Cultural Integration in Horizontal Mergers: A New Model for Measuring Effective Organizational Change

Geoffrey P. Morgan

CMU-ISR-22-100
April 2022

Institute for Software Research
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Kathleen M. Carley, Chair
Linda Argote
Brandy Aven
Guiseppe Labianca (University of Massachusetts, Amherst)

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Computation, Organizations, and Society

The research reported in this thesis has been supported in part by the Carnegie Mellon University Center for Computational Analysis of Social and Organizational Systems (CASOS). This research was also supported by the Office of Naval Research, grants N000142112229 and N00014-17-1-2675, as well as the Office of Naval Research Minerva Grant N000141512797. I am grateful for the assistance of the CHRO of "MergedCo," as well as all the employees who assisted the research team in this data collection (particularly within the HR and IT departments), without whom this research would not have been possible.

Keywords: horizontal merger, diffusion, social networks, natural language processing, computational organizational theory, simulation

For my family

Abstract

Horizontal mergers of two formerly competing companies often fail in the integration stage, when employees of each organization start working side by side with former competitors. I believe these failures occur because of ineffective diffusion of a new shared organizational culture and its corresponding public identity. Many organizations rely on motivated leadership propagating the identity to their direct reports. This practice leads to the new identity's uneven adoption, because while receptive employees will accept and spread it easily, those who are less receptive may block the new culture's spread to their work groups. This uneven adoption, in turn, causes mistakes and hinders both individual and group performance. Resistors seize upon poor group performance as further reason not to adopt the new organizational culture, which only lengthens the period of struggle. This phenomenon is particularly damaging in virtual organizations — those that rely on teams of subject matter experts to create value. Virtual organizations are growing more common in the knowledge-driven economy.

In this work, I address elements of horizontal merger failure across three chapters. Each chapter relies on real-world data from the period immediately following the horizontal merger of two large multinational corporations. I will refer to the resulting corporation as MergedCo. In Chapter 1, I use MergedCo's email data to generate windowed reciprocity networks, which represent the organization at work and are good indicators of how the new organizational identity will spread. In Chapter 2, I use the text of MergedCo's emails to characterize the spread of the new organizational culture and elaborate on the implications for MergedCo. In both Chapter 1 and Chapter 2, I compare analysis outcomes to actual survey data from employees at MergedCo.

In Chapter 3, I synthesize my theory of organizational culture diffusion into an active multilevel simulation, which I call the Unified Network Model. I validate the Unified Network Model in relation to its predecessors and demonstrate the model's ability to emulate six stylized facts that are important to organizations. I then instantiate the model with data from the MergedCo case study and comparing the simulation outcomes to actual MergedCo data from a later time period.

Across this work, I have developed new methods that may be re-used in future scholarship. I believe I have contributed three important ideas to the literature:

- The spread of a new organizational identity is a diffusion problem, and that problem helps explain the oft-described integration issues common in large horizontal mergers.
- Quantitative measurement of language tokens provides an empirical and multilevel measure of cultural change over time.
- An active multilevel simulation of organizational operation can predict empirical organizational outcomes of horizontal mergers.

These findings have important implications for organizational leaders and consultants who oversee mergers. For one, they show that leaders should preserve systems of value generation rather than single high-performing employees. Merger consultants must devise explicit strategies for spreading the new organizational culture. Both groups should recognize the importance of corporate culture compatibility.

Acknowledgments

Many people helped me on this journey. First and foremost among them is my adviser, Dr. Kathleen Carley, who offered me a place to study at Carnegie Mellon, supported my interest in organizational phenomena, provided me an opportunity to work with amazing data, and offered her valuable time and insights as I made intermittent progress on this work.

I want to thank Dr. Joe Labianca for the opportunity to work with the MergedCo data, and Dr. Jesse Fagan for all the hard work he did to make the data available for analysis in the first place. Without Jesse's and Joe's generous help, this dissertation could not exist in its current form.

I want to thank Dr. Linda Argote for supporting my quest to take every organizational theory class available at The Tepper School of Business at Carnegie Mellon and for letting me participate in an MBA capstone course on applied organizational change. I also want to thank Dr. Brandy Aven for allowing me to TA her social networks MBA class. Both experiences have been very useful to me in my professional journey and helped advance my interest in this research area.

I want to thank my many academic colleagues from my time at the Center for Computational Analysis of Social and Organizational Systems, particularly Dr. Michael Lanham, who provided me ten times the support I ever gave him; Dr. Kenneth Joseph, who influenced my ideas on how to make effective and inexpensive computational agents; and Dr. Terril Frantz, who helped show me what was possible with computational modeling.

I want to thank my professional colleagues, including Dr. Mitchell Sipus, who invited me to co-create not one but two beautiful dreams in the private sector, and Dr. Israel Alguindigue, who has made it easy to make completing this dissertation a priority.

I want to thank my family, especially my brother, Dr. Jonathan H. Morgan, who helped me even from an ocean away, and my mother, Renee Lautzenhiser, who was always the first to provide substantive feedback and suggestions for improvement.

And finally, I must thank Tory Wegerski, my wife and love, who believed in me and supported my progress on this "almost done dissertation" for our entire relationship. I cannot imagine the true depths of her patience.

Contents

List of Figures

List of Tables

List of Equations

Chapter 1: A Structural Analysis of an Organization in Transition

Abstract
In this work, I evaluate an ongoing horizontal merger using empirical longitudinal email and survey data. Because the mere sending of emails is rarely a true signifier of coworker connection inside a large organization, I develop a novel method, called windowed reciprocity, for identifying reciprocal links. Unlike traditional methods, windowed reciprocity incorporates response time into the definition of a network tie between two coworkers. To demonstrate the value of windowed reciprocity, I compare networks generated by this method to those based on raw email network data. I analyze how well each type of network correlates to data from employee surveys.

In both the traditional email and reciprocity networks, the number of connections an individual has correlates with many important survey measures. However, one survey measure—whether employees feel they receive recognition for their work—correlates well with connections in the reciprocity network but not with connections in the raw email network. The fact that connections are directional in a windowed reciprocity network is also illuminating. I find that connections *to* the individual in the reciprocity network correlate well with the survey measure of job satisfaction, even though the individual's number of reciprocal connections with others does not.

In total, I analyze 1.7 MM emails, as well as survey data that coincides in time with those emails. My analyses reveal how the organizational units associated with each pre-merger company fare over two time periods: Time-1, three months into the merger process, and Time-2, nine months after that. I describe the analytical processes used to come to my conclusions, which I consider a significant step forward for data-driven measurement of organizational change.

Introduction
Horizontal mergers—the merging of two organizations in the same industry—are notoriously difficult (Burke & Cooper, 2000; Dion, Allday, Lafforet, Derain, & Lahiri, 2007; Toma et al., 2012). For a horizontal merger to succeed, two distinct organizations and their associated cultures and climates must merge into a single mélange—their cultures blended into a new, unified culture that is distinct from either of its source cultures. Furthermore, core administrative and support functions including accounting, communications, human resources (HR), and infrastructure management each have their own subcultures and idiosyncratic climates that must merge in order to function within the unified company.

For a horizontal merger to succeed, leaders from each organization must set an example by reaching outside their own habitual interaction patterns to include their newest coworkers and co-leaders. These leaders, along with other specialists, must take stock of the tacit and explicit knowledge held by each of the source functional groups and reconcile these bodies of knowledge to create the new merged functional groups. Reconciling the forms, processes, and procedures of specialized groups requires enormous amounts of communication along previously nonexistent paths, however. Organizations that do not make the necessary effort to create new communication patterns and reconcile the cultures, climates, and knowledge of the merging organizations substantially reduce the probability that their merger will succeed.

MergedCo is a fictional name for a real organization resulting from the acquisition of one multinational firm, which I call StandardCo, by another multinational firm, which I call LuxuryCo. LuxuryCo and StandardCo operated in the same industry and at the same point in the value chain, making the merger a horizontal merger. Before the merger, both LuxuryCo and StandardCo were mature organizations with multiple locations and thousands of employees. To enable this observational study, MergedCo has provided metadata and text data from email communications, as well as the data from an annual employee survey designed to measure the state of the merger. In this work, I will focus on the metadata of the email (i.e., who emailed whom and when). I will use the survey data to cross-validate the results of the metadata analysis.

I group these emails into two time-periods, Time-1, three months into the merger process, and Time-2, a year after the merger and nine months after Time-1. At a broad level, I also consider three groups of employees:

1. LuxuryCo employees—people who were part of LuxuryCo when the merger was announced
2. StandardCo employees—people who were part of StandardCo when the merger was announced
3. MergedCo employees—people who were not part of either LuxuryCo or StandardCo when the merger was announced and were hired directly into MergedCo

Using Dynamic Network Analysis (Carley, 2003), I analyze the communications of LuxuryCo and StandardCo employees in both Time-1 and Time-2. This analysis attempts to answer several specific questions that I believe are valuable in understanding this merger in progress and that can also be generalized to help organizational leaders assess the efficacy of their merger efforts.

First: On a broad level, did interaction between LuxuryCo employees and StandardCo employees increase between Time-1 and Time-2? If the merger is proceeding successfully, interactions across these legacy organizations should be increasing. Indicators of trust should also be increasing, while contraindicators should be decreasing. While not every individual will have the opportunity or interest to interact with cross-legacy coworkers, the focus here is on identifying trends and discussing the implications for business processes and for future research.

Second: What is the typical reciprocity of messaging across all of MergedCo in both Time-1 and Time-2? By reciprocity, I mean the likelihood of receiving an emailed response from one or more of the recipients of the original message within a specified twenty-four (24) hour period after the original message was sent. In the social network literature, reciprocity is when a bidirectional link occurs between two nodes. Asymmetric reciprocation is a strong indicator of status differences between the two alters (Hallinan, 1978; Newcomb, 1961), and the speed of the reciprocation can also reflect status differences (Burgoon, Dillman, & Stem, 1993; Doreian, 2002; Subbian, Srivastava, Pinar, Singhal, & Kolda, 2013; Wang et al., 2011). I therefore consider this temporal data in addition to overall patterns of reciprocity to analyze the organization's status dynamics.

Third: Are interactions between LuxuryCo employees and StandardCo employees less reciprocal than interactions across MergedCo more generally? Low reciprocity is an indicator not only of status, but also of low trust and low information flow. This would suggest a suboptimal situation for MergedCo. A horizontal merger requires the unification of business functions in order to lower overhead and capture the financial promise of the merger. MergedCo's initial statements about the merger indicated the potential for tens of millions of dollars of cost savings through business function unification. Low reciprocity indicates that, while the formal organization has nominally consolidated these functions across legacy groups, divisions between the groups remain real and affect the day-to-day experience of the average employee.

Fourth: Are reciprocity and interactions across legacy organizations dependent on the functional role of the employees? Are there functional groups—such as accounting, HR, or sales—where interaction and reciprocity are particularly high or low? If these measures are not consistent between functional groups, that may be an indicator that the functional leaders tasked with unifying their various departments have not managed the change with equal effectiveness.

Fifth and finally: When I compare low reciprocity and low interaction across legacy organizations as predictors of poor climate, do those predictions correlate with actual survey measures of perceived fairness, commitment to the merger, and feelings of recognition by employees?

Motivating Theory

In this section, I provide an overview of organizational forms. I explain how the Weberian bureaucracy managed the formal network—the lines of authority and control within an organization—in a way that minimized how much any individual could disrupt the operation of the organization. This has changed with the rise of virtual organizations, however. In a virtual organization, individual experts must work together to create value. The connections between these individuals form the organization's informal network. Because each individual is an expert in their own domain, it is difficult to replace them. Furthermore, the *connections* between these experts represent earned knowledge about how to best collaborate to meet the organization's goals. A horizontal merger disrupts these connections, and some key individuals are lost entirely. These disruptions, in turn, harm the organization's ability to create value. Because virtual organizations must manage both the formal and informal networks throughout a horizontal merger, integrating successfully is much more difficult than it is for a more traditional Weberian bureaucracy, which needs only to focus on the formal network.

The Weberian Bureaucracy

How humans are organized at work has evolved over the centuries, with systems including farm-based communities; guilds within cities and mercantile states; caste systems and hereditary professions; and civil services in ancient Egypt, China, and other more modern civilizations. A recurring problem for many forms of organization is over-reliance on individual personalities, such that the departure of prominent individuals can negatively affect the operation and capabilities of the entire organization. By contrast, Max Weber described bureaucracy as an attempt to develop an organizational form that minimizes the impact of individual idiosyncrasies by defining tasks (e.g., execution, information passing and filtering, strategic thinking) in terms of interchangeable role sets (Robert K Merton, 1957; Weber, 1922). If prominent individuals leave the firm, the Weberian bureaucracy identifies new people to fill those roles in the organization, and ideally the firm's operations continue unchanged. Of course, sufficiently prominent individuals, such as chief executives, can still influence the outcomes of the organization, but the innovations of the Weberian bureaucracy—employed everywhere from ancient China's civil service to modern-day militaries—sought to minimize the number of individuals who could exert a strong pull on the organization's fate.

