, we have a heuristic mapping of the ground truth. We could measure how many of the top 17 stakeholders are clustered with our BI director and calculate a standard F1 score. As we scale, since our data are unlabeled—e.g. we do not know the ground truth—the two measures we use to assess the goodness of cluster fit is the external-internal index by cluster and an ERGM reponse test by cluster. As discussed in Chapter 5, EI index ranges from -1 to +1, with -1 signaling that every relationship is an in-group relationship and +1 signaling that every relationship is an external group relationship. Thus, when calculating the EI index on clusters, the best possible result is a -1 for the clustered network. Since this provides a relative measure, we use a simulation style approach to validate results. The approach is similar to a statistical t-test—we generate classical random networks of the same order and size as our network under study, apply the clustering routine to the randomly generated network, and visualize our results [19]. The distribution of number of community in our random graphs allows us to assess whether the clusters in our network under study are meaningful. We expect to observe that the number of communities clusters identified in our network under study falls on either tail of our randomly generated distribution.

This concludes our introduction to community clustering techniques. Next, we apply and assess these methods on our director level network and employee level network. Initially, the goal is to leverage our heuristic knowledge about the director level network to validate clustering performance—we devise this approach to act as a semi-labelled test set. We use this simple network to test and describe our assessment criteria. Then, we scale by assessing the employee level network clusters using these network metrics and simulation techniques.

7.3 Clustering and Assessing Community Structure

This section applies community detection algorithms to the director level network. If our subject organization is tethered to the current scheduling process where business unit directors set schedules for an entire organization, our finding is that community clustering at the director level will provide a **74%** improvement over the current process. We believe that simple recommendations to directors would greatly improve alignment outcomes—a simple recommendation to the BI director would state who falls in her community based on the data. She could then communicate with the right directors and align schedules. While this would not produce optimal work group alignments, it would be a step in the right direction. This section provides details on how to cluster the director level network and our method to assess goodness of cluster assignment.

Each clustering routine can be implemented in R or Python using the appropriate package. To describe our methods, we detail the process using the *Spinglass* algorithm and provide a table comparing algorithms in the results section. We apply the *Spinglass* algorithm to the director level network and color code by community to produce the network seen in Figure 7-5. The *Spinglass* algorithm determines the optimal number of clusters and encodes nodes according to the optimization formulation.

A logical starting point after executing a clustering routine is to aggregate the descriptive statistics from the entire director network and subsequently calculate similar

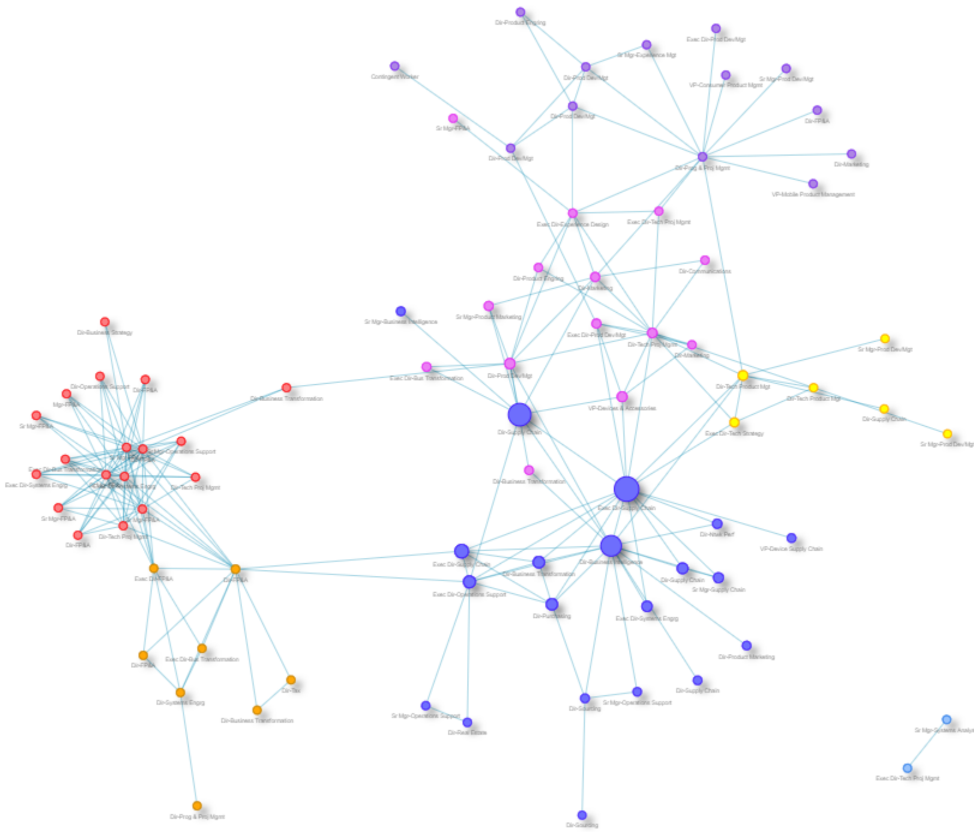


Figure 7-5: Director Network Clustered

metrics within cluster assignments. The global descriptive statistics after implementing the *Spinglass* algorithm are detailed below in Table 7.1.

Nodes	Modularity	Density	Transitivity	Clusters	Avg Degree	Avg Centrality
83	0.592	0.060	0.410	8	4.9	0.06

Table 7.1: Director Level Network—Descriptive Network Metrics, *Spinglass*

The first point to notice when assessing the quality of clustering is the modularity measure. After clustering using the *Spinglass* algorithm, the director network displays a modularity of 0.592 which is well above the generally accepted benchmark of 0.30 for "good" community assignments [19]. Of course, our goal is to demonstrate how to triangulate goodness of community alignment more rigorously. The additional descriptive metrics in Table 7.1 show the basic structure of the network under study and lay the ground work as a representative structure when simulating random graphs to assess cluster significance.

The next logical step in analyzing community clusters is to build granular network metrics by cluster. We accomplish this by inducing sub-graphs from our global graph. Then we can calculate a variety of network metrics for each community cluster. Table 7.2 presents network metrics for each of the seven community clusters identified by the *Spinglass* algorithm. We expect to see density higher within clusters when compared to the global network—this means our clusters are well connected internally, or "cohesive." Analyzing the data in Table 7.2 leads us to discount cluster 7 and 8, these are outlier dyads disconnected from the rest of the graph. Looking within the 6 primary clusters, the density of each cluster is greater than the global network density in Table 7.1. This is a positive indicator: the communities have meaning because nodes are more densely connected in the community versus the global network.

Cluster	**Nodes**	**Avg Degree**	**Avg Centrality**	**Density**	**Transitivity**
Cluster 1	18	7.6	0.001	0.45	0.56
Cluster 2	6	2.0	0.028	0.40	0.30
Cluster 3	13	2.5	0.001	0.21	0.21
Cluster 4	20	3.9	0.22	0.21	0.41
Cluster 5	8	2.0	0.028	0.40	0.3
Cluster 6	14	3.4	0.047	0.43	0.26
Cluster 7	2	1	0.0	1	N/A
Cluster 8	2	1	0.0	1	N/A

Table 7.2: Cluster Analysis—Director Level Network

The second step in assessing community significance is simulating similar random networks and clustering these randomly generated networks. Clustering results are then compared with the network under study. The simulation algorithm is detailed in Algorithm 4, it is adapted from [19].

The distribution of results is plotted and helps assess the significance of clustering in the network under study. In the director network clustering results, we choose to exclude the two unconnected dyads and posit that there are six distinct communities in the network. The distribution of communities in simulated graphs displayed in Figure 7-6 provides a reference. Similar to a t-test, if the actual number of communities in the network under study falls on either tail of the distribution, the null hypothesis

Algorithm 5 Simulating Randomly Generated Networks and Clustering

```
procedure SimulateRandomGraphs(network)
    Determine number of trials
    In each trial,
    Build Random Classical Graph, each edge has equal probability of developing
    Build Random Graph with same degree characteristics as network under study
    Cluster communities in each type of graph and log data
```

is rejected. In this case, the null hypothesis is that no meaningful community structure exists in the network under study. Analyzing where six communities falls on the distribution in Figure 7-6 provides another positive indicator that the communities are meaningful and not due to random chance.

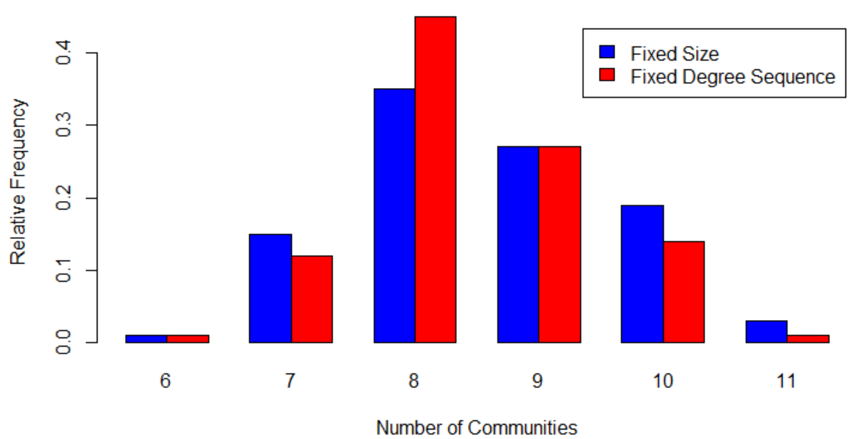


Figure 7-6: Distribution of Number of Communities in Simulations

Now that the significance of community clusters has been baselined against random networks, we can use the external-internal index and an ERGM model to confirm that the clustering aligns with relationships in the network. When aggregrated by cluster, the director network EI index is **-0.649**—meaning the majority of relationships occur internally in clusters. This is the third positive indicator that our community clusters capture the right relationships.

The final tool is the ERGM. After implementing a clustering routine, we assign "cluster" as a nodal attribute and build an EGRM model to test the significance of clusters. The model parameters are presented in Table 7.3, and it is clear that cluster assignment is significant. As discussed earlier with ERGMs, the model coefficients are interpreted using log-odds. After converting the coefficients, we can say that there is a 95% probability a relationship exists between two nodes in the same cluster.

Nodal Match/Covariate	Coeff. Est.	Std. Error	Z-Value	P-Value
Edges	-4.1441	0.2520	-16.45	.0001***
NodeMatch(Cluster)	3.0982	0.2991	10.36	.0001***
Null Deviance: 1698.2				
Residual Deviance: 394.1				
AIC: 398.1 BIC: 408.3				

Table 7.3: ERGM 6 - Director Network Match on Cluster Assignment

Although we have demonstrated throughout this study that assigning schedules at the director level is not optimal, if the subject organization is set on continuing this practice, then a significant improvement to the process would be clustering at the director level. Using our proof of concept sample data, the *Spinglass* algorithm assigns 14 out of 17 BI stakeholders to cluster 4. Although this example only serves as a semi-labelled test set—e.g., we know that the groups work together but do not have a full picture of each stakeholders extended network—our finding is that unsupervised clustering can identify meaningful work groups at scale. This example improves alignment from the 13% baseline to 82% alignment after clustering—a stakeholder alignment increase of over 6 times.

The methods presented in this section are used throughout the remainder of this chapter. Before extending these methods to the more complex employee level network, we compare performance of our clustering algorithms on the director network. Following this exploration of clustering methods, we apply clustering to the employee network and detail our improved system design at scale.

7.4 Clustering Analysis—Director Network

This section focuses on applying the three different clustering methods and assessing performance. We choose to investigate clustering algorithms because there is limited research around which community clustering algorithm is best suited for analyzing work communities—much of the literature focuses on clustering social networks. We follow the methods detailed in Section 7.4 and report results for three clustering algorithms—*FastGreedy*, *Spinglass* and *Leading Eigenvector*. The clustering results are shown in Table 7.4.

Algorithm	**Clusters**	**# of Clusters 95% CI**	**Modularity**	**Density**	**EI Index**	**ERGM Coef.**	**ERGM P-Value**
FastGreedy	5	7.2 - 7.3	0.605	0.33	-0.698	2.88	.0001***
Spinglass	6	7.3 - 7.4	0.592	0.35	-0.649	3.09	.0001***
Leading Eigenvector	14	7.4 - 7.9	0.554	0.34	-0.512	2.75	.0001***

Table 7.4: Algorithm Comparison—Director Level Network

For clarity, we detail the meaning of key columns in Table 7.4 and highlight findings. Column 2—clusters, specifies how many clusters the algorithm identified in the director network. This can be compared with column 3—# of clusters 95% CI, or the 95% confidence interval of clusters identified in our simulation significance test. Columns 7 and 8—ERGM coef. and ERGM p-walue—present the parameters of an ERGM matching nodes by cluster. The coefficient can be used to understand how likely it is that nodes in the same cluster have relationships, and the p-value demonstrates significance of this finding.

On balance, we observe that community structure is present in our director network. The first indicator supporting this claim is the quantity of clusters identified by each algorithm. All three algorithms produce a cluster quantity that falls outside of the 95% confidence interval for clusters in a randomly simulated network. This means that the communities are not due to random chance; there are exogenous factors at play. Second, the modularity produced by each algorithm is greater than 0.30, the generally accepted threshold for good community structure. Third, the average cluster density is approximately 8 times greater than the overall director network density—signifying dense interaction within clusters. Fourth, the EI index aggregated by cluster is less than -0.5. This means the majority of the relationships occur within clusters versus between clusters. Finally, the coefficients of the ERGM matching by cluster, are positive and strongly significant—a key indicator that relationships are much more likely within cluster assignments.

Between the algorithms, we observe three notable differences in performance. First, the *Leading Eigenvector* algorithm appears to over estimate the number of communities present in the network. This results in lower performance across all goodness of fit indicators and confirms a similar observation made using social networks in [35]. Second, the *Spinglass* algorithm provides the largest ERGM coefficient but still lags behind the *FastGreedy* algorithm on the remaining cluster goodness metrics. This comparative analysis drives our finding that the *FastGreedy* and the *Spinglass* algorithm are effective on our proof of concept data, and are relatively efficient as organizations look to build a clustering pipeline to schedule work groups. Next, we extend this analysis methodology to the larger employee level proof of concept network.

7.5 Communities in the Employee Network

The next step towards scaling is testing clustering on a more complex network. In this section we use the *FastGreedy*, *Spinglass* and *Leading Eigenvector* algorithms to cluster the larger employee level network from our proof of concept. We follow the same process to assess community significance, test our hypothesises about algorithm performance at scale, and set the stage for a simplified and scalable scheduling pipeline.

The natural starting point is, again, the descriptive networks metrics for the network under study. The employee level network is approximately seven times larger than the director network and displays a higher transitivity, as seen in Table 7.1.

Nodes	Density	Transitivity	Avg Degree	Avg Centrality
528	0.031	0.714	16.1	0.01

Table 7.5: Employee Level Network—Descriptive Network Metrics Pre-Clustering

To get a sense of the clustering results after applying the *FastGreedy* algorithm, Table 7.6 details key network metrics in each recommended cluster.

Cluster	**Nodes**	**Avg Degree**	**Avg Centrality**	**Density**	**Transitivity**
Cluster 1	135	13.2	0.005	0.099	0.54
Cluster 2	89	19.4	0.389	0.221	0.84
Cluster 3	28	10.9	0.001	0.405	0.81
Cluster 4	39	17.9	0.164	0.47	0.98
Cluster 5	41	10.73	0.001	0.268	0.68
Cluster 6	9	4.7	0.001	0.583	0.84
Cluster 7	14	12.4	0.012	0.96	0.97
Cluster 8	10	4.0	0.002	0.44	0.72
Cluster 9	59	16.9	0.007	0.29	0.80
Cluster 10	11	7.3	0.001	0.73	0.91
Cluster 11	7	5.7	0.0	0.73	0.91
Cluster 12	6	5.0	0.0	1.0	1.0
Cluster 13	37	27.4	0.0	0.761	0.94
Cluster 14	6	2.0	0.0	0.93	0.92
Cluster 15	5	2.0	0.0	0.50	0.43
Cluster 16	4	3.0	0.0	1.0	1.0
Cluster 17	4	2.5	0.0	0.83	0.75
Cluster 18	3	2.0	0.0	1.0	1.0
Cluster 19	3	2.0	0.0	1.0	1.0
Cluster 20	3	2.0	0.0	1.0	1.0
Cluster 21	3	2.0	0.0	1.0	1.0
Cluster 22	2	1.0	0.0	1.0	N/A
Cluster 23	2	1.0	0.0	1.0	N/A
Cluster 24	2	1.0	0.0	1.0	N/A
Cluster 25	2	1.0	0.0	1.0	N/A
Cluster 26	1	0.0	0.0	N/A	N/A
Cluster 27	1	0.0	0.0	N/A	N/A
Cluster 28	1	0.0	0.0	N/A	N/A
Cluster 29	1	0.0	0.0	N/A	N/A

Table 7.6: Cluster Analysis—Employee Level Network, *FastGreedy* Clusters

The first observation is that many clustering routines will naturally group unconnected nodes into single communities. This phenomena is more common in our data because we only have access to the Business Intelligence data—with a full data set, we believe this would not occur frequently. Table 7.6 details all 29 clusters, but eliminating these "disconnected" communities leaves 15 meaningful clusters in the network. Within these 15, we find remarkably high transitivity and density. This is our first sign that community clusters are indicative of how groups interact.

The next step of our method is simulation—this provides a reference for how communities randomly develop in networks. Clustering results from 1000 simulated trials are presented in Figure 7-7. The simulations highlight that our resulting 15 communities are clearly significant when compared to the random reference distribution.