While well-suited to the manufacturing and corporate work of the industrial age, however, the Weberian bureaucracy is vulnerable to structural rigidity — an inability to reorganize itself to cope with external shocks, or sudden changes in the world that reduce the value of the products the firm was previously offering. What good is being the finest carriage maker in the world, for example, once the automobile arrives? Flexible and visionary carriage makers conducted exploratory learning (March, 1991) and adapted their existing capabilities into venues that were still profitable, such as detailing luxury vehicles or even making automobiles. Those with too rigid a bureaucracy, however, were unable to make the shift.

The Weberian bureaucracy is highly susceptible to problems associated with its formal network. Because these organizations often add new levels of management over time, one common cost-cutting strategy is to reduce the number of managers at each level. However, this often leaves the remaining managers overstretched and unable to supervise their direct reports adequately, because there is simply too much information for each individual to receive and process. Under these circumstances, organizations can develop 'hot spots' (Carley, 1991, 1992) — positions in which any employee would appear to be underperforming, no matter how superb their individual capabilities. The typical response to this problem is to cycle many individuals through the position, finding all wanting, until the organization finally chooses to rework its formal structure, which often requires years of effort.

Structural rigidity is also evidenced by the highly directed flows of information in bureaucracies, as well as promotion policies that generate homogeneity among top decision makers. This situation generally contributes to group-level confirmation biases such as The Matthew Effect and groupthink (Janis, 1972; Robert King Merton & Merton, 1968).

Virtual Organizations

With the dawn of the information age, how organizations did work began to change. Information technology enabled work to be done remotely, and cross-functional teams could focus on solving specific problems more efficiently (Drucker, 1988). As these teams became more common, informal networks re-emerged as important contributors to group performance and helped mitigate the worst effects of bureaucracy by encouraging information flow from low to high levels, as well as across organizations (Tushman & Nadler, 1978). A side effect of this mitigation, however, was that the informal networks could — without management — become harmful to the organization (Krackhardt & Stern, 1988) if connections were made for the benefit the individuals involved, rather than for the organization as a whole. Exclusive cliques, the hiring of important individuals' relatives, and promotions based on "who you know" are all signs of an organization whose informal network is harmful to its operations.

The success of cross-functional teams spurred a transition toward "virtual organizations" — those that rely on such teams to produce their work product (Ahuja & Carley, 1999). As multinational corporations began adopting characteristics of virtual organizations, however, their leadership quickly learned they needed to adjust their management structures to remain effective (Maznevski & Chudoba, 2000). As organizations and suborganizations evolve into hybrids of Weberian bureaucracies and groups of cross-functional teams, they must ensure that their leaders and members have a sense of what is possible, as well as of the strengths and weaknesses of different team organization strategies (Bergiel, Bergiel, & Balsmeier, 2008). As organizations become more virtual, however, presenting and using that information effectively remains a challenge (Contractor & Seibold, 1993; Huber, 1990, 1991). This work addresses how to use particular data to address these challenges in the context of managing a horizontal merger.

One of the defining characteristics of virtual organizations is that they rely on an element of trust. Without trust, the deliberate delegation of task execution without low-level supervision would be untenable. To create trust, reciprocity between and among employees is critical (Berg, Dickhaut, & McCabe, 1995). Reciprocity and trust are key mechanisms for building interpersonal relationships, as well as for establishing expectations and norms (Gouldner, 1960). Because a horizontally merged corporation is no longer in competition with its source organizations, cooperation — not competition — is necessary and expected. Helping those who help, and avoiding harm to those who have helped, are norms (Gouldner, 1960) that need to supersede the previous mindset of competition. The available data does not quite allow analysis of whether the measured reciprocity is a byproduct of personnel managing their image (Tyler & Tang, 2003), but I'll discuss the relevant observables and implications shortly.

In social network analysis and organizational literature, a dominant characteristic of reciprocity is its dichotomous nature — i.e., you can be reciprocal to me, but I may not be reciprocal to you — and the implications and consequences of that dichotomy (Baker, Day, & Salas, 2006; Berg et al., 1995; Boyd & Richerson, 1989; Burgoon et al., 1993; Coyle-Shapiro, 2002; Criswell, 1946; Gouldner, 1960; Katz, 1955; Kenis & Knoke, 2002; Seeley, 1948; Tyler & Tang, 2003). However, because it is dichotomous and connected with status, reciprocity between individuals is difficult to measure over time using traditional survey methods. The rise of virtual, technology-enabled organizations is closing that gap, because these organizations now store empirical interaction data in well-structured databases and can make them available for over-time analysis.

Over-time analysis, supported by survey data, has indicated that some reciprocity is not about building trust — showing that you are a good teammate eager to help your coworkers — but in fact about performance and image management (Tyler & Tang, 2003), or ensuring that you look reliable to your managers. Other over-time data indicates that response time is predictive of the nature of the person-to-person tie (Wuchty & Uzzi, 2011). I developed windowed reciprocity as a way to take into account both of these findings and apply them to the study of the organization's informal network and how it changes over the course of a horizontal merger.

MergedCo as a Hybrid Organization

The MergedCo organization, like each of its legacy source organizations, is a modern hybrid of a Weberian bureaucracy and a virtual organization. It has a demonstrable formal structure, clear reporting lines, defined areas of responsibility, and other markers of bureaucracy. However, the organization also makes deliberate efforts to create and develop cross-functional teams, uses communications and technology to enable task execution across geographic regions and time zones, and empowers decision-makers at lower echelons — hallmarks of a virtual organization.

The transition from Weberian bureaucracies to virtual organizations has exacerbated a common stumbling block that hinders the success of horizontal mergers. Horizontal mergers — traditionally envisioned as the merging of two organizations and their formal networks, based on the input of the C-suite — are often failing to achieve their goals in the modern age. Merger management professionals have often attributed these failures to 'human factors' (Cartwright, 2002; Cartwright & Cooper, 2014; Marks & Mirvis, 2001). These factors include the interplay between the organization's formal and informal social networks (Frantz, 2012) and its cultural, tacit, and explicit knowledge networks (Frantz & Carley, 2009).

In this work, I explore the use of dynamic network analysis to characterize the informal network of an organization undergoing a horizontal merger. Email records provided by MergedCo represent one communication network, where each node of the network is an employee, and each edge is a directed edge representing emails sent between nodes. Edge weights represent the number of emails sent between the nodes. Because employees often use email to communicate both information and directives relevant to a changing organization, I will use the characteristics of the email network over two time periods to evaluate the impact of the merger.

Data

MergedCo provided both email data and employee survey data from two distinct time periods after the merger announcement. I will first describe how I prepared the email data, then follow up with the survey data. The email data included 1,717,021 distinct emails. The messages were systematically anonymized but left the entirety of the metadata and message content intact. The survey data, collected at corporate headquarters, is from the same distinct time periods as the emails. A total of 2,159 persons responded to the survey, and their participant IDs were cross-linked with unique email IDs so that I could correlate survey measures and node-level network measures.

I note that to further anonymize survey data, I stylized names of the functional groups rather than using the same specific labels as the corporation.

Email Data

The first time-period I examined, Time-1, is approximately three months into the formal merger process and lasts thirteen weeks. The second time-period, Time-2, is nine months later and lasts four weeks. The final corpus of emails includes 1,263,320 emails from Time-1 and 453,701 emails from Time-2. Email senders and recipients were anonymized prior to processing by assigning them new unique identifiers (i.e., To, From, cc). I also performed systemic anonymization of all nouns within email bodies, as described in Appendix A and discussed more in Chapter 2. I removed emails from the count if the recipient was identical to the sender, because these emails represent a form of record-keeping rather than interpersonal communication.

I used Apache Tika (Apache Software Foundation, 2016), a natural language processing toolkit developed at Stanford University, to identify nouns for anonymization and to identify the language of emails. MergedCo has significant corporate representation in countries where English is not the primary language. Because I used natural language processing tools trained on English texts, and I needed to be certain I was removing sensitive corporate information and personally identifiable information (PII), I removed emails that Tika could not positively identify as being in English.

Figure 1 is a stacked distribution of emails by language, as identified by Tika. The *x*-axis indicates email body length in characters, and the *y*-axis indicates how many emails fell into that length and language category. The preponderance of emails in the data set — especially the longer emails — are in English. This pattern is not unusual for multinational firms whose employees are predominantly from or located in a single country (in this case, the United States).[1] As shown in Figure 1, Tika identified most emails as being in English.

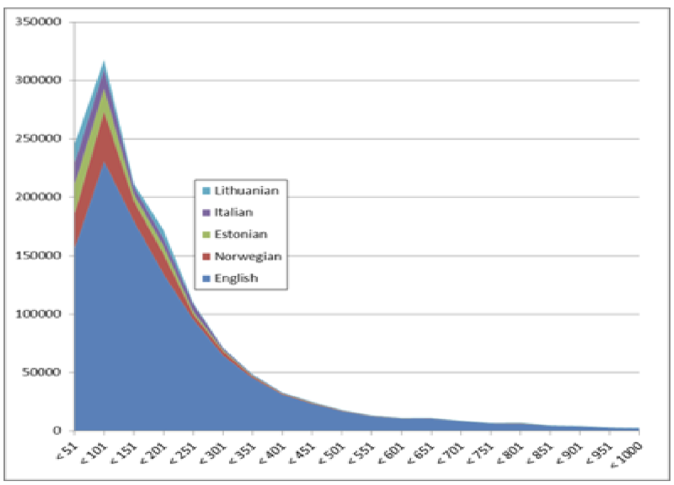


Figure 1. Distribution of email languages by content length

Not all emails are created equal. I wanted to limit the set of emails to interpersonal communications between members of MergedCo, so I established three filter criteria:

- That the email be primarily in English, as identified by Apache Tika

[1] There are unavoidable limitations to using only English only emails. This analysis was unable to focus on the structural implications for individuals who did not write their emails in English, and thus may unfairly privilege the English-fluent members of the organization. Unfortunately, the need to maintain privacy and confidentiality must outweigh other concerns.

- That the number of recipients be fewer than seven and not include MergedCo's mailing lists
- That the sender of the email is an individual, not a mailing lists

With these limitations in place, I tested that the filter criteria had not introduced problematic bias by examining the distribution of the emails over time. In a typical large organization, email traffic is much heavier over the course of the workweek and lighter during weekends (Karagiannis & Vojnovic, 2009; Tyler & Tang, 2003). In Figure 2, I show how Time-1 and Time-2 emails were distributed across time. In Time-1, there are thirteen (13) spikes, which correspond to thirteen workweeks. There are two deeper-than-typical dips — one between the 4th and 5th workweeks, and one between the 12th and 13th workweeks. These correspond to United States holidays. In Time-2, there are four workweek spikes, which matches my expectation. These patterns are congruent with those found by other analyses (Karagiannis & Vojnovic, 2009).

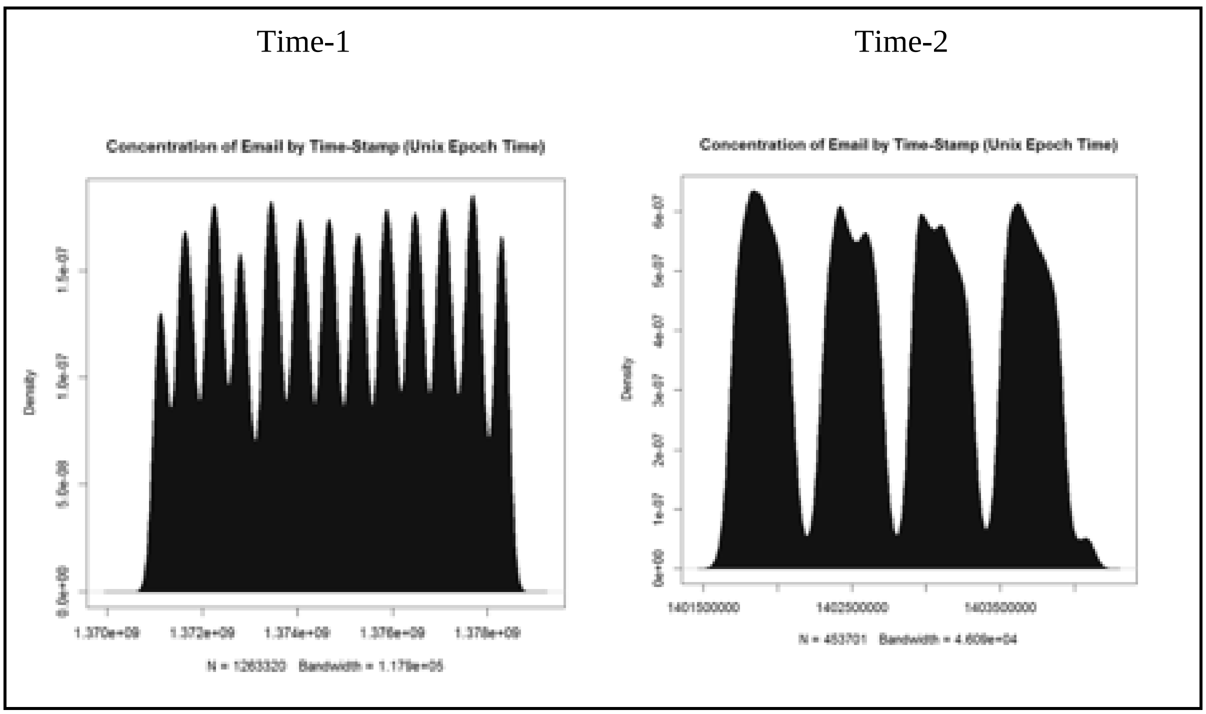


Figure 2. The distribution of emails over time during the key time-periods

Survey Data, an Indicator of Organizational Turbulence

As mentioned earlier, MergedCo conducted annual employee surveys over the course of the merger. I report demographic characteristics of participants in the Time-1 and Time-2 surveys in Table 1. Slightly more than 1,600 participants submitted responses to each survey.