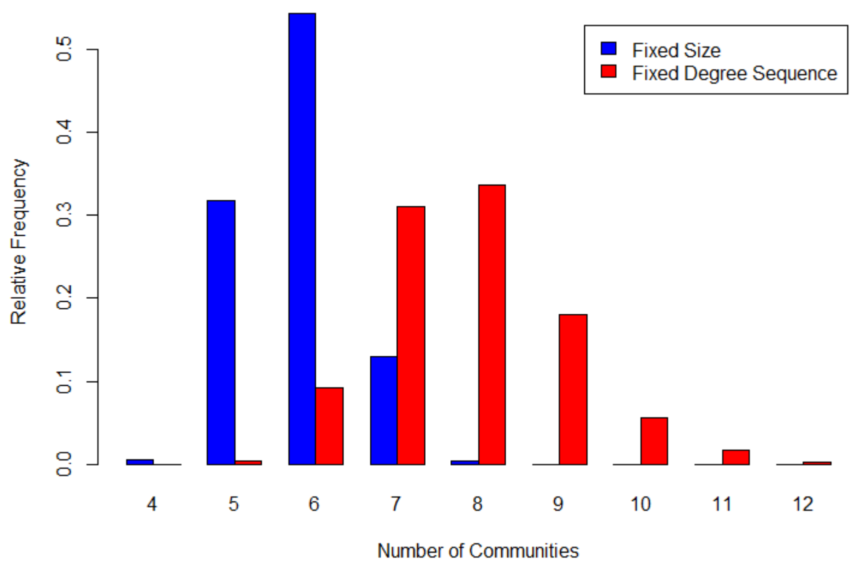


Figure 7-7: Distribution of Number of Communities in Simulations

After analyzing details of the *FastGreedy* algorithm, we test our performance hypothesis—our goal is identify the best clustering method for a scaled scheduling pipeline. We apply our five goodness of fit metrics—simulation clusters, modularity, average cluster density, EI index on cluster, and an ERGM matching nodes by cluster. The results are detailed in Table 7.7.

Algorithm	**Clusters**	**Sim. 95% CI**	**Modularity**	**Density**	**EI Index**	**ERGM Coef.**	**ERGM P-Value**
FastGreedy	29	6.8 - 6.9	0.715	0.55	-0.759	4.11	.0001***
Spinglass	35	7.5 - 7.8	0.765	0.75	-0.772	4.83	.0001***
Leading Eigenvector	32	7.4 - 7.9	0.723	0.69	-0.718	4.57	.0001***

Table 7.7: Algorithm Comparison—Employee Level Network

These results confirm that *Spinglass* is the optimal clustering method as our networks scale in size and complexity. For ease of interpretation, the EI index can be translated to the percent of internal relationships that the cluster assignments capture. In the case of the *Spinglass* algorithm, 88.6% of work relationships fall within cluster assignments. Based on the effectiveness of clustering in these initial trials, we next focus on how to structure a data pipeline that leverages clustering to align schedules.

7.6 The Case for a Redesigned Scheduling Pipeline

Up until this point, we have presented a variety of data and analyses highlighting design limitations in the current scheduling process. Throughout, we have highlighted simple changes that the subject organization can implement to improve results without a complete overhaul of the scheduling system. Now, we present a newly designed process that leverages data and removes the scheduling burden from executives. The overarching finding of this study is that the scheduling system is too complex and forces executives to make decisions without appropriate data.

We propose a simple, five step system that matches appropriate work groups while balancing office capacity across high demand weekdays. Most importantly, this is a scaleable and unsupervised pipeline that can be trained quarterly to match the dynamic nature of knowledge work. We present this process at the employee level because we find this will deliver optimal results: however if the subject organization wants to continue scheduling at the director, level this process can still be implemented. All of the data, networks and clustering would simply be executed at the director level.

The high level process is:

1. Use lagging quarter meeting data to recommend in-office frequency at senior manager level.

2. Senior managers select schedule type—monthly or weekly.

3. Build campus network using collaboration data from previous quarter, assign frequency as a nodal attribute.

4. Cluster campus network into work communities.

5. Group communities to build three proportional "communities of communities."

Throughout this section, we detail each step of the process using the proof of concept employee network. But before delving into the process, we apply three key simplifications upfront: 1) eliminate the bi-weekly schedule option, 2) eliminate the custom schedule option, and 3) only offer Tuesday/Wednesday/Thursday as in-office options.

The bi-weekly option is a key cause of misalignment between work groups. First, this selection is confusing to employees and real estate capacity planners alike. Not every month of the year has four weeks, causing the bi-weekly rhythm to change frequently. Even the booking system implemented by the subject organization does not accurately implement a bi-weekly schedule when an employee selects the bi-weekly option. Our finding is that bi-weekly scheduling undermines the stated objective of setting in-office schedules—predictability. Second, there is much research around the impact of choice architecture on system design; Thaler summarizes this in his 2008 book, Nudge, which shows the prevailing finding across studies is simplified default options lead to better global outcomes [30]. We embrace this logic and encourage the subject organization to present four in-office options to hybrid teams: 1 day per/month, 1 day/week, 2 days/week, and 3 days/week. The default selection for each team should be based on the Team Collaboration Index and day of the week selection limited to Tuesday/Wednesday/Thursday.

Similar to the bi-weekly schedule, the custom option also undermines the stated goal of predictability. While our study did discover higher adherence to custom schedules, we believe the value is isolated internally to that team. The custom patterns submitted are not simple or intuitive, limiting collaborators from aligning in-office days. Our findings in Chapter 5 and 6 drive this logic—employee networks are highly coupled systems,and thus custom schedules will make leaders reactionary.

Finally, our data showed a large bias towards Tuesday/Wednesday/Thursday schedules. To simplify real estate operations, organizations should acknowledge workforce preferences and close offices accordingly. In our subject organization this would mean closing the office on Mondays and Fridays. Hybrid organizations can test capacity concerns by analyzing the average days in-office per week and using Monte Carlo simulations as shown in Chapter 4. In our subject organization, the average number of scheduled in-office days was approximately one day per week. This data point was coupled with simulation to verify that a Tuesday/Wednesday/Thursday model would not present capacity concerns. Furthermore, workplaces provide value via network effects. By limiting the number of week days the office is open, we can more densely populate the correct work groups to enable the desired collaboration.

With these upfront recommendations clarified, we move into each step of our recommended scheduling pipeline.

7.6.1 Step 1: Frequency Recommendation

Our first step in the new scheduling process is anchoring in-office frequency recommendations to collaboration data at the senior manager level. The subject organization does not have a data backed reason as to why schedules are being set by directors—our hypothesis is that setting schedules at the director echelon was an attempt to maintain control and/or simplify the process. This section discusses why organizations should empower first line leaders to set schedules and anchor decisions using collaboration data.

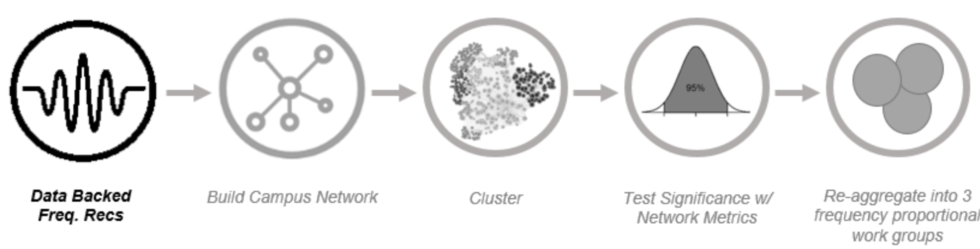


Figure 7-8: Five Step Process

Regardless of why the decision was made to set schedules at the director level in our subject organization, we encourage leaders to rethink scheduling assumptions in light of the data in this study. First, prevailing research in Chapter 2 highlights the need for autonomy in hybrid work designs—this means empowering the lowest level of leaders in an organization to set schedules. Second, Chapter 5 and 6 proves that knowledge work does not unfold neatly along lines on an organizational chart—directors simply have too many subordinates engaging with too many colleagues. Finally, rooting frequency recommendations in data eliminates employer concerns around empowering decisions below the director/executive level. An automated data pipeline can scale recommendations to the senior manager population and properly anchor junior leader's frequency decisions so teams are not coming into the office too few or too many times a week.

The frequency calculation is simple. For each senior manager's team, calculate the average time spent collaborating in groups of three or more per week with employees assigned to the same campus. Then calculate the average meeting length per engagement on that team. These two components combine to form a "team collaboration index," similar to the calculation in Chapter 4:

$$Collaboration\ Index = \frac{Collaboration\ Hrs/Week\ * Average\ Meeting\ Duration}{Average\ Total\ Work\ Hrs/Week} \tag{7.1}$$

The team collaboration index is simply the ratio of amount of work hours with potential in-person collaboration versus total available work hours. The index is translated to a frequency recommendation using a scale similar to the one presented below in Table 7.8—business leaders could adjust the sensitivity of this scale.

Collaboration Index (CI)	**Estimated Collaboration**	**Recommended Frequency**
0.00 < CI < 0.10	0 - 4 Hours/Week	1 Day/Month
0.10 < CI < 0.30	4 - 12 Hours/Week	1 Day/Week
0.30 < CI < 0.50	12 - 20 Hours/Week	2 Days/Week
0.50 < CI < 1.0	20 - 40 Hours/Week	3 Days/Week

Table 7.8: Collaboration Index Translation

Our finding is that once we isolate employees by campus, the frequency of in-person interaction is less than leaders estimate using intuition alone. We saw this phenomena is Chapter 4 of this study. In our Business Intelligence unit, we found that employees at the HQ campus have on average 16.8 hours of potential in-person collaborations and an average meeting duration of 45 minutes. The calculations for our sample population are as follows:

$$Collaboration\ Index = \frac{16.8\ Hrs/Week\ \cdot 0.75\ Hrs}{40\ Hrs/Week} = 0.32 \tag{7.2}$$

Finally, the team collaboration index is converted to a frequency recommendation to enable the leader's decision. In this case, the algorithm recommends that the team be scheduled in-office twice a week based on historical data from the previous quarter.

7.6.2 Senior Manager Selects Frequency

We envision the recommendation algorithm driving the default selection, but not removing the business leader from the loop. A simple data dashboard would display the collaboration statistics for the leader's team and provide the data required to make an informed decision. Each subordinate's frequency would default based on the team collaboration index from the previous quarter. The senior manager would manually adjust selections to vary from the default.

Again, we recommend four simple options: 1 day/month, 1 day/week, 2 days/week, 3 days/week. While this may seem too streamlined, we found that over complicating the scheduling options creates more opportunity for misalignment. If leaders believe they fall in between these frequency categories, our solution is flex reservations; our subject organization, and most large organizations without capacity concerns, are not going to prevent a team from coming in one additional time per month if leadership thinks it is necessary.

7.6.3 Steps 2, 3 and 4: Build Campus Network, Cluster and Test

The next three steps in the process have been discussed in detail—we use employee collaboration data to build the campus network, cluster employees into work communities and test the clusters.

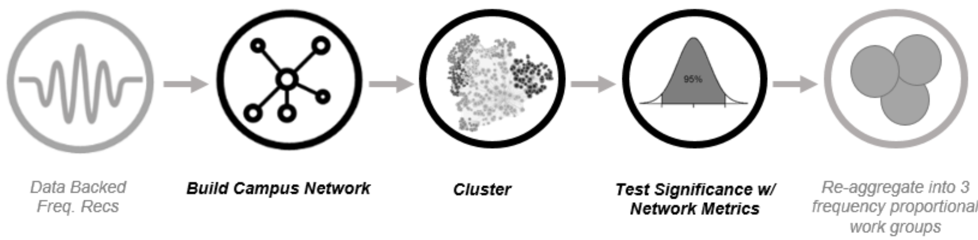


Figure 7-9: Five Step Process

We apply the *Spinglass community* algorithm since it demonstrated optimal performance on larger networks and delivered the highest modularity of the three clustering methods assessed. The resulting clusters are visualized in Figure 7-10.

Figure 7-10: BI Employee Network Clustered—Color Coded by Cluster Assignment

The clustering results after applying the *Spinglass community* algorithm are presented again in Table 7.9. Initially, as we apply this method at a large site, an analytics team can interpret the significance of clustering.

Algorithm	**Clusters**	**Sim. 95% CI**	**Modularity**	**Density**	**EI Index**	**ERGM Coef.**	**ERGM P-Value**
Spinglass	35	7.5 - 7.8	0.765	0.75	-0.772	4.83	.0001***

Table 7.9: Employee Network Clustering Results

To scale across many sites and execute on a quarterly basis, we recommend setting thresholds that would prevent poor clusters from being recommended. The two intuitive thresholds are modularity and EI index. Our hypothesis is that clustering results with a modularity less than 0.3 will not increase in employee alignment and should not be implemented. Similarly, a EI index from 0.0 to 1.0 is an indicator that the clusters are not meaningful—this indicates relationships are more prevalent outside of the cluster assignments versus internally, a sign of poor cluster assignments. Based on the success of our proof of concept, we believe that large campuses will display meaningful clustering results—these thresholds would function as an anomaly indicator for the few campuses that do not present meaningful clustering results.

Once the campus network model is built and clustered, a frequency attribute needs to be added for each employee to ensure that the final work group optimization model can balance capacity across Tuesday/Wednesday/Thursday. A practical way to accomplish this is encoding the monthly frequency—so, for an employee assigned 3 days/week, monthly frequency is 12 times per month. This transitions into step 5 of the pipeline and embeds an attribute into the network that allows organizations to balance capacity without sacrificing work group alignment.

7.6.4 Step 5: Build Three Work Communities

In an effort to design a simple system, the final step is to aggregate the granular cluster assignments into three work communities. In our subject organization, the three work communities align with the recommendation to open offices on Tuesday/Wednesday/Thursday—for other organizations it may make sense to have

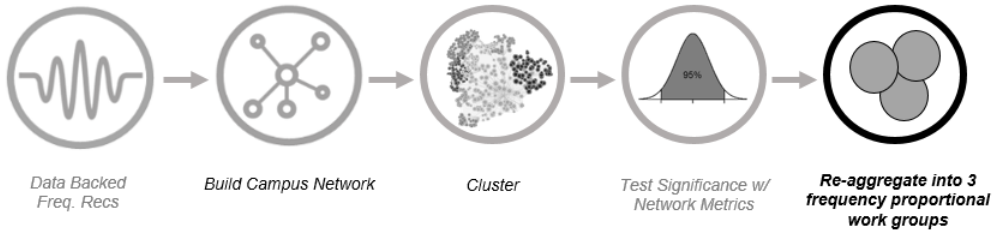


Figure 7-11: Step Five

four or five work communities. This section formulates an optimization that re-clusters granular communities into three, frequency proportional work groups. Each employee's work community assignment is coupled with his/her frequency to produce a day of the week assignment—we present a simple mapping at the close of this section. As the conclusion of this pipeline, each senior manager receives an assignment for employees such as "Employee A—Work Community #1: Weekly on Tuesdays."

In our first attempt to build three macro work communities, we tryto simply use the *Spinglass community* algorithm and set the objective to find three large clusters—unfortunately, this method only optimizes for modularity and does not take into account in-office frequency. The three resulting work communities are not balanced across the week or month. Based on this discovery, we devise a method which starts by using the *Spinglass* algorithm to build granular clusters and then finds three connected macro communities—each made up of a combination of granular clusters. In this approach, in-office frequency can be balanced proportionally across the three macro work communities—e.g., not all the high frequency employees assigned to the same work community.

Our revised approach builds three macro communities by finding the three groupings of granular clusters that are the most "connected." The metric to assess the "connectedness" between macro communities is the EI index discussed in Chapter 5. As a refresher, the EI index can be calculated on any nodal attribute and provides a normalized measure of external/internal relationships based on that nodal attribute. This is useful when building macro communities because we can calculate the EI index for a grouping of granular clusters to assess "connectedness." The goal is to find macro communities that are most connected and assign these to the same work community. For example, a macro community of 10 granular clusters that has an EI index of -1 has no connections between granular clusters. This is not a particularly good macro community assignment—the goal is to assemble a macro community of granular clusters that share connections and should work in-office on the same day—mathematically, we want to maximize the EI index across three macro communities.

Practically, we accomplish this task using an approach similar to a grid search in hyper parameter optimization—however, we place constraints on the solution space to keep macro communities proportional. The objective is to find three macro communities which maximize the average EI index while maintaining less than 5% deviation from the in-office frequency mean of four days per week and less than 10% deviation from the mean of 178 nodes per macro community. This is accomplished by Algorithm 6.

Algorithm 6 Maximize EI Across Three Macro Work Communities

1: **procedure** FIND3OPTIMALWORKCOMMUNITIES(*NetworkObject*,*ClusterOptions*)
2: Instantiate "Best Macro Communities" as "BCs"
3: Create 3 Combos of *ClusterOptions* - Current Macro Communities
4: Build 3 Induced Sub-Graphs; 1 for Each Macro Community
5: Calculate Mean EI Index of the 3 Sub-Graphs
6: Calculate SD of Sub-Graph EI Index
7: Calculate SD of Sub-Graph Mean Attendance Freq.
8: Calculate SD of Sub-Graph Node Count
9: **IF** Current Mean EI Index $>$ BCs' EI Index
10: **AND** Current EI Index SD $<=$ 3 * BCs' EI SD
11: **AND** Current Attendance Freq. SD < 0.15 ▷ 3.75% from mean $= 4$
12: **AND** Current Node Count SD < 20 ▷ 11.6% from mean $= 178$
13: **THEN** Store current macro communities as BCs
14: Repeat Steps 3 through 13

The algorithm iteratively searches macro community assignments to find "connected" assignments while maintaining in-office frequency proportionality between communities. The procedure starts by randomly sampling three lists of clusters—in our data there are 35 clusters so the algorithm generates three lists of 11 to 12 cluster each—each list is a potential macro community. Then, for each macro community an induced sub-graph is created—this is simply a network that only includes the clusters in that macro community. For each of the three macro communities, we calculate network metrics to assess "goodness" of assignment. Again, the goal is to maximize the EI index across the three communities while constraining solutions that do not provide three relatively proportionally work communities. The steps are repeated until a desired EI threshold is achieved—in our study, trials stabilized at a local maximum of approximately -0.90.