Table 1. Survey demographics[2]

Feature	Time-1	Time-2
n	1,660	1,693
Minority Status	10.2% reported minority status, 39.2% preferred not to say	11.8% reported minority status, 36.3% preferred not to say
Gender (M/F)	37.8% M, 23.1% F 39.1% preferred not to say	38.6% M, 25.2% F 36.2% preferred not to say
Age (years)	Mean: 44 (M), 43 (F), 49 overall 39.2% preferred not to say	Mean: 45 (M), 44 (F), 44 overall 36.2% preferred not to say
Tenure (years)	Mean: 9.7 (M), 7.6 (F), 8.9 overall 39.1% preferred not to say	Mean: 8.3 (M), 7.5 (F), 8.0 overall 36.3% preferred not to say

Exogenous to the data, I know that employees departed both legacy organizations between Time-1 and Time-2, which precluded their participation in the Time-2 survey. Also, as noted earlier, MergedCo hired new employees after merger activities began; some of those employees took the survey in Time-2 but not Time-1[3].

Survey participants in both time periods also provided their current functional groups. In a stable organization, changing functional groups from year to year is relatively rare (Fisher, 2013). Because of the merger, however, many MergedCo survey respondents changed functional groups in the year between time periods of the survey. Of the 1,194 participants who took the survey in both periods, 390 (32.7%) changed functional groups between Time-1 and Time-2. Because this was not a complete census of the organization, however, I will focus on larger functional group transitions that are less likely to be artifacts of sampling.

Viewed from Time-2, MergedCo created five new functional groups and filled three existing ones with majority new staff from other functional groups. The five new functional groups — again with stylized names — are Manufacturing, Accounting, Forecasting and Planning, Direct-to-Consumer (DTC), and Financial Planning Analysis. The three functional groups with more than half of their sampled staff coming from other functional groups are Strategic Logistic Solutions, Sourcing, and Product Engineering & Safety.

Between Time-1 and Time-2, MergedCo reorganized several existing functional groups and broke up others to form new functional groups. Of functional groups in Time-1 that still existed in Time-2, Sourcing, Finance, Marketing, and Product Engineering & Safety experienced the most disruption. Sourcing experienced the most change, with 95.3% of its original staff leaving

[2] The prevalence of 'preferred not to say' answers to demographic questions made it infeasible to discern any clear patterns of respondents along multiple dimensions.
[3] The available data does not support meaningful analysis of the distribution of the new hires and employee departures across the functional areas.

the functional group. Most of them moved to a new functional group called Manufacturing. Finance, similarly, lost over two-thirds of its staff between Time-1 and Time-2. Most of those employees moved to a new Accounting functional group. Marketing was broken apart, with many employees moving to the new DTC functional group. Finally, many staff members from Global Product Engineering & Safety moved to the Sourcing functional group by Time-2.

Table 2, below, summarizes major transitions in functional roles. Time-2 functional groups are identified alongside a ranked list of their contributing functional groups from Time-1.

Table 2. Time-2 Functional groups as the end result of changes from Time-1 functional groups.

Time-2 Functional Group	% Change from Time 1	Major Time-1 Source Functional Groups (Ranked)[4]
Accounting	100%	Finance, Other
Direct to Consumer (DTC)	100%	Marketing
Financial Planning Analysis	100%	Finance
Forecasting and Planning	100%	Sourcing, Strategic Logistic Solutions
Manufacturing	100%	Sourcing, Other
Sourcing	95.3%	Product Engineering & Safety, Sourcing
Product Engineering & Safety	77.8%	Product Engineering & Safety, Sourcing, Product Development
Logistic Solutions	70.6%	Sourcing, Other, Logistic Solutions

Merging Email Metadata and Survey Data to create Meta-Networks

I used the email data and survey data to create a dynamic meta-network representing the organization. A dynamic meta-network is a set of related meta-networks, each of which provides a snapshot of the organization at a specific time (called a key frame). This dynamic meta-network includes two key frames: Time-1 and Time-2. To make the analysis tractable and easier to compare to the yearly surveys, I chose to aggregate emails across time periods rather than dividing them by days or weeks.

A meta-network is distinct from a traditional social network analysis network because it is both multimodal (composed of distinct sets of entities, such as agents and organizations) and multiplex (including multiple relationship types, such as membership and proximity). This makes it possible to analyze multiple relationships simultaneously and identify connections between various relationship types. Meta-networks are therefore particularly valuable for understanding an organization’s formal and informal networks because they provide an

[4] Time-1 Source functional groups are listed in order of contribution to the new functional group.

opportunity to evaluate many relevant factors, including individual and group access to resources, information, beliefs, and other people (Carley, 2003; Tsvetovat, Reminga, & Carley, 2004).

In this analysis, I focus on the relationships among two node types: agents and organizations. Individual employees are represented as agents. Functional groups and each of the three legacy organizations are collections of individual agents. I attribute network structure to organizations based on employee membership and each agent's activity, so if:

- Agent *a* is a member of Functional Group *G1* and Legacy *L1*
- Agent *b* is a member of Functional Group *G2* and Legacy *L2*
- Agent *c* is a member of Functional Group *G2* and Legacy *L3*
- *a* communicates with *b*
- *b* communicates with *c*

then when interactions are grouped by functional group, *G1* will have a link (weight 1) with *G2* based on the interaction *a* and *b*, and *G2* will have a self-link (also weight 1) from the interaction of *b* with *c*. Similarly, when grouped at the legacy organization level, *L1* will have a weight-1 link with *L2*, and *L2* will have a weight-1 link with *L3*.

In matrix multiplication notation, if agents are represented by the set A, functional groups by the set G, and legacy organizations by the set L, then this aggregation can be represented as shown in Equation 1.

Equation 1. Matrix Multiplication to create Functional Group and Legacy Aggregations

$$GG = (AG)^T * AA * AG$$

$$LL = (AL)^T * AA * AL$$

Combining the survey and email data into one sample is particularly challenging because there are many individuals who were not surveyed in either time period but who are nonetheless active communicators at MergedCo. However, everyone who participated in a survey did send at least one email included in the data set.

Because individuals in the organization may frequently communicate with external actors who are not relevant to this analysis (vendors, competitors, potential hires, family members, etc.), I removed individuals who were at the periphery of the network. I defined peripheral individuals as those who communicated with fewer than 3 others (< 3 alters) within the organization over the course of each time period. This filtering removed many external actors and drastically reduced the size of the network to roughly the actual size of the organization, as reported in exogenous interviews with the CEO.

Table 3. Combined Network and Survey Data, entity counts who fit various criteria

Time	All Unique IDs	Took Survey (%)	Core Unique IDs	Core Survey Takers (%)
Time-1	16,374	867 (5.3%)	6,589	847 (12.9%)
Time-2	19,046	2,182 (11.5%)	7,373	2,122 (28.8%)

In each case, removing non-core actors more than doubled the percentage of those within the network who had taken the survey, suggesting that the survey successfully focused on people who were at the heart of MergedCo in both time periods. Limiting the analysis to the core removed only a small percentage of survey takers (~ 2.5% of survey takers in each case).

Figure 3, below, shows the network as it evolved between Time-1 and Time-2. Red nodes listed LuxuryCo as their affiliation, teal nodes listed StandardCo as their affiliation, and yellow nodes listed MergedCo as their affiliation. Links represent emails to or from each person. Comparing the two networks reveals that MergedCo has integrated much more deeply into both StandardCo and LuxuryCo in Time-2 compared to Time-1.

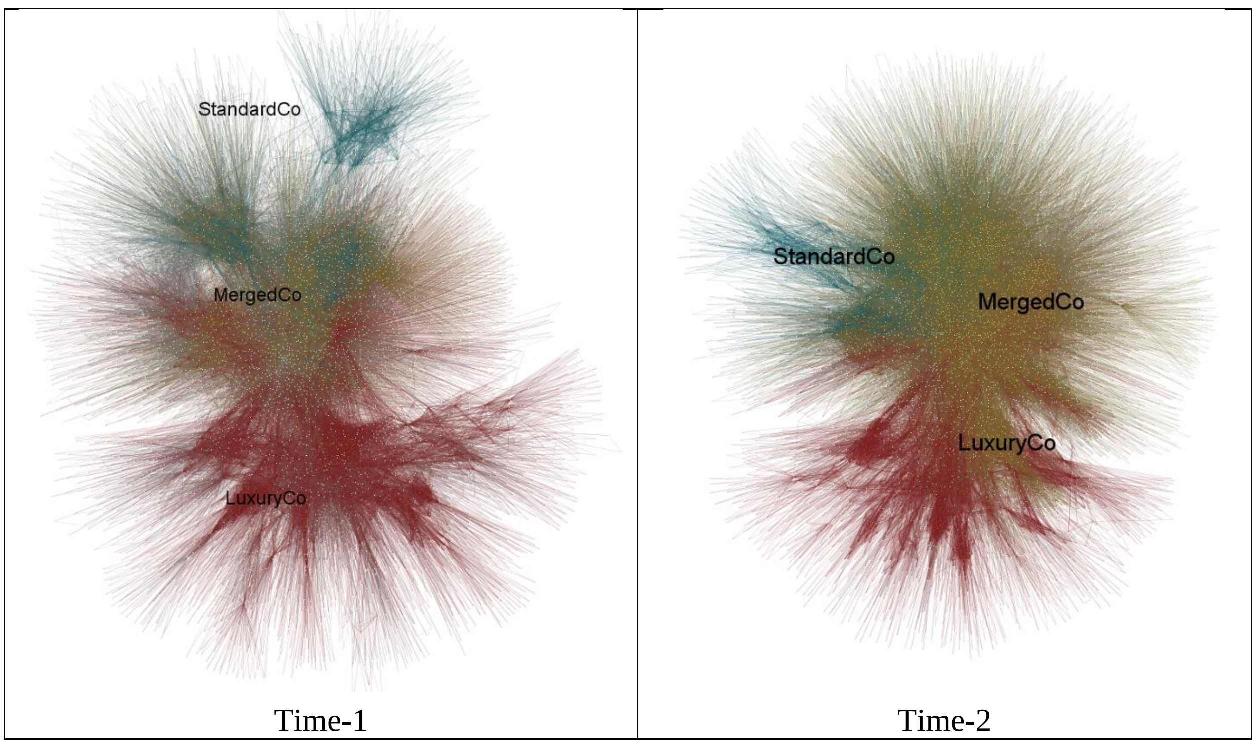


Figure 3. The network of MergedCo at Time 1 and Time 2.

Methods

In this section, I describe the methods I used to answer the questions identified in the Introduction. In addition to the standard network analysis methods that are applied to dynamic meta-networks, I also conducted inter- and intra-organizational network analysis. I explain how I computed my new measure — windowed reciprocity — and describe it in more detail. Finally, I conclude with a list of key survey outcomes I use to validate the findings regarding legacy organizations and functional groups.

Network Analysis Measures

I posit that leaders of organizations undergoing horizontal mergers would find information about their informal leaders and informal leadership networks interesting, especially as those networks evolve within the legacy organizations and new informal leaders rise within the integrated environment of MergedCo.

To characterize these individuals, I focus on three common social network analysis measures:

- In-Degree: The number of alters who, during the time period, sent at least one email to this individual. High in-degree shows that the individual receives messages from many different people.
- Out-Degree: The number of alters to whom, during the time period, this individual sent at least one email. High out-degree shows that the individual sends messages to many different people.
- Betweenness Centrality: The number of shortest paths this individual is on across the entire organization. In a binary network, the shortest path is the connection between any two individuals in the network that requires the fewest jumps through other individuals. High betweenness centrality — or being on many shortest paths — indicates that an individual is a broker or gatekeeper of information within the organization.

Using these methods, informal leaders of legacy organizations would be identifiable and could therefore be leveraged to assist in a successful merger. The informal leadership dynamics identified in Time-1 would likely change by Time-2, and MergedCo's corporate leadership would need to stay aware of those changes. Moreover, people or groups who lost power in the informal dynamics over time could contribute to a microculture of discontent.

In this context, the key entities report of the Organizational Risk Analyzer (ORA) is ideal for rapidly identifying key nodes across multiple networks and multiple measures (Altman, Carley, & Reminga, 2018; Carley, 2017). A high ranking in the key entities report indicates that a node is among the top ten nodes across more than 20 network measures.

Over time, the collection of informal leaders in MergedCo would grow to include employees who were hired since the merger began. With only a year's worth of data, however, there is no expectation that these MergedCo leaders would rise to the same level of importance as those rooted in the legacy organizations.

Windowed Reciprocity

In every context, individuals will receive messages from others. They will be predisposed to view or ignore these messages as they arrive depending on who is sending them. Unlike other acts for which reciprocity is often measured, such as the sharing of valuable information or favorable commercial transactions, modern communications are often immediately responded to or not at all (Wuchty & Uzzi, 2011). Many unknowable factors can affect the speed of an "immediate" response, including the recipient's meeting schedule, whether they need to get information from a third party before responding, and whether organizational actors follow up with each other. I therefore allow a response within twenty-four hours[5] to count as "immediate."