The impact of this re-clustering routine is best articulated visually. The same network presented with granular clusters in Figure 7-10 is now displayed in Figure 7-12 with colors corresponding to three macro communities.

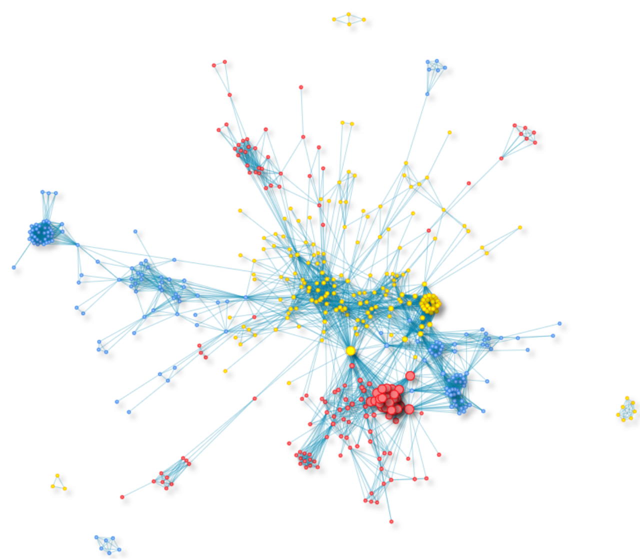


Figure 7-12: BI Employee Network— Three Macro Community Clusters

Furthermore, the impacts of our optimization routine can be visualized by breaking down the statistics of each macro community. The makeup of each macro community is detailed in Table 7.10. First, we validate the clustering routine and articulate what this means for hybrid work architects. The distribution of number of nodes in each macro community yield a standard deviation of 14.73 or approximately 8% from the mean—within our threshold. Furthermore, similar results are achieved in community EI and average monthly attendance. The key take away for business leaders is that networks have enabled us to build a model that can balance capacity without sacrificing work group collaboration. The standard deviation tolerances are simply parameters that an analytics team can adjust to meet the goals of real estate strategy leaders.

Macro Community	**Nodes Assigned**	**Community EI**	**Attendance**
Community 1	168	-0.89	4.28 Days/Month
Community 2	193	-0.88	4.06 Days/Month
Community 3	167	-0.94	4.13 Days/Month
Mean	**176**	**-0.90**	**4.16**
Standard Deviation	**14.73**	**0.03**	**0.11**

Table 7.10: Macro Community Statistics

Second, the work communities are highly significant and capture over **90%** of relationships in the network. The macro work communities yield an EI index of negative 0.86, an 11% EI improvement from the granular clusters assigned in step 3 of the pipeline—the maximum we could achieve is -1 or all relationships internal to macro community assignment. Translating the impact of macro communities into percentages is helpful. In this study, the three macro community assignments capture **93%** of relationships, leaving only 7% of relationships external to an employee's assigned community. Finally, using ERGM's, we estimate that the probability of a node having a meaningful relationship in his/her assigned macro community is **0.967**—near certainty that employees will be in the office with someone in their network.

The key finding of this study is that network clustering provides a scaleable method to balance capacity and optimize for work community collaboration. Although our method of re-aggregrating clusters at the end of the pipeline could be formulated as a more sophisticated optimization, we feel that the marginal gains are not relevant considering that the macro communities properly align **93%** of work relationships. To solidify the impact on hybrid work, we transition the macro community assignments into a simulation where we match schedules to communities to quantify the improvement.

7.6.5 Results of Data Backed Schedules

We simulated the typical month to assess the impact of the new scheduling pipeline. We begin by mapping work communities to schedules. Then, we simulate a month at the HQ campus and compare the current schedules with the future state schedules. Overall, the new scheduling pipeline achieves **88%** schedule alignment with stakeholders—an increase of **68** percentage points from the current state—and results in an over **4x** increase in the estimated network value of in-office work throughout the typical month. Furthermore, office space demand is spread across the core portion of the week and real estate operations expenses can be reduced by approximately 40%.

The first step in implementing community assignments is mapping work communities to day of the week assignments. Because our subject organization demonstrated a high preference against Monday/Friday scheduling, our finding is that creating three work groups simplifies scheduling and naturally aligns with Tuesday/Wednesday/Thursday. We assign a core day to each work community—community 1 = Tuesday, community 2 = Wednesday, community 3 = Thursday—and then build in overlap between communities using the remaining days of the week. The mapping is laid out below in Table 7.11.

Schedule Type	**Frequency**	**Community 1**	**Community 2**	**Community 3**
Weekly	1 Day	Tue	Wed	Thu
Weekly	2 Days	Tue/Wed	Tue/Wed	Wed/Thu
Weekly	3 Days	Tue/Wed/Thu	Tue/Wed/Thu	Tue/Wed/Thu
Biweekly*	1 Day/Wk	Tue	Wed	Thu
Monthly	1 Day	Tue	Wed	Thu
Monthly	2 Days	Tue/Wed	Tue/Wed	Wed/Thu
Monthly	3 Days	Tue/Wed/Thu	Tue/Wed/Thu	Tue/Wed/Thu
Custom*	1 Day/Wk	Tue	Wed	Thu

Table 7.11: Mapping Work Communities to Days of the Week
**Biweekly and Custom are reverted to 1 Day Weekly for simulation*

A key improvement with this method is that monthly scheduled employees now fall on the same day as the rest of their work group—maximizing meaningful in-person interaction while minimizing the need to come into the office more. The monthly assignments can be further aligned by week of the month—e.g., community 1 monthly employees come in on the first week of the month, community 2 aligns to the second week of the month, community 3 the third week of the month. Again, this prevents the misalignment between work group identified throughout this study. Furthermore, by mapping an overlap day on Wednesday, interaction between communities is still facilitated. In the simulation, we will see that only 15-20% of employees come in more than one day per week, but those that do will benefit from being aligned with other work communities on the overlap day.

The next step is simulating the current state. Each node in the BI network contains information on the employee's schedule. These plain text schedules are translated into a binary encoding that models the typical four week month. Furthermore, each node has an associated macro community from the clustering pipeline—this attribute enables visualization by community before schedules are adjusted using our mapping. The simulated results of the current state are presented in Figure 7-13.

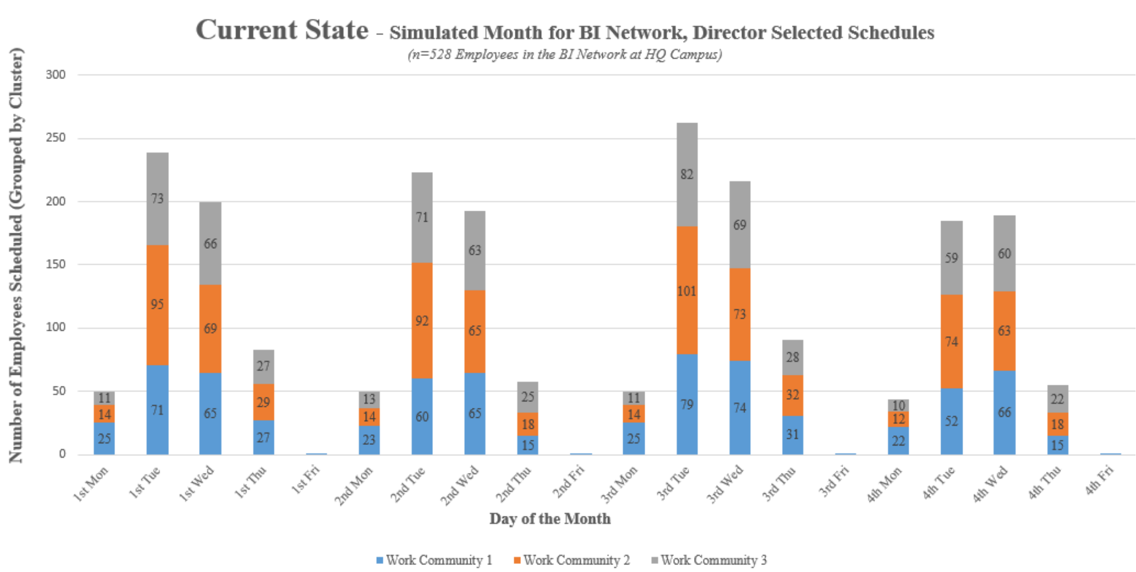


Figure 7-13: Current State Simulation

This matches our hypothesis that work communities would be relatively evenly distributed across the work week—this would be expected of a random process. Based on our research, this is the opposite of what organizations should strive for with hybrid work design. The goal needs to be maximizing in-person interaction between work groups while minimizing the need for extra days in the office. The current state does not achieve this objective—approximately one-third of each community is assigned in office throughout the week.

Using the simulated current state, we estimate that the **current average in-community alignment is 20%**. This establishes the reference baseline organizations can use to test whether clustering provides measurable scheduling improvement. The alignment value is calculated by averaging the percent of each community assigned across the days of the month.

Furthermore, to estimate the concept of a network effect in the office, we look to Metcalf's law [26]. Metcalf's law states that the value of a network equals the square of the number of users—more users results in a non-linear increase in value. This concept is generally accepted across platforms that rely on network effects. Applying this logic, we calculate the monthly network value by community—**community 1 = 40,275; community 2 = 54,695; community 3 = 40,394**. The baseline network value allows us to estimate the increase in network value when we compare the current state simulation to the future state simulation.

The same process is used to simulate the future state, but now the schedule map from Table 7.11 is implemented to schedule employees by communities. Future state results are visualized in Figure 7-14.

The key finding is that aligning employees according to macro community assignments captures **93%** of meaningful relationships, achieves **88%** alignment with stakeholders and provides a **4.6x increase** in office space network value/day of scheduling. These results match our objective of maximizing meaningful in-person collaboration while minimizing—or at least not increasing—the amount of time an employee is required to be in-office.

Furthermore, an ERGM can be used on the network behind the simulation to test whether employees scheduled on the same day are more likely to share work relationships—this is the overall goal of scheduling optimization. We fit two ERGMs to test this hypothesis; both demonstrate that community aligned schedules properly align employees. The first model matched nodes scheduled on the third Wednesday

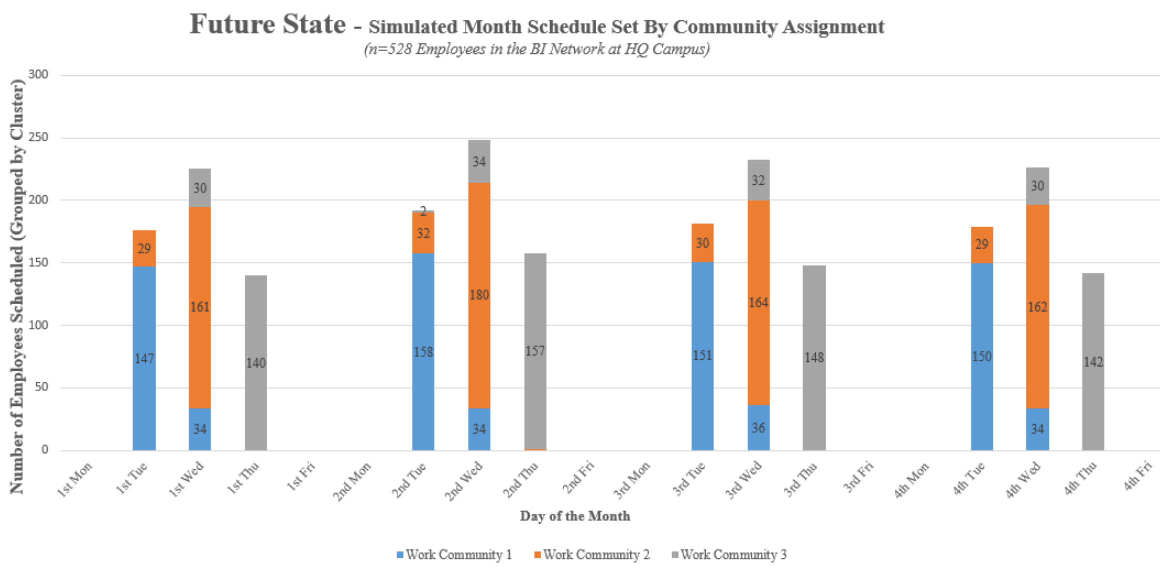


Figure 7-14: Current State Simulation

of the month and produces the estimate in Table 7.12.

Nodal Match/Covariate	Coeff. Est.	Std. Error	Z-Value	P-Value
Edges	-3.83	0.02648	-144.52	.0001***
NodeMatch(Scheduled 3rd Wed)	0.628	0.03286	19.11	.0001***
Null Deviance: 192142				
Residual Deviance: 37322				
AIC: 37326 BIC: 37346				

Table 7.12: ERGM 7 - Simulation Match on "Scheduled 3rd Wednesday"

The coefficient of this model can be interpreted as a 0.65 probability that two randomly selected nodes scheduled on the third Wednesday of the month share a meaning work relationship. This result makes sense, since we have three work groups over lapping on Wednesdays. Our expectation on Thursdays is a much higher model coefficient since the entirety of community three falls on Thursday. The model estimate is below in Table 7.13.

As expected, the probability of two nodes sharing a relationship when scheduled on the third Thursday is 0.84—meaning our scheduling pipeline is producing measurable increases in alignment.

Nodal Match/Covariate	Coeff. Est.	Std. Error	Z-Value	P-Value
Edges	-4.77	0.046	-103.31	.0001***
NodeMatch(Scheduled 3rd Thu)	1.72	0.049	34.97	.0001***
Null Deviance: 192142				
Residual Deviance: 35888				
AIC: 35892 BIC: 35912				

Table 7.13: ERGM 8 - Simulation Match on "Scheduled 3rd Thursday"

In summary, the network modeling pipeline presented in this chapter is a scaleable method that provides quantifiable improvement over the current manual heuristic scheduling solutions. Furthermore, the pipeline could be automated and executed on a quarterly basis. This would provide business leaders updated recommendations by campus as work relationship change due to project alignment. We do not argue that this method is perfect; however, we do believe it will create alignment across the majority of the organization while maintaining employee satisfaction and productivity.

7.6.6 Detailed Simulation Data

The current state simulation is presented in tabular format in Tables 7.12 through 7.16. The future state simulation data follows in Tables 7.17 through 7.21. These data are used to build the simulation visuals and quantify the improvements presented in the previous section. The key take away is that assigning simple communities can increase the network value while minimizing the number of days employees allocate in-office time. This allows organizations to quantify the value per/day of office operations to gauge if opening the office for more than three days per week is valuable to the network.

Table 7.14: Simulated 1st Week of Month—Current State

Community	**1-Mon**	**1-Tue**	**1-Wed**	**1-Thu**	**1-Fri**
1 (# Scheduled)	25	71	65	27	1
2 (# Scheduled)	14	95	69	29	0
3 (# Scheduled)	11	73	66	27	0
Comm. 1 Aligned	15%	42%	39%	16%	0%
Comm. 2 Aligned	7%	49%	36%	15%	0%
Comm. 3 Aligned	7%	44%	40%	16%	0%
Comm. 1 Value (n^2)	625	5041	4225	729	1
Comm. 2 Value (n^2)	196	9025	4761	841	0
Comm. 3 Value (n^2)	121	5329	4359	729	0

Table 7.15: Simulated 2nd Week of Month—Current State

Community	**2-Mon**	**2-Tue**	**2-Wed**	**2-Thu**	**2-Fri**
1	23	60	65	15	1
2	14	92	65	28	0
3	13	71	63	25	0
Comm. 1 Aligned	14%	36%	39%	9%	0%
Comm. 2 Aligned	7%	48%	34%	9%	0%
Comm. 3 Aligned	8%	43%	38%	15%	0%
Comm. 1 Value (n^2)	529	3600	4225	225	1
Comm. 2 Value (n^2)	196	8464	4225	324	0
Comm. 3 Value (n^2)	169	5041	3969	625	0

Table 7.16: Simulated 3rd Week of Month—Current State

Community	**3-Mon**	**3-Tue**	**3-Wed**	**3-Thu**	**3-Fri**
1	25	79	74	31	1
2	14	101	73	32	0
3	11	82	69	28	0
Comm. 1 Aligned	15%	47%	44%	18%	0%
Comm. 2 Aligned	7%	52%	38%	17%	0%
Comm. 3 Aligned	7%	49%	41%	17%	0%
Comm. 1 Value (n^2)	625	6241	5476	961	1
Comm. 2 Value (n^2)	196	10201	5329	1024	0
Comm. 3 Value (n^2)	121	6724	4761	784	0

Table 7.17: Simulated 4th Week of Month—Current State

Community	**4-Mon**	**4-Tue**	**4-Wed**	**4-Thu**	**4-Fri**
1	22	52	66	15	1
2	12	74	63	18	0
3	10	59	60	22	0
Comm. 1 Aligned	13%	31%	39%	9%	0%
Comm. 2 Aligned	6%	38%	33%	9%	0%
Comm. 3 Aligned	6%	35%	36%	13%	0%
Comm. 1 Value (n^2)	484	2704	4356	225	1
Comm. 2 Value (n^2)	144	5476	3969	324	0
Comm. 3 Value (n^2)	100	3481	3600	484	0

Table 7.18: Current State Summary

Community	**Average Alignment**	**Network Value**	**Freq. In-Office**	**Daily Value**
1	21% /day	40,275 /month	20 days/month	2,013 /day
2	20% /day	54,695 /month	16 days/month	3,418 /day
3	21% /day	40,394 /month	16 days/month	2,524 /day