I will measure the predisposition to view or ignore a message as a likelihood of response within twenty-four (24) hours — the window — as it applies to messages in both Time-1 and Time-2. Between any two alters, there are three common implications from the two possible values of likelihood of response. If X responds frequently to Y and Y responds frequently to X, then X and Y are likely collaborative coworkers who routinely complete work together. If X responds frequently to Y but Y does not respond frequently to X, then X likely reports to or is inferior in status to Y; at the very least, Y is not concerned with being responsive to X. If both X and Y are nonresponsive to each other, they are probably not close coworkers but are instead messaging each other for some exogenous reason — perhaps because a manager suggested they work together, or because it is X's job to notify individuals like Y about some aspect of X's job.

Measuring windowed reciprocity is straightforward. For every message from X to Y, is there a message within a given window (e.g., 1 hour, 12 hours, 24 hours, 1 week) from Y to X? I do not attempt to determine whether a given message is a response to another specific message, because that would be difficult and likely to produce noise, and it would ignore the reality that a series of messages may all be related to the same idea, or a single message may be a response to multiple unrelated messages.

More formally, reciprocity is sum of all replies, r, within a given window, w, divided by the sum of all messages, m, between the dyad ij. When i sends a message j at time t, there is a reply to i if j sends a message to i within time $t + w$.

Equation 2. Reciprocity for person i for time period t is the number of replies within time window, w, against the number of sent messages

$$Reciprocity_{ij}^{p} = \frac{\sum r_{ijt}}{\sum m_{ijt}},\ m_t + w > r_t > m_t$$

This reciprocity value defines a network link and is therefore dyadic, in that it measures a statistic per dyad. When aggregated at the person level, reciprocity values represent that person's response rate to all others. When aggregated at the group level, I retain the dyadic aspect and consider the reciprocity of a group to other groups. I aggregate at two group levels: the original legacy organization reported by the individual, and the functional group to which the employee belongs. I measure windowed reciprocity for all dyads and individuals to establish baselines, but most of my analyses consider only surveyed individuals.

In later sections of this paper, I compare the 'raw' email network — where I don't remove ties that are nonreciprocal based on windowed reciprocity — to the windowed reciprocity network, where I do remove these ties.

[5] The distributions of reciprocity values were insensitive from one (1) hour up to twenty-four (24) hours. The distribution begins to look different at 72 hours, and at 1 week (168 hours) it is quite distinct from the 1-24 hour windows.

Survey Measures

I pulled eight measures from the survey data that I think may be influenced by either legacy organization or functional group membership. All measures were averaged aggregations of multiple survey questions — both positively and negatively correlated with the concept in question — and were scored on a 7-point Likert scale ranging from 1 (strong disagreement) to 7 (strong agreement). I selected only those questions from the survey that had sufficient response rates. Several of the aggregated measures are taken from existing survey instruments. These are:

- NeedForChange: From a study of organizational citizenship behaviors (Van Dick, Grojean, Christ, & Wieseke, 2006), NeedForChange measures the perception that the organization must change to succeed. The concept, as a whole, can be described as agreement with the statement, "Our organization needs to change."
- MergComm: This measure was adapted from an existing measure of goal commitment (Klein, Wesson, Hollenbeck, Wright, & DeShon, 2001). The concept can be described as agreement with the statement, "I am committed to this merger."
- DistJust: A measure of distributive justice (Niehoff & Moorman, 1993), this considers whether the individual perceives the outcomes they receive to be fair. The concept can be described as agreement with the statement, "I am rewarded fairly by this organization."
- Satis: A measure of job satisfaction (Cammann, 1983), this considers whether the person is content with their work. The concept can be described as agreement with the statement, "I am satisfied with my job."

Other statement prompts were written to support this particular research. These are:

- ClearExp: How well the person believes they understand their job duties. The concept can be described as agreement with the statement, "I understand what is expected of me."
- SuperFair: How fairly the individual feels their supervisors treat people on the basis of inherited characteristics (e.g., gender, age, race). The concept can be described as agreement with the statement, "My supervisors treat people fairly."
- SuperInfo: How well the individual feels their supervisors provide accurate and timely information that the individual needs to do their work. The concept can be described as agreement with the statement, "My supervisors provide me the information I need."
- Recognition: Whether the individual feels they receive recognition for the work they perform. The concept can be described as agreement with the statement, "I receive recognition for the work I do."

When I aggregate data at the legacy organization level, I treat membership in a legacy organization as a categorical variable with three possible values: "MergedCo," "LuxuryCo," and "StandardCo."

Results

In this section, I cover the four areas of results. In the first section, I analyze the changes in intra-group links between Time-1 and Time-2 at both the legacy organization and functional group levels. In the second section, I analyze changes in reciprocity between Time-1 and Time-2 at both the legacy organization and functional group levels. In the third section, I compare how raw email node-level network measures and windowed reciprocity node-level network measures correlate to important survey outcomes. Finally, in the fourth section, I discuss the identification of key entities who are negatively affected by the changes in the organization between Time-1 and Time-2.

Intra-Organization Link Analysis Based on Raw Email Data

As we saw in the visual representation of the raw email networks in Figure 3, it is evident that the merged organization has undergone substantial change between Time-1 and Time-2. An aggregate analysis of links between the legacy organizations, summarized in Table 4, shows similar results.

Table 4. Intra-Legacy Org Messages for Time-1 and Time-2

	% of Messages Intra-Org		
LegacyOrg	Time-1	Time-2	Change
StandardCo	51%	33%	-18%
LuxuryCo	86%	59%	-27%
MergedCo	59%	79%	+20%
Overall	73%	70%	-3%

In comparing Time-2 to Time-1, I can see that the MergedCo organization — composed of people hired after the merger — has become increasingly self-contained, leaning less on expertise of actors that identify with the original LuxuryCo and StandardCo. By contrast, both StandardCo and LuxuryCo actors are interacting more and more outside of their original organizational envelopes. Overall, the proportion of links that are within one legacy organization remains similar, but in Time-2 their distribution has shifted in favor of MergedCo at the expense of both LuxuryCo and StandardCo.

From a management perspective, these trends appear to be healthy. Understanding what motivates LuxuryCo employees to direct 2/3 of their communications to fellow LuxuryCo alumni would require more pointed questions in surveys or other means of gathering data. Chapter 2, where I use text analysis to understand the organization's changing culture, will delve more into these ideas.

By incorporating the survey data, I was able to identify eight functional groups that experienced substantial change during the merger. These functional groups were either newly created or the result of a major reorganization effort, as the majority of their staff in Time-2 had been in different functional groups in Time-1. In Table 5, I list the percentage of intra-org email connections in Time-2 for each of these changed functional groups.

Table 5. Time-2 functional groups and their intra-org links

Time-2 Functional group	Change	% Intra Time-2
Direct to Consumer (DTC)	Created	40%
Manufacturing	Created	33%
Logistic Solutions	Reorganized	23%
Accounting	Created	23%
Sourcing	Reorganized	16%
Forecasting and Planning	Created	15%
Financial Planning Analysis	Created	11%
Product Engineering and Safety	Reorganized	9%

Notably, all these functional groups have intra-org percentages below the entire organization's rate of 44% — most of them far below. This reflects the behaviors I would expect of a virtual organization coping with abrupt change, as individuals reach out across newly drawn organizational boundaries to complete work together.

This analysis shows me how the organization is evolving over the year between the two time periods. I see that the MergedCo organization is becoming more self-contained, but individuals in new functional groups are reaching out to others outside their functional groups to complete their work. LuxuryCo, which was almost entirely self-interested in Time-1, is now sending about 40% of its messages to others. StandardCo — the acquired company and perceived junior partner at Time-1 — is becoming well integrated into MergedCo by Time-2. The functional group analysis shows that the groups that experienced significant change — either being created or massively reorganized — are still recovering from that change, as individuals within these groups communicate internally much less than average.

Inter-Organization Windowed Reciprocity Analysis

Using the calculation of reciprocity, I am able to measure the reciprocity of individuals to other individuals and aggregate that information to both the legacy organization and functional role levels. These networks will play a large role in Chapter 2 and Chapter 3.

For these analyses, I used a 24-hour response window. In selecting a 24-hour response window, I may be ignoring some Monday responses to Friday emails. This has a negligible effect, however, since response patterns within 1 hour and within 24 hours are similar. As we can see in Figure 4, if a person is going to get back to you, it will typically be within hours.

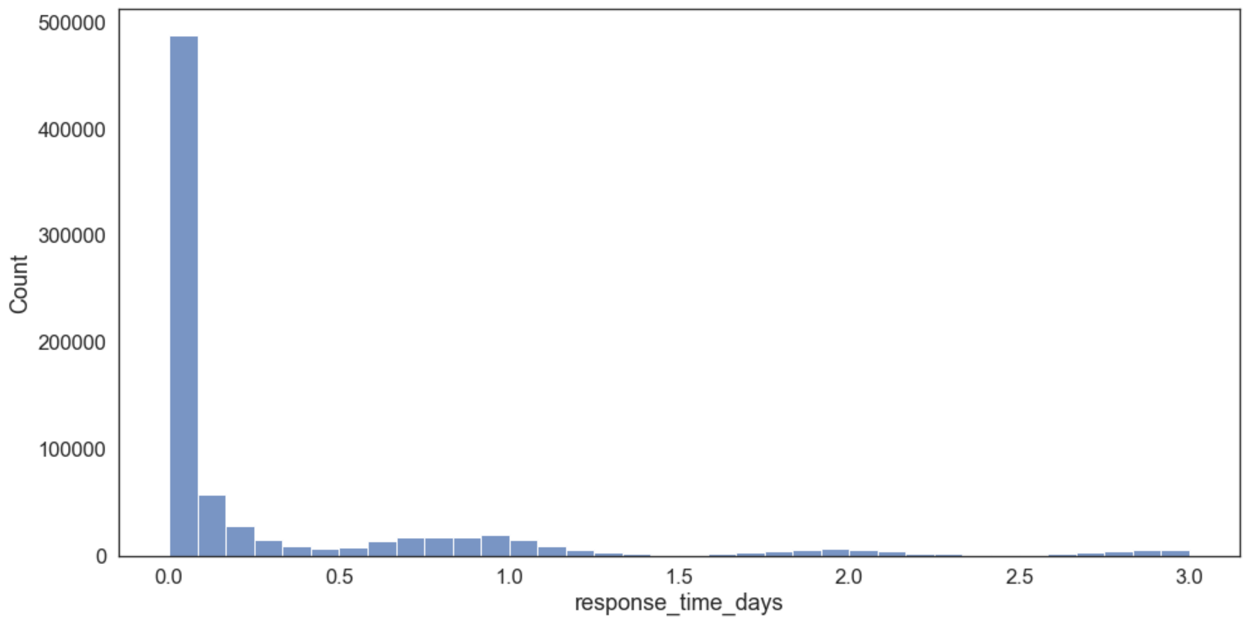


Figure 4. The distribution of response times to individual messages at Time-1 from at least one recipient, up to 72 hours or 3 days

Furthermore, in most analyses that rely on the binarized network, any single communication that resulted in a response within 24 hours is sufficient to create a tie. There would be cause for concern if there were strong evidence that a large number of individuals interact *only* on Friday-to-Monday communication cycles. If that were the case, the analysis would lose these connections, which would affect the overall observed structure of the informal network. However, that communication pattern is unlikely to be common — and I would also argue that choosing to send an email late on a Friday is rarely coincidental. At such, choosing to 'skip' the weekend when measuring response time may actually distort the understanding of the network.

Legacy Organization Windowed Reciprocity

The intra-organizational analysis in the prior section showed me that MergedCo, LuxuryCo, and StandardCo primarily communicated within their organizational boundaries. Using windowed reciprocity, I wanted to compare the interactions between these organizational units in terms of how many messages were responded to within a given time frame.

Because the measure of reciprocity is inherently dyadic, I created a network of all individuals and then aggregated the link weights together based on legacy organization affiliation. I then normalized those link weights by the row-sum for the legacy organization. Table 6 shows the resulting matrices for both Time-1 and Time-2.

Table 6. Legacy Organization response rates to each other normalized by their row-sum in both Time-1 (left) and Time-2 (right). The row indicates the sender, while the column indicates the receiver.

Time-1	MergedCo	LuxuryCo	StandardCo
MergedCo	64.37%	22.06%	13.57%
LuxuryCo	10.97%	87.72%	1.31%
StandardCo	41.56%	8.04%	50.40%

Time-2	MergedCo	LuxuryCo	StandardCo
MergedCo	81.22%	9.54%	9.24%
LuxuryCo	33.34%	64.57%	2.10%
StandardCo	60.49%	3.57%	35.94%

In considering these two matrices, I see that the organization has changed between Time-1 and Time-2. LuxuryCo, the acquiring company, is being subsumed into the MergedCo identity by Time-2, with members of all three organizations responding in a timely fashion to LuxuryCo members less frequently than they did in Time-1. StandardCo, similarly, is responded to less frequently by both MergedCo and StandardCo, but LuxuryCo members become slightly more responsive. By contrast, everyone becomes more responsive to MergedCo by Time-2.

Functional Group Windowed Reciprocity

Table 2 and Table 5 described the functional groups that had undergone substantial change from Time-1 to Time-2. In Table 7, I report the intra-group reciprocity of these functional groups, along with the alter functional group to which they are most responsive.

Table 7. Disrupted Time-2 sroups, their intra-group reciprocity, and the Time-2 group to which they are most responsive.

Time-2 Group	Intra-Group Reciprocity	Most Responsive To (%)
Direct to Consumer (DTC)	67%	Customer Service (19%)
Logistic Solutions	50%	Manufacturing (13%)
Accounting	46%	Corporate Finance (13%)
Manufacturing	41%	Accounting (10%)
Sourcing	39%	Retail Sales (8%)
Forecasting and Planning	35%	Manufacturing (20%)
Financial Planning Analysis	35%	Corporate Finance (21%)
Product Engineering and Safety	31%	Manufacturing (35%)

If I measure the windowed reciprocity of the other, non-disrupted Time-2 groups, I see that the average intra-group windowed reciprocity is 55%, while the average intra-group reciprocity of these changed groups is lower at 43%. These results suggest to me that many of these groups, particularly those toward the bottom of the table, are still learning how to function as coherent groups. Product Engineering and Safety is particularly interesting, as they are now most responsive to Manufacturing — more so than they are to themselves, in fact. Both Manufacturing and Product Engineering and Safety acquired many of their new people from the Time-1

Sourcing group (which was almost entirely composed of new individuals by Time-2), so perhaps these individuals are still interacting frequently, even across new organizational boundaries. It would not be surprising if the company decides to reorganize Product Engineering and Safety to better align with Manufacturing.

Table 7 listed each functional group's intra-group reciprocity and its most responsive alter. Figure 5, below, depicts the complete matrix across functional groups. The diagonal represents intra-group reciprocity. The darker the cell, the more reciprocity between the row and column cells.

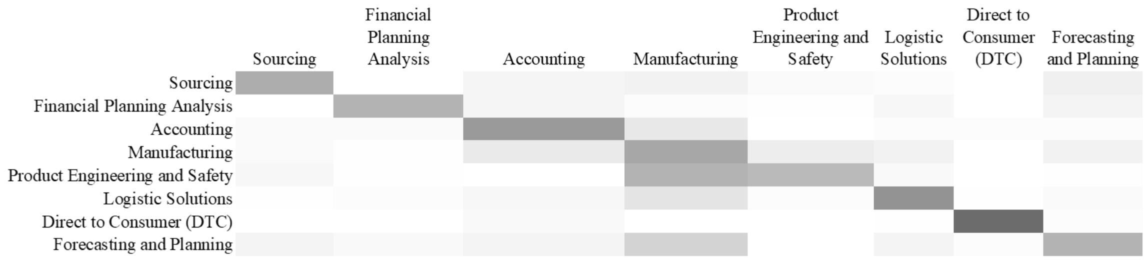


Figure 5. Heat map of normalized windowed reciprocity across functional groups

As one would expect, intra-group reciprocity (on the diagonal) is almost always the highest. The only exception, as previously discussed, is that Product Engineering and Safety is more responsive to Manufacturing than to themselves.

Figure 6 is a network visualization of reciprocity for most Time-2 functional groups. Links are weighted by aggregate reciprocity. Links have arrows, indicating that X is often responsive to Y.

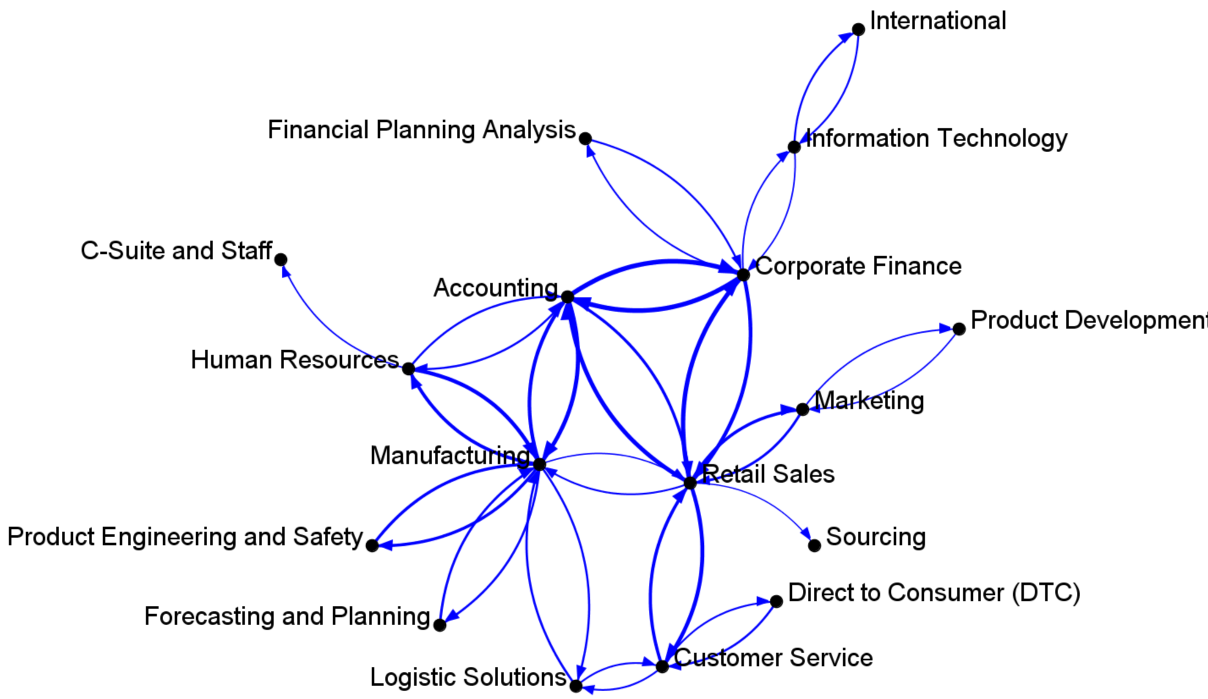


Figure 6. Windowed reciprocity of Time-2 functional groups.

In Figure 6, I can see two front-facing functional groups — Manufacturing and Retail Sales — that connect with many other functional groups. Presumably, these two functional groups drive forward and coordinate key operations of the organization. I also see two back-office groups, Accounting and Corporate Finance, that connect to many other groups across the organization. If I interpret high windowed reciprocity as indicating a high likelihood of information flow, then all other functional groups have very limited reach within the organization. Finally, two links are asymmetric: Human Resources, not surprisingly, has a high response rate to messages from C-Suite and Staff, but not vice versa. Similarly, Retail Sales has high response rate to messages from Sourcing, but not vice versa.

Figure 5 also reveals an interesting four-group communication cycle between Manufacturing, Retail Sales, Customer Service, and Logistic Solutions. This cycle is central to delivering value to MergedCo's customers, and inside MergedCo it is colloquially referred to as the "Make it, Sell it, Ship it, Support it" function.

Comparing Legacy and Functional Group Survey Outcomes

In this final subsection, I examine survey outcomes at both the legacy organization and functional levels. I analyzed eight (8) measures drawn from the larger survey.

Legacy Survey Outcomes between Time-1 and Time-2

In the two prior result sections, I reported the dynamics of the organization's change as measured by how group members interact. The network analysis showed MergedCo emerging as a self-contained entity, with more actors at LuxuryCo and StandardCo needing to interact with MergedCo stakeholders in order to do their work. The reciprocity analysis showed not only that everyone is more responsive to MergedCo in Time-2 than in Time-1, but also that individuals became less responsive, overall, to LuxuryCo individuals.

I expected that I would see corresponding trends in the survey measures for each legacy organization between Time-1 and Time-2. I used Tukey tests to compare the means of each outcome variable for the legacy organizations. In the far right column of Table 8, I report the lowest p-value reported by Tukey for each measure and time period. Below the values for each measure, I show the delta between Time-1 and Time-2.

Table 8. Survey outcomes for Time-1 and Time-2 for each legacy organization[6]

Measure and Concept	Time	MergedCo	LuxuryCo	StandardCo	Sig.
NeedForChange "Our organization needs to change to succeed."	Time-1 $\bar{x}$	4.65 (1.54)	3.49 (1.32)	3.71 (1.74)	< 0.001
	Time-2 $\bar{x}$	4.70 (1.49)	3.82 (1.47)	4.41 (1.44)	< 0.001
	Time Δ	+ 0.05	+ 0.33	+ 0.70	

[6] Because LuxuryCo-affiliated individuals often chose not to respond about their commitment to the merger in Time-2, the interpretation is unreliable here — presumably, only people who were committed to the merger chose to respond.

Measure and Concept	Time	MergedCo	LuxuryCo	StandardCo	Sig.
MergComm "I am committed to this merger."	Time-1 x̄	5.47 (0.95)	4.74 (1.06)	5.16 (0.88)	< 0.001
	Time-2* x̄	5.72 (0.83)	5.34 (0.99)	5.50 (0.94)	< 0.001
	Time Δ	+ 0.25	+ 0.60	+ 0.34	
		MergedCo	LuxuryCo	StandardCo	Sig.
DistJust "I am rewarded fairly by this organization."	Time-1 x̄	4.66 (1.22)	4.78 (1.15)	5.04 (1.15)	0.03
	Time-2 x̄	4.82 (1.26)	4.81 (1.26)	4.65 (1.31)	0.23
	Time Δ	+ 0.16	+ 0.03	- 0.39	
ClearExp "I understand what is expected of me."	Time-1 x̄	5.72 (1.37)	5.94 (1.10)	5.98 (1.44)	0.14
	Time-2 x̄	5.88 (1.16)	5.86 (1.23)	5.83 (1.12)	0.88
	Time Δ	+ 0.16	- 0.08	- 0.15	
SuperFair "My supervisors treat people fairly."	Time-1 x̄	5.69 (1.67)	5.42 (1.52)	5.78 (1.49)	0.07
	Time-2 x̄	6.08 (1.33)	5.32 (1.66)	5.77 (1.39)	< 0.001
	Time Δ	+ 0.39	- 0.10	- 0.01	
SuperInfo "My supervisors provide me the information I need."	Time-1 x̄	5.40 (1.68)	5.39 (1.49)	5.62 (1.60)	0.34
	Time-2 x̄	5.66 (1.33)	5.36 (1.60)	5.33 (1.62)	< 0.001
	Time Δ	+ 0.26	- 0.03	- 0.29	
		MergedCo	LuxuryCo	StandardCo	Sig.
Recognition "I receive recognition for the work I do."	Time-1 x̄	5.10 (1.56)	5.27 (1.47)	5.04 (1.71)	0.33
	Time-2 x̄	5.32 (1.53)	5.16 (1.56)	4.82 (1.68)	0.001
	Time Δ	+ 0.22	- 0.09	- 0.22	
Satis "I am satisfied with my job."	Time-1 x̄	5.60 (1.26)	5.85 (1.09)	5.90 (1.23)	0.07
	Time-2 x̄	5.82 (1.13)	5.79 (1.20)	5.41 (1.28)	< 0.001
	Time Δ	+ 0.22	- 0.06	- 0.49	

In Table 8, there is evidence for what I have inferred from the structural and reciprocity analyses. By Time 2, four measures — job satisfaction, recognition, supervisors being fair, and supervisors providing information — have diverged statistically significantly between legacy organizations, while they were not distinctive in Time 1. In three of these cases, MergedCo has gone up on the measure, while StandardCo has gone down. In the remaining case, supervisor fairness, MergedCo has gone up, while LuxuryCo has gone down.

On two other measures — the need for change and commitment to the merger — there were significant differences in opinion across legacy organizations in both Time-1 and Time-2. MergedCo employees saw a need for change in Time-1, while StandardCo and LuxuryCo employees did not. StandardCo agrees much more with the need for change by Time-2, however. MergedCo employees were highly committed to the merger in Time-1. Everyone's commitment to the merger has increased by Time-2.

Fewer than 20% of LuxuryCo respondents who answered the other survey questions were willing to answer questions related to their commitment to the merger. LuxuryCo initiated the merger, and by Time-2, it an irreversible fact of life at MergedCo. The many refusals to answer this question suggest to me that LuxuryCo individuals were no longer as committed to the merger as they had been, but did not feel comfortable reporting that fact.

Two measures remain. In perceptions of the organization's fairness, StandardCo believed the organization was fair slightly more than the other two legacy organizations during Time-1, but all three organizations are in agreement by Time-2. Finally, legacy organization seems to have little affect on whether the individual understands the expectations of their job.

To me it is clear that these survey outcomes reflect the major changes to the organization's informal network and how the three legacy organizations interact. With LuxuryCo's change from almost complete self-sufficiency to heavy reliance on MergedCo, LuxuryCo's employees are not convinced of the need for change, and many refused to answer questions on their commitment to the merger. Meanwhile, StandardCo employees' commitment to the merger and recognition of the need for change increased.

Comparing Survey Outcomes to Raw and Windowed Reciprocity Node-Level Network Measures

In addition to examining the mean values of survey responses for each legacy organization (MergedCo, LuxuryCo, and StandardCo), which showed me organizations that differ in their views of most of the survey measures, I wanted to know if the calculated measures of the raw email network and the windowed reciprocity network were correlated with survey measures at the individual level. Specifically, I believe that the windowed reciprocity network is more reflective than the unprocessed email network of how individuals experience work.