Table 7.19: Simulated 1st Week of Month—Future State

Community	**1-Mon**	**1-Tue**	**1-Wed**	**1-Thu**	**1-Fri**
1 (# Scheduled)	0	147	34	0	0
2 (# Scheduled)	0	29	161	0	0
3 (# Scheduled)	0	0	30	140	0
Comm. 1 Aligned	N/A	88%	20%	N/A	N/A
Comm. 2 Aligned	N/A	15%	83%	N/A	N/A
Comm. 3 Aligned	N/A	N/A	18%	84%	N/A
Comm. 1 Value (n^2)	0	21609	1156	0	0
Comm. 2 Value (n^2)	0	15	25921	0	0
Comm. 3 Value (n^2)	0	0	900	19600	0

Table 7.20: Simulated 2nd Week of Month—Future State

Community	**2-Mon**	**2-Tue**	**2-Wed**	**2-Thu**	**2-Fri**
1	0	158	34	0	0
2	0	32	180	1	0
3	0	2	34	157	0
Comm. 1 Aligned	N/A	94%	20%	N/A	N/A
Comm. 2 Aligned	N/A	17%	93%	0%	N/A
Comm. 3 Aligned	N/A	0%	20%	94%	N/A
Comm. 1 Value (n^2)	0	24964	1156	0	0
Comm. 2 Value (n^2)	0	1024	32400	1	0
Comm. 3 Value (n^2)	0	4	1156	24649	0

Table 7.21: Simulated 3rd Week of Month—Future State

Community	**3-Mon**	**3-Tue**	**3-Wed**	**3-Thu**	**3-Fri**
1	0	151	36	0	0
2	0	30	164	0	0
3	0	0	32	148	0
Comm. 1 Aligned	N/A	90%	21%	N/A	N/A
Comm. 2 Aligned	N/A	16%	85%	N/A	N/A
Comm. 3 Aligned	N/A	N/A	19%	89%	N/A
Comm. 1 Value (n^2)	0	22801	1296	0	0
Comm. 2 Value (n^2)	0	900	26896	0	0
Comm. 3 Value (n^2)	0	0	1024	21904	0

Table 7.22: Simulated 4th Week of Month—Future State

Community	**4-Mon**	**4-Tue**	**4-Wed**	**4-Thu**	**4-Fri**
1	0	150	34	0	0
2	0	29	162	0	0
3	0	0	30	142	0
Comm. 1 Aligned	N/A	89%	20%	N/A	N/A
Comm. 2 Aligned	N/A	15%	84%	N/A	N/A
Comm. 3 Aligned	N/A	N/A	18%	85%	N/A
Comm. 1 Value (n^2)	0	22500	1156	0	0
Comm. 2 Value (n^2)	0	841	26244	0	0
Comm. 3 Value (n^2)	0	0	900	20164	0

Table 7.23: Future State Summary

Community	**Average Alignment**	**Network Value**	**Freq. In-Office**	**Daily Value**	**Improvement**
1	90% on core day	96,638 /month	8 days/month	12,080 /day	6.0x /day
2	86% on core day	115,068 /month	8 days/month	12,785 /day	3.75x /day
3	88% on core day	90,301 /month	8 days/month	10,033 /day	3.98x /day

Chapter 8

Testing the New System and Future Work

At the close of this six month study, leaders in HR and Real Estate began formally collaborating to implement a large scale experiment with the new system. Our final chapter recaps the findings in this study and provides input on experimental structure. Furthermore, we discuss potential follow on research applying networks to real estate strategy and employee collaboration.

8.1 Summary of Contributions

At its core, hybrid work is a coordination problem—how frequently and when should hybrid employees come into the office? We have shown that collaboration data and employee network models can help organizations answer these questions. Ultimately, to achieve optimal results for an organization, a hybrid work system needs to minimize the amount of time hybrid employees are required to be in the office while maximizing the correct collaborations during office visits. When designed in this manner, employees can capitalize on the deep focus that comes with work outside of the office.

In sum, this study comes to a number of conclusions that can help organizations implementing hybrid work. In this section, we recap the findings for Verizon and then think about what these findings imply more generally.

8.1.1 Recap of Study Findings at Verizon

This study reaches four broad conclusions in the subject organization:

1. At the HQ campus, employees in our sample collaborate on average less than four hours per week with co-located colleagues. Based on this sample, we conclude that one day in the office per week can provide the requisite collaboration if an employee is properly aligned with his/her work community.

2. Simulations show that office space capacity is not a concern across the major hub campuses—on average, office space demand will fall 60% below capacity once hybrid workers return to the office.

3. Hybrid schedules are currently misaligned. At least partially, this is because schedules are being set at an unrealistically high level—directors have too many employees to understand schedule intricacies. The second factor contributing to misalignment is over complicated scheduling options.

4. Employee collaboration data can be used to build networks and align schedules. When coupled with appropriate in-office frequency, a network model matches work communities and achieves the goal of minimizing in-office frequency, while maximizing necessary collaborations. Furthermore, network models can help spread demand for office space throughout the work week without separating employees who need to collaborate.

These findings are specific to the subject organization but hold insights for every large organization shifting to hybrid work. This study provides a road map organizations can use to decipher the optimal hybrid work system.

8.1.2 Generalized Findings

More generally, this study provides three key contributions to the field of hybrid work design:

1. When analyzing the data, knowledge workers spend less time collaborating with co-located colleagues than leaders may intuitively imagine. Workplace data and calendar data provide a means to estimate how many hours per week hybrid workers need in the office to satisfy minimum collaboration requirements.

2. Hybrid work introduces a number of variables that may impede collaboration between teams that should be working together in-person—especially in matrixed organizations. We have shown that network models and simple schedule recommendations can ensure key links are maintained and employee demands for flexibility are satisfied.

3. Finally, organizations gain a competitive advantage by using data to find and orchestrate the optimal balance between in-person and remote work—employees gain deep focus from time working remotely and maintain interpersonal connections when in the office.

8.2 Running a Hybrid Scheduling Experiment

We encourage organizations to run experiments when hybrid workers return to the office in the post COVID19 era—rigorously test the ideas presented in this study. The subject organization began designing a large scale business experiment at the close of 2021. Fueled by continuing employee feedback around alignment, HR combined forces with our study to test the concepts at scale. In this section we develop an experimental structure—the three key points are level of the intervention, duration of the experiment and what to measure.

The ideal experiment needs to be executed at a high enough level in the organizational hierarchy to provide a diverse and random sample of employees. We believe that the correct level in the subject organization hierarchy is a vice president's business unit. Specifically, we recommend using the operations vice president as the test population and marketing vice president as the control population. These organizations each have an average in-office frequency of approximately one day per week and contain a variety of employee specialties. Furthermore, the scope of the experiment should not be contained to a single campus as in our proof of concept—opening up the problem will test if there are benefits for geographical dispersed organizations.

While it may be tempting to shorten experimental duration to deliver results, the hybrid employee return to office will take time to stabilize. We recommend a minimum experiment duration of six months. This timeline would begin when the subject organization encourages employees to return to offices according to designed schedules. Ideally, this allows for three to four months of data analysis prior to calendar year 2023 in the case of our subject organization. The results of the experiment could then influence system changes in the next calendar year.

A vital piece of the experiment structure is how to measure impact of the new scheduling system. We recommend using a combination of quantitative and qualitative data sources to assess the performance of the new scheduling system. The quantitative data can be drawn from the Book-A-Space reservation system. Analysts can use Book-A-Space data to calculate four quantitative metrics: 1) schedule adherence, 2) flex day bookings, 3) network cohesion metrics and 4) number of teleconferences when employees are in the office. Each metric can be calculated for the test and control population to help leaders understand how the new system is impacting hybrid work. We expect that the test population will display a higher schedule adherence, lower flex behavior, higher cohesion and less teleconferences when they are reserved in the office. These are objective data points that can demonstrate if the new system achieves alignment. Finally, capacity data between the groups can tell real estate planners if the new system is spreading demand across the work week.

Qualitative data can provide insight into the employee perspective on how hybrid work is evolving. A survey methodology can test employee satisfaction between the legacy system and the new system design. Furthermore, the subject organization can use retention data to test whether the legacy scheduling system contributes to more attrition. Finally, a telling qualitative data point will be the number of schedule changes that occur six months into hybrid work—our hypothesis is that the new system will yield fewer changes because employees are aligned with their networks.

8.3 Potential Areas for Future Work

There are a number of extensions to this study. Two areas of research which would help businesses orchestrate hybrid work are network recommender systems and hybrid worker productivity. Recommender systems are ubiquitous in today's society and could eliminate the need for a scheduling system. A network recommender system could use the reservation data and employee calendar data to understand employee networks and recommend the optimal days to schedule in-person collaboration. The other area is productivity—testing what mix of in-person and remote work is most productive. While the common response is that this answer will depend by job function, it has not been formally investigated. Once workers return to the office and adopt different hybrid patterns, a team could connect with HR systems such as WorkDay to test if there is correlation between productivity and frequency of in-person collaboration. Each of these research areas could answer business leader's questions and help organizations build the optimal hybrid work design.