To evaluate this, I examined whether there are meaningful correlations between survey outcomes and the raw email and windowed reciprocity node-level network measures. Because the survey outcomes are ranked categorical data, I used Kendall's Tau, a rank correlation coefficient. In Table 9, I report Tau estimates and their significance for the structural and reciprocity measures for Time-2 survey outcomes at the individual level. Positive Tau estimates indicate that there is concordance between the input and the output (i.e., A and B go up together), while a negative Tau estimate indicates that there is discordance (i.e., B goes up when A goes down). The p-value indicates the statistical significance of the reported Tau, with bold font used for p-values less than 0.05. Significance for a given row-cell is indicated by + for p-values of 0.1, * for p-values < .05, ** for p-values < 0.01, or *** for p-values < 0.001. Entries are shaded by significance value as well.

Table 9. Tau estimates and p-value (in parentheses) of the structural and reciprocity measures for Time-2 survey outcomes at the individual level.

Survey Measure	Email In-Degree	Email Out-Degree	Email Betweenness	Reciprocity In-Degree	Reciprocity Out-Degree
Need For Change	**0.29*** **	**0.25*** **	**0.21*** **	**0.24*** **	**0.23*** **
Commitment to the Merger	**0.07** **	**0.04***	**0.05***	**0.09*** **	**0.09*** **
DistJust	**-0.07*** **	**-0.07*** **	**-0.07*** **	**-0.06*** **	**-0.06*** **
ClearExp	-0.04+	-0.03	-0.03+	-0.01	-0.00
SuperFair	**0.16*** **	**0.13*** **	**0.08*** **	**0.11*** **	**0.12*** **
SuperInfo	0.03	0.02	0.00	0.03+	0.03+
Recognition	0.04+	0.04+	0.03	**0.04***	**0.04***
Satisfaction	**-0.06*** **	**-0.04***	**-0.05** **	**-0.04***	-0.03

In Table 9, we see the correlations between node-level network measures and survey outcomes for surveyed individuals. Each row indicates a survey measure. Each column indicates a node-level network measure. There are five node-level network measures: three from the raw email network (In-Degree, Out-Degree, and Betweenness Centrality), and two from the reciprocity network (Reciprocity In-Degree and Reciprocity Out-Degree). The cell value is the Kendall's Tau and its significance for a given possible correlation between the survey measure and node-level network measure.

Both the raw email network measures and the reciprocity network measures provided statistically significant outcomes in most cases. I was surprised that the raw email network node-level measures performed this well, with significant correlations in both In-Degree and Out-Degree for five measures — Need For Change, Merger Commitment, Organizational Justice, Supervisors are Fair, and Job Satisfaction. Betweenness reciprocity in the raw email network does not provide any advantage over the degree measures, which are much easier to gather and calculate.

Windowed reciprocity also provided statistically significant correlations to six survey measures — all of the measures the raw email network was able to correlate to, plus Recognition. In addition, for Merger Commitment, the Kendall's Tau concordance is higher for windowed reciprocity measures than for structural measures, meaning the relationship is more statistically — and practically — significant. People who are well embedded in the organization, with highly reciprocal relationships, are more likely to be committed to the merger. In addition, I think it is valuable that reciprocity, unlike raw email measures, has distinct incoming and outgoing measures, which are relevant in analyzing the survey measure of job satisfaction. Reciprocity in-degree — a measure of how often the individual respond to others — has a statistically significant negative correlation with job satisfaction, but reciprocity out-degree — a measure of how often others respond to the individual — does not. This reflects the difference between being "chained to the desk by email," which is typically unsatisfactory, and the satisfactory experience of having others quickly get back to you. In the raw email network, this distinction is lost.

These results tell us interesting things about the merger and how an individual's informal network relates to their feelings about the organization and merger. Individuals with more reciprocal ties tend to recognize the need for change, feel committed to the merger, feel that their supervisors are fair, and feel that they are recognized at work. At the same time, individuals with more reciprocal ties also feel less fairly rewarded by the organization, and, as previously discussed, tend to feel less job satisfaction. Taken together, these facts paint a complicated picture of MergedCo's employees: Those who are highly connected feel like the organization is unfair even as they recognize the need for change and do their best to support the merger, while those who are less connected feel their local leadership has failed them and the acquisition was a terrible idea, but they don't believe that rewards are distributed unfairly.

I also wanted to see if these correlations at the individual level would hold at the functional group level for Time-2 functional groups. There are only twenty (20) Time-2 functional groups, so any significant correlations would indicate a strong relationship at the group level. I used the mean of the values for a group's members to calculate the group's value for each survey outcome. When aggregated to the functional group level, job satisfaction was the only measure with a statistically significant correlation to the network measures. Table 10 showing the Tau estimate.

Table 10. Tau-Estimates and P-Value for Job Satisfaction at the functional group level

Group Mean of Measure	Tau-Estimate and P-Value For Satisfaction
In-Degree	**0.39 (0.03)**
Out-Degree	**0.39 (0.03)**
Betweenness Centrality	0.15 (0.42)
Reciprocity In	**0.41 (0.02)**
Reciprocity Out	**0.43 (0.02)**

Groups with many internal and external connections, and that return messages to others often, also tend to have higher job satisfaction. Given the small set of data points available, I am pleased with this result. It also suggests that job satisfaction, unlike commitment to the merger or recognition of the need for change, is a more universal quantity. As discussed earlier in the analysis, this merger affected different functional groups in wildly varying ways.

Key Entities Analysis

Network methodologies allow analysis at multiple levels of granularity. Throughout this work, I have focused on two distinct group levels: the legacy organization level and the functional group level. However, I can also analyze network outcomes at the individual level.

ORA (Altman et al., 2018) supports this type of analysis with key entity reports and can identify changes across time in various measures of social network influence. A key entity report uses a given meta-network and generates many different measures of node importance in that meta-network. An individual may be on many shortest paths (betweenness), which suggests access to information. An individual may have connections to key resources, giving them total or partial exclusivity. An individual may have many connections (degree) and use those to wield influence on the organization. The key entity report calculates these measures and many others and then integrates them into a holistic view of each individual’s importance based on rank across all measures.

These top-ranked actors are distinct from the typical individual at MergedCo. Table 11 shows the three measures we’ve discussed throughout this work, but the full key entities report includes 16 measures.

Table 11. Distributions of measures in the raw email network, comparing the typical value to those of the top 100, top 25, and top 10.

Measure	Median Value	Top 100 Value	Top 25 Value	Top 10 Value	Top Value
In-Degree	3	97	152	169	214
Out-Degree	1	134	245	346	2,699
Betweenness (Normalized)	0	0.003	0.007	0.011	0.154

I used the key entity report to compare Time-1 and Time-2 and identify changes in the top-ranked individuals based on the raw email network. I identified sixteen (16) individuals who were greatly affected by the changes occurring to MergedCo’s informal network. Such individuals, sensing a loss of influence, may well respond with strategies that are maladaptive for the organization as a whole. They may choose to emphasize the failures of the merger in individual conversations. They may leave, taking many other individuals with them, to work for a competing firm. They may choose not to reinforce messages of unity and solidarity that come from central leadership. Opportunities abound for mischief in the corporate sphere.

Interestingly, none of these sixteen individuals took the survey in either Time-1 or Time-2. This suggests that actors who perceive a change in their status or fear an upcoming change may choose to avoid notice of their beliefs.

I also compared the key entities identified by analysis of the raw email network to those identified in the windowed reciprocity network. None of the top 10 actors are consistent between the two networks. Only 3 of the top 10 in the raw email network took the survey in any year, and all of them were in Information Technology. Two were from LuxuryCo and one was from StandardCo. Only one of the three in the raw network was a manager. Of the top 10 in the windowed reciprocity network, only 2 of the top 10 took the survey, both from StandardCo. One of the two was from the Manufacturing group, and the other was from the International group. Neither was a manager.

Limitations

This work has focused on email correspondence. Of course, email is not the only way individuals communicate with each within a large organization. Email often supports other forms of interpersonal communication, however. I believe email correspondence can meaningfully identify key entities within formal and informal groups inside organizations, but because this is a study of a single company, further work that employs windowed reciprocity in similar contexts is necessary.

Further research should focus on replicating these analyses and results across different forms of technology-mediated or enabled communications, such as instant messaging, group chats, team collaboration tools, and VOIP logs.

Much work remains to better understanding the impact of a horizontal merger on the organization. In addition to their effects on the informal network, I know that horizontal mergers also involve the clash and reconciliation of sometimes contradictory corporate cultures, which I focus on in Chapter 2. Further research must focus on identifying how horizontal mergers reconcile their contrasting cultures to perform work.

Conclusion

This work has focused on an organization undergoing a horizontal merger and assessed whether network analysis and windowed reciprocity could capture the dynamics of the merger across two time periods. Because the organization is a hybrid of a Weberian bureaucracy and a virtual organization, its formal and informal structures influence the efficacy of the merger. I examined survey outcomes by functional group — the primary formal unit of the organization — and by legacy organization. By analyzing these outcomes alongside individuals’ email interactions, I can reveal how the formal and informal organizational structures affect survey outcomes for MergedCo.

From the structural analysis, I see that the new company identity, MergedCo, has arisen as a self-contained unit. As a result, LuxuryCo’s employees have become more dependent on others within MergedCo, which employees perceive as a loss in status from acquirer to partner. Because functional groups are intended to be at least somewhat self-contained, I interpret the eight new functional groups in the organization — all of which have strong ties outside the group — as struggling to create a cohesive in-group even as they perform their work for MergedCo’s customers. With the windowed reciprocity analysis, the dominance of MergedCo by Time-2 is revealed even more starkly: MergedCo receives responses from other legacy organizations much more often than they do from MergedCo. At the functional group level, I see employees reaching across new functional groups to perform work but not being responsive to others within their own groups.

The key-entity analysis on the raw email network identified some cause for concern for MergedCo leadership. Sixteen important leaders in the informal network at Time-1 — all of whom avoided the survey in both Time-1 and Time-2 — lost influence by Time-2. Such leaders, even if they don't hold obvious levers of influence, can cause substantial damage to an organization if they feel they are treated unfairly or if they no longer have MergedCo's interests in mind. Alternatively, those leaders may have been isolated through happenstance and would succeed with more support. In either case, MergedCo should investigate and take appropriate action.

From the survey measures, I see that legacy organization has a strong influence on how people feel about the organization. MergedCo employees, hired into the new merged entity, see the merger as more necessary and are more committed to the merger than older employees. LuxuryCo employees, at Time-2, are still uncertain of the need for change — not surprising, given their fall from acquiring partner to merely another element of the larger MergedCo operation. StandardCo employees, from the acquired organization, are more committed to the merger by Time-2 than they were in Time-1.

Based on this analysis, the MergedCo merger is not an entirely happy one. In the typical narrative of a successful merger, the acquiring organization's employees would be more committed to the merger at Time-2 than they were at Time-1. Instead, many of LuxuryCo's employees seem to feel that the merger was a mistake and that they were better off pre-acquisition.

Finally, I see that network dynamics and windowed reciprocity both correlate strongly with most of the survey measures. These correlations, while weak, suggest that how the organization acts and how its people respond to each other can weave subtle but persistent effects on feelings of organization fairness, on job satisfaction, and on feelings of recognition.

References

Ahuja, M. K., & Carley, K. M. (1999). Network structure in virtual organizations. *Organization Science, 10*(6), 741-757.

Altman, N., Carley, K. M., & Reminga, J. (2018). *ORA User's Guide 2018* (CMU-ISR-18-103). Retrieved from Pittsburgh, PA: http://www.casos.cs.cmu.edu/publications/papers/CMU-ISR-18-103.pdf

Apache Software Foundation. (2016). Apache Tika. tika.apache.org: Apache Software Foundation.

Baker, D. P., Day, R., & Salas, E. (2006). Teamwork as an Essential Component of High-Reliability Organizations. *Health Services Research, 41*(4p2), 1576-1598. doi:10.1111/j.1475-6773.2006.00566.x

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity,and Social History. *Games and Economic Behavior, 10*, 122-142. Retrieved from https://www.sciencedirect.com/science/article/pii/S0899825685710275

Bergiel, B. J., Bergiel, E. B., & Balsmeier, P. W. (2008). Nature of virtual teams: a summary of their advantages and disadvantages. *Management Research News, 31*(2), 99-110.

Boyd, R., & Richerson, P. J. (1989). The Evolution of Indirect Reciprocity. *Social Networks, 11*(3), 213-236.

Burgoon, J. K., Dillman, L., & Stem, L. A. (1993). Adaptation in dyadic interaction: Defining and operationalizing patterns of reciprocity and compensation. *Communication Theory, 3*(4), 295-316. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-2885.1993.tb00076.x

https://s3.amazonaws.com/academia.edu.documents/42743686/Interpersonal_Adaptation_Dyadic_Interper20160216-6848-eb3xkj.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1550438962&Signature=cSTe2N0hiEKIaospSjgFntGWrlA%3D&response-content-disposition=inline%3B%20filename%3DInterpersonal_Adaptation.pdf

Burke, R. J., & Cooper, C. L. (Eds.). (2000). *The Organization in Crisis: Downsizing, Restructuring, and Privatization*. Oxford, UK: Blackwell Publishers, Ltd.

Cammann, C. (1983). Assessing the attitudes and perceptions of organizational members. *Assessing organizational change : A guide to methods, measures, and practices*, 71-138. Retrieved from https://ci.nii.ac.jp/naid/10027435723/en/

Carley, K. M. (1991). Designing organizational structures to cope with communication breakdowns: A simulation model. *Industrial Crisis Quarterly, 5*(1), 19-57.

Carley, K. M. (1992). Organizational learning and personnel turnover. *Organization Science, 3*(1), 20-46.

Carley, K. M. (2003). Dynamic Network Analysis. In R. Breiger, K. M. Carley, & P. Pattison (Eds.), *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, Committee on Human Factors, National Research Council* (pp. 133-145). Washington, DC: National Research Council.

Carley, K. M. (2017). ORA: A Toolkit for Dynamic Network Analysis and Visualization. In R. Alhajj & J. Rokne (Eds.), *Encyclopedia of Social Network Analysis and Mining* (pp. 1-10). New York, NY: Springer New York.

Cartwright, S. (2002). Why mergers fail and how to prevent it. *Business: The Ultimate Resource*.

Cartwright, S., & Cooper, C. L. (2014). *Mergers and acquisitions: The human factor*: Butterworth-Heinemann.

Contractor, N. S., & Seibold, D. R. (1993). Theoretical frameworks for the study of structuring processes in group decision support systems: Adaptive structuration theory and self-organizing systems theory. *Human Communication Research, 19*(4), 528-563.

Coyle-Shapiro, J. K., I. (2002). Exploring reciprocity through the lens of the psychological contract: Employee and Employer Perspectives. *European Journal of Work and Organizational Psychology, 11*(1), 69-77.

Criswell, J. H. (1946). Measurement of Reciprocation Under Multiple Criteria of Choice. *Sociometry, 9*, 126-127.

Dion, C., Allday, D., Lafforet, C., Derain, D., & Lahiri, G. (2007). Dangerous liaisons, mergers and acquisitions: The integration game. *Hay Group report, 10*, 2007.

Doreian, P. (2002). Event sequences as generators of social network evolution. *Social Networks, 24*(2), 93-119. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.523.2283&rep=rep1&type=pdf

Drucker, P. F. (1988). The coming of the new organization. *Harvard Business Review*, 3-12.

Fisher, C. (2013). The Average Employee Turnover in Large Organizations. *Chron.com*, 1. Retrieved from http://work.chron.com/average-employee-turnover-large-organizations-29521.html

Frantz, T. L. (2012). A social network view of post-merger integration. In C. L. Cooper & S. Finkelstein (Eds.), *Advances in mergers and acquisitions* (pp. 161-176): Emerald Group Publishing Limited.

Frantz, T. L., & Carley, K. M. (2009). Computationally modeling the effect of organizational complexity on post-merger integration. In C. L. Cooper & S. Finkelstein (Eds.), *Advances in Mergers and Acquisitions* (pp. 79-101): Emerald Group Publishing Limited.

Gouldner, A. W. (1960). The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review, 25*, 161-179. Retrieved from https://www.jstor.org/stable/pdf/2092623.pdf?casa_token=NRuzNzzvN40AAAAA:SfjieNDkhcm

X7jqMHA6u5eKXhl1S7uU027bjTrGm1HygECJy1ZUacB7XOBGdelciZ8DCJr2SzQWwoTatFb1qNHASWN9-e8QXvYJRKUhE6QZLGnZuYczd
Hallinan, M. T. (1978). The process of friendship formation. *Social Networks, 1*(2), 193-210. doi:10.1016/0378-8733(78)90019-9
Huber, G. P. (1990). A theory of the effects of advanced information technologies on organizational design, intelligence, and decision making. *Academy of Management Review, 15*(1), 47-71.
Huber, G. P. (1991). Organizational learning: The contributing processes and the literatures. *Organization Science, 2*(1), 88-115.
Janis, I. L. (1972). *Victims of groupthink: a psychological study of foreign-policy decisions and fiascoes*: Houghton-Mifflin.
Karagiannis, T., & Vojnovic, M. (2009). *Behavioral profiles for advanced email features.* Paper presented at the Proceedings of the 18th international conference on World wide web.
Katz, L. (1955). Measurement of the Tendency Towards Reciprocation of Choice. *Sociometry and the Science of Man, 18*, 659-665.
Kenis, P., & Knoke, D. (2002). How organizational field networks shape interorganizational tie-formation rates. *Academy of Management Review, 27*(2), 275-293. Retrieved from https://www.jstor.org/stable/pdf/4134355.pdf
Klein, H. J., Wesson, M. J., Hollenbeck, J. R., Wright, P. M., & DeShon, R. P. (2001). The assessment of goal commitment: A measurement model meta-analysis. *Organizational behavior human decision processes, 85*(1), 32-55.
Krackhardt, D., & Stern, R. N. (1988). Informal networks and organizational crises: An experimental simulation. *Social Psychology Quarterly*, 123-140.
March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*(1), 71-87.
Marks, M. L., & Mirvis, P. H. (2001). Making mergers and acquisitions work: Strategic and psychological preparation. *Academy of Management, 15*(2), 80-92. doi:2001.4614947
Maznevski, M. L., & Chudoba, K. M. (2000). Bridging space over time: Global virtual team dynamics and effectiveness. *Organization Science, 11*(5), 473-492.
Merton, R. K. (1957). The role-set: Problems in sociological theory. *The British Journal of Sociology, 8*(2), 106-120.
Merton, R. K., & Merton, R. C. (1968). *Social theory and social structure*: Simon and Schuster.
Newcomb, T. M. (1961). The acquaintance process as a prototype of human interaction. doi:10.1037/13156-015
Niehoff, B. P., & Moorman, R. H. (1993). Justice as a mediator of the relationship between methods of monitoring and organizational citizenship behavior. *Academy of Management Journal, 36*(3), 527-556.
Seeley, J. R. (1948). The net of reciprocal influence: A problem in treating sociometric data. *Canadian Journal of Psychology, 3*, 234-240.
Subbian, K., Srivastava, J., Pinar, A., Singhal, A., & Kolda, T. G. (2013). *On Reciprocity in Massively Multi-player Online Game Networks*. Retrieved from http://cds.cern.ch/record/1533004
Toma, A., Roghé, F., Noakes, B., Strack, R., Kilmann, J., & Dicke, R. (2012). Organization of the future–designed to win. Flipping the odds for successful reorganization. In: New York.
Tsvetovat, M., Reminga, J., & Carley, K. (2004). *DyNetML: Interchange Format for Rich Social Network Data*. Retrieved from Pittsburgh, Pa.: http://reports-archive.adm.cs.cmu.edu/anon/isri2004/CMU-ISRI-04-105.pdf
Tushman, M. L., & Nadler, D. (1978). Information processing as an integrating concept in organizational design. *Academy of Management Review, 3*, 613-624.
Tyler, J. R., & Tang, J. C. (2003). *When can I expect an email response? A study of rhythms in email usage.* Paper presented at the ECSCW 2003.

Van Dick, R., Grojean, M. W., Christ, O., & Wieseke, J. (2006). Identity and the extra mile: Relationships between organizational identification and organizational citizenship behaviour. *British Journal of Management, 17*(4), 283-301. doi:10.1111/j.1467-8551.2006.00520.x

Wang, C., Strathman, A., Lizardo, O., Hachen, D., Toroczkai, Z., & Chawla, N. V. (2011, 8 March 2019). *Weighted reciprocity in human communication networks*. (preprint arXiv:1108.2822). arXiv.

Weber, M. (1922). *Economy and society: An outline of interpretive sociology* (Vol. 1): Univ of California Press.

Wuchty, S., & Uzzi, B. (2011). Human communication dynamics in digital footsteps: A study of the agreement between self-reported ties and email networks. *6*(11), 8. Retrieved from https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0026972&type=printable

Chapter 2: An empirical examination of a horizontal merger via textual analysis

Abstract

Horizontal mergers require two collections of individuals to interoperate and eventually integrate in order to capture the benefits of the merger for corporate stakeholders. To accomplish this goal, the new organization's leadership team must define a new organizational identity and associated culture, and then communicate that new identity to all members of the organization through a diffusion process. Meanwhile, while this new identity spreads, individuals must make decisions about which information to value and from whom to accept new information. This paper is an empirical over-time study of the diffusion of a new corporate identity. I use network and textual analysis to identify individual employees who changed in culture and those who did not, and I explain what that implies about these individuals' commitment to the merged organization. This work also introduces a new method for quantitatively evaluating differences between two textual corpora and explains how this method complements, but does not replicate, existing methods. I find that network centrality and legacy organization affiliation are strongly associated with language change, an important measure of cultural adoption. Language change is also associated with the sentiment of the tokens an individual uses; token sentiment, in turn, is correlated with recognizing that the organization needs to change, as reported on employee surveys. I also found that more negative language sentiment in Time-1 was strongly associated with not being present in the data by Time-2. This work concludes with a discussion of the implications and areas to be explored further through computational simulation.

Introduction

In this work, I am attempting to apply an empirical lens to the organizational changes and adaptations that occur over the course of a horizontal merger. I conceptualize the merged company's new unified identity as an innovation that diffuses across the merging organizations through the interactions of individuals. As such, I am communicating an essentially Constructuralist (Carley, 1991) perspective — one in which the words people use are often linked to the people with whom they are working day-to-day.

However, in large multinational organizations, individuals are sometimes reassigned *en masse* to new duties, responsibilities, and coworkers. These reassignment decisions are made by relatively few individuals in leadership positions, most of whom likely have very little personal insight into the essential coworking networks they are disrupting and reshaping. This fact of organizational life is doubly true when an organization is going through a horizontal merger — the merging of two organizations that occupy the same market niche and previously competed for market share, requiring similar resources to do their work and selling the resulting product to a similar set of customers (Koehler, 2009).

In successful horizontal mergers, the unified organization is able to establish a coherent organizational identity (Adolph et al., 2001), which leaders must communicate in the symbols, or tokens, they use with employees and the public. To be fiscally prudent and realize the economies of scale projected by the merger's advocates, most such mergers also require the streamlining and deduplication of organizational competencies, such as human resources, information technology, and marketing (Pablo, 1994). MergedCo's press releases for the merger identified tens of millions of dollars of cost-savings from deduplication of competencies. Such streamlining and deduplication inevitably involve significant reorganization and restructuring of the social context within which the organization's members work.

Reorganizing and restructuring are not without risks. Multiple industry analyses over the years have estimated that over half of horizontal mergers fail (Adolph et al., 2001; Porter, 1987). Rather than two organizations merging into one unified and market-dominating whole, they become a fragmented organization constantly at internal war with itself, unable to move swiftly and coherently to grasp new market opportunities. While these failures are not new, the continuing rise and dominance of virtual organizations (Drucker, 1988), predicted more than thirty years ago, has exacerbated the trend. More and more work is now dominated by knowledge specialists who must coordinate with specialists in different fields to perform the work of the organization. This coordination is threatened by the structural reorganization that is necessary to achieve the merger's goals.

Failed mergers are often attributed to human or cultural problems. Organization managers are often advised in general terms to improve their communication of new values, to make more explicit plans, and to follow laid-out processes more strictly. As good as these ideas might be, they have been the standard guidance for more than 20 years, and yet there is no evidence that merger outcomes have been improving.

In this work, I take a Computational Organizational Theory (COT) perspective (Morgan & Carley, 2015). Armed with this perspective, I conceptualize a horizontal merger as a diffusion task (Rogers, 2010). Measuring cultural change is usually only possible at the group level, using survey data (Bechky, 2011), but I present a method for measuring it at the individual level by taking advantage of data from the organization at work. I also examine factors that correlate highly with this cultural change.

Furthermore, this paper takes aim at that general advice that management improve merger communication. I suggest instead that it is possible to take advantage of the data from the organization at work to identify areas where patterns of communication are changing, where patterns of language are changing, and where general sentiment is growing more negative, so that management can potentially intervene. I then show that the aforementioned measurable changes have a significant correlation with employee survey answers that reveal sentiments relevant to the success of the merger — namely, agreement on the need for change and commitment to the merger.

Motivating Theory

In this section, I will review literature about organizations from multiple perspectives. First, I will discuss the development of a unified organizational culture as a diffusing innovation and introduce the individuals within an organization as boundedly rational actors who must make the best decisions they can moment to moment. I will also consider the organization as a whole as a boundedly rational actor and discuss how organizations learn and improve in that light. I will consider literature that suggests that the organization's culture relates strongly to its language, and I will consider research into organizational socialization, which helps me understand how individuals orient themselves within an organizational context. Finally, I will bring these research areas together into five testable assertions that relate to horizontal mergers.

The Horizontal Merger as a Diffusing Innovation

Individuals within an organization act as processors of information. This processing is limited by bounded rationality (Simon, 1991) — that is, each individual has limited time and attention for processing and retain information and for creating and maintaining social connections. These individuals constantly make decisions about whom to listen to, whom to learn from, which new information sounds correct, and which new information should be retained. An individual may make decisions of this type multiple times a day, and even more within a virtual organization (Drucker, 1988), where lines of authority are more ambiguous and everyone's expertise is required to produce value.

Effective communication and knowledge sharing within a virtual organization requires a common context and a common understanding of the organization's values and goals (Ahuja & Galvin, 2003). However, a horizontal merger disrupts this common context and common understanding until a new organizational identity can be distributed throughout the organization. The longer the organization's identity is in turmoil, therefore, the more likely the organization is to experience critical failures in communication that leave it unable to adapt to changing market conditions.