Bibliography

[1] Thomas J. Allen and Günter Henn. *The organization and architecture of innovation: managing the flow of technology*. OCLC: ocm72868452. Amsterdam ; Boston: Elsevier ; Butterworth-Heinemann, 2007. ISBN: 9780750682367.

[2] George Anders. *How the pandemic changed us: Our fastest rising priority is job flexibility | LinkedIn*. 2022. URL: `https://www.linkedin.com/pulse/how-pandemic-changed-us-our-fastest-rising-priority-job-george-anders/` (visited on 01/17/2022).

[3] Jose Maria Barrero, Nicholas Bloom, and Steven Davis. *Why working from home will stick*. Working Paper 28731. National Bureau of Economic Research, 2021. URL: `https://wfhresearch.com` (visited on 12/14/2021).

[4] Nicholas Bloom et al. "Does Working from Home Work? Evidence from a Chinese Experiment". In: *The Quarterly Journal of Economics* 130.1 (Nov. 2014). _eprint: https://academic.oup.com/qje/article-pdf/130/1/165/30629971/qju032.pdf, pp. 165–218. ISSN: 0033-5533. DOI: `10.1093/qje/qju032`. URL: `https://doi.org/10.1093/qje/qju032`.

[5] P Bonacich. "Factoring and weighting approaches to status scores and clique identification". In: *Journal of Mathematical Sociology* 2.1 (1972), pp. 113–120.

[6] *Cisco Study Finds Telecommuting Significantly Increases Employee Productivity, Work-Life Flexibility and Job Satisfaction*. en-US. URL: `https://newsroom.cisco.com/press-release-content?type=webcontent&articleId=5000107` (visited on 09/01/2021).

[7] Matthew Claudel et al. "An exploration of collaborative scientific production at MIT through spatial organization and institutional affiliation". In: *PLOS ONE* 12.6 (June 2017). Ed. by Igor Linkov, e0179334. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0179334`. URL: `https://dx.plos.org/10.1371/journal.pone.0179334` (visited on 09/03/2021).

[8] Dawn Foster. *Cisco's Take on Telecommuting and Productivity - NYTimes.com*. 2009. URL: `https://archive.nytimes.com/www.nytimes.com/external/gigaom/2009/06/26/26gigaom-ciscos-take-on-telecommuting-and-productivity-81155.html` (visited on 09/01/2021).

[9] Michael Gibbs, Friederike Mengel, and Christoph Siemroth. *Work from Home & Productivity: Evidence from Personnel & Analytics Data on IT Professionals.* Working Paper 2021-56. Becker Friedman Institute for Economics at UChicago: University of Chicago, July 2021, p. 40. URL: `https://bfi.uchicago.edu/working-paper/2021-56/` (visited on 08/31/2021).

[10] Mark S. Granovetter. "The Strength of Weak Ties". In: *American Journal of Sociology* 78.6 (May 1973), pp. 1360–1380. ISSN: 0002-9602, 1537-5390. DOI: `10.1086/225469`. URL: `https://www.journals.uchicago.edu/doi/10.1086/225469` (visited on 09/03/2021).

[11] Adam Grant. "The Real Meaning of Freedom at Work". en-US. In: *Wall Street Journal* (Oct. 2021). ISSN: 0099-9660. URL: `https://www.wsj.com/articles/the-real-meaning-of-freedom-at-work-11633704877` (visited on 01/17/2022).

[12] Rob Cross and Peter Gray. *Optimizing Return-to-Office Strategies With Organizational Network Analysis.* 2021. URL: `https://sloanreview.mit.edu/article/optimizing-return-to-office-strategies-with-organizational-network-analysis/` (visited on 10/18/2021).

[13] Robert Hanneman and Mark Riddle. *Introduction to Social Network Methods.* epublished, 2005. URL: `http://faculty.ucr.edu/~hanneman/nettext/index.html`.

[14] Jenine Harris. *An Introduction to Exponential Random Graph Modeling.* 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc., 2014. ISBN: 9781452220802 9781452270135. DOI: `10.4135/9781452270135`. URL: `http://methods.sagepub.com/book/an-introduction-to-exponential-random-graph-modeling` (visited on 11/02/2021).

[15] Michaela Hoffman et al. "Detecting Clusters/Communities in Social Networks". In: *Multivariate behavioral research* 53.1 (2018), pp. 57–73. ISSN: 0027-3171. DOI: `10.1080/00273171.2017.1391682`. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6103523/` (visited on 11/17/2021).

[16] David R. Hunter et al. "ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks". In: *Journal of statistical software* 24.3 (May 2008), nihpa54860. ISSN: 1548-7660. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2743438/` (visited on 11/02/2021).

[17] *Introduction to ERGMs.* en-US. URL: `https://eehh-stanford.github.io/SNA-workshop/ergm-intro.html` (visited on 11/02/2021).

[18] L Katz. "A new status index derived from sociometric analysis". In: *Psychometricka* 18.1 (1953), pp. 39–43.

[19] Eric D. Kolaczyk and Gabor Csardi. *Statistical Analysis of Network Data with R.* 2nd ed. Vol. 1. Switzerland AG: Springer, 2020. ISBN: 978-3-030-44129-6. (Visited on 08/01/2021).

[20] Cory Maklin. *Spectral Clustering Algorithm Implemented From Scratch*. en. July 2019. URL: https://towardsdatascience.com/unsupervised-machine-learning-spectral-clustering-algorithm (visited on 11/18/2021).

[21] Tsedal Neeley. *Remote work revolution: succeeding from anywhere*. New York: Harper Business, 2021. ISBN: 9780063068308.

[22] M. E. J. Newman. "Fast algorithm for detecting community structure in networks". eng. In: *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics* 69.6 Pt 2 (June 2004), p. 066133. ISSN: 1539-3755. DOI: 10.1103/PhysRevE.69.066133.

[23] M. E. J. Newman. "Finding community structure in networks using the eigenvectors of matrices". In: *Physical Review E* 74.3 (Sept. 2006). arXiv: physics/0605087, p. 036104. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.74.036104. URL: http://arxiv.org/abs/physics/0605087 (visited on 11/18/2021).

[24] *R Documentation*. URL: https://www.rdocumentation.org/packages/igraph/versions/0.4.3/topics/leading.eigenvector.community (visited on 11/18/2021).

[25] Holger Reisinger and Dane Fetterer. "Forget Flexibility. Your Employees Want Autonomy." In: *Harvard Business Review* (Oct. 2021). ISSN: 0017-8012. URL: https://hbr.org/2021/10/forget-flexibility-your-employees-want-autonomy (visited on 01/17/2022).

[26] C. Shapiro and H.R. Varian. *Information Rules: A Strategic Guide to the Network Economy.* Harvard Business Review Press, 1998.

[27] Dai Shizuka. *Measuring Networks Part 2: Community Structure and Assortment.* https://dshizuka.github.io/networkanalysis/05$_c$*ommunity.html*. June 2019. URL: https://dshizuka.github.io/networkanalysis/05_community.html (visited on 09/01/2021).

[28] Meridith Somers. *A 4-step framework for returning to the office*. 2021. URL: https://mitsloan.mit.edu/ideas-made-to-matter/a-4-step-framework-returning-to-office (visited on 12/14/2021).

[29] Jaime Teevan, Brent Hecht, and Sonia Jaffe. *The New Future of Work: Research from Microsoft on the Impact of the Pandemic on Work Practices*. Tech. rep. Microsoft, 2021, p. 65. URL: https://aka.ms/newfutureofwork (visited on 09/02/2021).

[30] Richard Thaler and Cass Sunstein. *Nudge.* New Haven: Yale University Press, 2008.

[31] Shailesh Tripathi et al. "Comparison of module detection algorithms in protein networks and investigation of the biological meaning of predicted modules". In: *BMC Bioinformatics* 17 (Mar. 2016), p. 129. ISSN: 1471-2105. DOI: 10.1186/s12859-016-0979-8. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4797184/ (visited on 11/18/2021).

[32] Stanford University. *The productivity pitfalls of working from home in the age of COVID-19*. Mar. 2020. URL: https://news.stanford.edu/2020/03/30/productivity-pitfalls-working-home-age-covid-19/ (visited on 09/03/2021).

[33] Nithya Vaduganathan. *The How-To of Hybrid Work*. Apr. 2021. URL: https://www.bcg.com/publications/2021/identifying-postpandemic-work-model (visited on 12/14/2021).

[34] *Verizon History*. URL: https://www.verizon.com/about/sites/default/files/Verizon_History_0916.pdf.

[35] Zhao Yang, René Algesheimer, and Claudio J. Tessone. "A Comparative Analysis of Community Detection Algorithms on Artificial Networks". In: *Scientific Reports* 6.1 (Aug. 2016), p. 18. ISSN: 2045-2322. DOI: 10.1038/srep30750. URL: https://www.nature.com/articles/srep30750 (visited on 08/31/2021).