When considering a horizontal merger as an information diffusion process, it is important to identify the typical success path and common failure points along the diffusion curve. In a leadership-driven horizontal merger, an initial senior leadership group — the early adopters — will be responsible for identifying the new unified cultural norms, as well as the associated values and symbols. Then, after a period of review, the leaders push the new identity out to the organization. Members of the organization are expected to align with the new organizational identity, but this realignment often taking significant time and effort. Typically, those closest to senior leadership will be the next to adopt the new norms, and those further removed from senior leadership will take longer. Eventually, the new identity will be adopted by most of the organization, and it will become the new de facto identity of the organization.

Explicitly framing the new shared identity as something that must diffuse across the organization helps me understand ways in which the unification process after a merger can go wrong: Perhaps the new identity is not adopted by anyone outside senior leadership, or perhaps it is only attractive to a certain subset of employees. This framing also helps me understand why senior management is often surprised by merger failure. Firstly, senior management's understanding of the progress of the merger will be informed by the people around them, who are the most likely to adopt the new identity rapidly. Secondly, the organization's ability to complete tasks and communicate successfully is increasingly at risk as the organization reaches the midpoint of the adoption curve, when each individual is likely to be interacting with others who have not yet adopted the new identity. Thirdly, this stall in successful operations — caused by many suboptimal interactions — sows doubt among employees about the merger effort, making individuals who have not yet adopted the change even more likely to resist it. This leaves the organization at its most precarious point for an extended period of time.

The Organization as an Information Processor

I approach this research from the perspective that organizations are information-processing systems (Cyert & March, 1963). Furthermore, these organizations must fit themselves to the market environment in which they work (Burton, Obel, & DeSanctis, 2011). In addition — and most importantly to this work — organizations and the individuals within them learn over time from the work they do (Argote, Beckman, & Epple, 1990; Darr, Argote, & Epple, 1995). From this perspective, it follows that the key to a successful horizontal merger is the ability to retain, transfer, and unify organizational knowledge throughout the merger effort. As such, a successful horizontal merger requires a successful information diffusion process (Rogers, 2010).

Organizational knowledge is embedded in an organization's members, tasks, and tools, as well as in the relationships between those members, tasks and tools (Argote & Hora, 2017; Argote, McEvily, & Reagans, 2003). The streamlining and deduplication of organizational competencies (Pablo, 1994) is therefore destructive to the knowledge of the organization. The uncertainty of a merger often reduces employee attachment to the organization, and the turnover rate increases (Sung et al., 2017). Previous experimental work indicates that even relatively mild turnover in an existing workgroup can hinder task performance (Argote, Aven, & Kush, 2018).

Even without turnover, the restructuring imposed by the merger often creates new coworking relationships. This requires individuals to learn how best to work with their new coworkers, reducing their productivity in the meanwhile (Liang, Moreland, & Argote, 1995). Restructuring also potentially destroys the complementarity of individuals' knowledge (Denrell, Fang, & Levinthal, 2004), making them less effective at team tasks as they adjust to the new group structure. Until social cohesion forms between these individuals, it will be difficult for them to share knowledge effectively (Reagans & McEvily, 2003). Even when newly-joined individuals learn to work together, identifying with different legacy organizations may prevent them from doing so in the most effective ways (Kane, Argote, & Levine, 2005).

Even if turnover *and* structural changes could be avoided, the horizontal merger itself changes the context of an organization's work, rendering previous knowledge less useful and harder to transfer within the organization (Argote & Ingram, 2000). Finally, the effort of doing the work itself begins to create new community-of-practice boundaries within organizations, limiting knowledge transfer within the evolving organization (Brown & Duguid, 2001).

Managing organizational knowledge has become even more difficult as more work has become virtual (Drucker, 1988), or driven by teams with specific expertise. Merging virtual organizations requires coordinating not only individual employees, but also how they are connected as teams. Awareness and management of these team structures is crucial to retaining their effectiveness (Ahuja & Carley, 1999; Carley, 2002). Previous work has shown that turnover affects teams more than it affects traditional rigid hierarchies (Carley, 1992).

A central concern for a virtual organization is how to manage the organization's informal network. An individual's position within the network affects their ability to benefit from their individual knowledge (Cook & Emerson, 1978), to broker between groups (Burt, 2000) and to manage teams (Rodan & Galunic, 2004). Informal network structures also affect the performance of workgroups: Centrally-located workgroups tend to perform better than those that are less connected (Tsai, 2001). Workgroups composed of people with a diversity of structural positions benefit more from knowledge-sharing (Cummings, 2004) and are more effective (Oh, Chung, & Labianca, 2004; Oh, Labianca, & Chung, 2006) than less diverse workgroups.

In addition to the intraunit effects described above, the structure of interunit ties also affects the ability of these organizational units to learn and perform. Weak interunit ties, such as those formed when representatives from various units meet regularly, help teams find new information. Strong interunit ties, such as long-term shared commitments to specific projects, make it easier to transfer complex knowledge between units (Hansen, 1999). Strong ties also tend to reduce conflict within the organization (R. E. Nelson, 1989). Strong interunit ties are not a panacea, however, because maintaining them often comes at the expense of organizational productivity (Hansen, 2002). Levin and Cross (2004) argue that tie strength is less important for successful knowledge transfer than is the presence of trust relationships — i.e., even weak ties support knowledge transfer as long as they are trustworthy. Moreover, organizational units with a high ratio of external to internal ties are likely to do better in crises than those with mostly internal ties (Krackhardt & Stern, 1988).

In this section, I briefly summarized the challenges a horizontal merger would impose on an organization. I considered the organization as an information-processing entity, where both the organization and its members learn over time, and the horizontal merger as a diffusion process. I discussed the additional challenges imposed by the growing trend toward virtual organizations and outlined how the informal structure of the organization affects the performance of both intra-organization units and individuals within the organization.

The Organization as its Culture

The structure and processes of an organization are incredibly important during a merger, but so is its organizational culture. As Davies, Nutley, and Mannion (2000) outline, there are two broad schools of thought within the organizational culture academic community. One considers culture as something an organization *is*, and the other considers it as something that an organization *has*. The latter has traditionally been more conceptually tractable for those trying to change organizations, but I believe, as Davies does, that the organization's culture emerges from the interaction of its conceptual parts (its members, tasks, and tools), which makes it extremely difficult to control — but also perhaps amenable to influence. Notably, we lack a multilevel and empirical organizational theory of culture that would help us understand *how* people do their work within the organization (Bechky, 2011).

In the context of a merger, it matters whether individuals need to change a lot or a little to align with the new diffusing organizational identity. Merging firms whose employees perceive their cultures as being more different tend to have lower ultimate shareholder gains (Chatterjee, Lubatkin, Schweiger, & Weber, 1992), while firms that merge with very similar firms are more successful (Finkelstein & Haleblian, 2002).

Traditionally, organizational culture has been assessed with survey instruments (e.g., Cooke and Rousseau, 1988), which can be administered over time to characterize the changing culture of the organizational unit of interest. However, surveys cannot connect organizational culture change to change within individual members. Survey practice often discounts individual responses as being noisy, instead trusting group means and trends to reflect the real change. In large organizations, these analyses are only actionable at the group level, and employees expect their individual survey responses to be held in confidence (and will often find reasons to leave organizations that break this trust). These limitations on actionable insights are a key reason why organizational management practice focuses merely on broad suggestions and guidance, as opposed to specific, concrete evidence-based mitigations.

To address this difficulty, I turn to Pettigrew (1979). In one of the earliest articles on organizational culture, Pettigrew reminds me that an organization's culture is reflected in more than what can be assessed with a survey. Pettigrew (p. 574) writes, "The offspring of culture are symbol, language, ideology, belief, ritual, and myth." Using this insight, I can generate an assessment of how an individual's language has changed over time using techniques and methods that were not computationally tractable until this century. Where an individual's language has changed, it is likely that their organizational cultural context has changed too. This definition of organizational culture is sympathetic to the definition in organizational socialization research, which identified an organizational culture as including, in part, the specialized language and ideology that bring meaning to a member's everyday experience (Van Maanen & Schein, 1979).

As Schweiger and Goulet (2005) were able to demonstrate with their approach, I hope that work in this vein makes it possible to understand where cultural differences are emerging over time in individuals — and to manage these differences more successfully.

Organizational Culture and Organizational Identity

As discussed in the previous section, I define organizational culture as something that an organization *is*, and I posit that an organization's culture can be delineated by its choices of symbols, language, ideologies, beliefs, rituals, and myths. An organization's *identity*, by contrast, consists of the "stylized narratives about the 'soul' or essence of the organization" (Ashforth and Mael, 1996, p. 21). Thus, the organization's identity, which must be capable of change over time, is driven by elements of its culture, including the "myths" the organization chooses to propagate about itself, its past, its ambitions, and about the world as it understands it.

Individuals express membership in a group through the stories they tell and the ways they make decisions. Goffman (1959) expressed this idea through his concept of frontstage and backstage performances. Similarly, Simon (1976) noted that identification with a group occurs when people make choices with that group in mind. Linguistically, individuals express their own identities by choosing words and stories that are distinct from what would be expected of a group member, and they express their group membership by adhering to the group’s expectations in this regard (Foote, 1951).

Note that individuals may express identification with a group despite a complete lack of knowledge about that group (Mael and Ashforth, 1995). As such, individuals who are attempting to show allegiance to a given organization and its culture will adopt symbols, languages, and myths perceived to be affiliated with that organization — often, but not always, correctly.

In this work, I use data from annual employee surveys to determine which legacy organization each individual identifies with. Because I am interested in how group cultures changed over time, I chose to consider the earliest available identification as enduring over time, even if that individual identified with a different legacy organization at a later time. There are alternative treatments that could be fruitful to explore in future work, which I will discuss later.

An Overview of Organizational Socialization

Organizational socialization is the process by which an individual acclimates to the context of the organization in which they work and the position they hold within that organization (Comer, 1991; Feldman, 1981). While the most visible context change of any individual occurs when they join the organization as a new hire, this process recurs repeatedly as the person’s professional life unfolds (Feldman, 1976; Schein, 1988).

Much of the literature on organizational socialization has focused on new hires. Many research studies have investigated the information acquisition tactics of individual joiners, and some have looked at how the organization can make this process more or less successful. Peers and coworkers are valuable sources of information for individuals going through socialization (Comer, 1991; Louis, Posner, & Powell, 1983; Morrison, 1993a, 1993b; Ostroff & Kozlowski, 1992; Van Maanen, 1978). Individuals going through socialization seek information from coworkers, and they attempt to learn different types of information in different ways. For example, individuals openly request task-related knowledge, but they tend to learn organizational norms merely through observation (Comer, 1991; Morrison, 1993b). More recently, enterprise social network sites have enabled better socialization, but newcomers tend to find these sites more useful than more established members of the organization do (Thom-Santelli, Millen, & Gergle, 2011).

Another line of research has highlighted the importance of organizational socialization by focusing on the outcomes of successful and unsuccessful socialization. Successful socialization increases job satisfaction (Feldman, 1976, 1981; Louis et al., 1983), commitment to the organization (Louis et al., 1983), internal work motivation (Feldman, 1981), feelings of empowerment at work (Feldman, 1976), and intention to remain at the organization (Feldman, 1981; Louis et al., 1983).

One of the early definitions of organizational socialization is that socialization is the interaction between a stable social system and the new members who enter it (Schein, 1988). This definition invites the question: what if the social system is not stable? Relatively few empirical research efforts have focused on socialization amidst the shifting context organizations experience as they undertake major organizational change efforts, such as an acquisition or merger. Aguilera, Dencker, and Yalabik (2006) developed a theoretical framework for maximizing the effectiveness of integration after an acquisition (rather than a horizontal merger), arguing that socializing newly acquired staff into the acquiring organization's identity was critical for the merger's financial success. Klindzic, Braje, and Galetic (2015) laid out the available research on organizational culture and identity during mergers and called for more empirical research into how organizational identities change during post-merger integration.

An Empirical Analysis of a Horizontal Merger through Five Testable Assertions

In the previous sections, I discussed the framing of the horizontal merger as a diffusion process among cognitively bounded individuals, and I identified why senior leadership may often be surprised by merger failures. I summarized relevant literature regarding the organization's role as an information processor, the ways an organization is embodied in its culture, and how socialization processes — led predominantly by coworker interactions — help individuals obtain the cultural and functional knowledge they need to be a member of the organization. I also identified two research lacunae: 1) the lack of work on measuring cultural change at the individual level, largely due to its difficulty (Bechky, 2011), and 2) the rarity of studying organizational socialization in a changing context such as a merger (Klindzic et al., 2015).

Drawing on Constructuralism (Carley, 1990) and symbolic interactionist theory, which posits that the self reflects society (Stryker, 2008), I argue that changes in what people know and believe are in large part due to the people they interact with, otherwise known as their alters (Carley, Martin, & Hirshman, 2009). Changes in knowledge and beliefs occur gradually over time, but as Goffman (1961) writes, changes in presentation and language choice may happen much more quickly as an organization changes and those changes become real to the individuals within the organization. This suggests that people may begin adopting new symbols before they know what those symbols mean for them, and before they have fully adopted the innovation of the horizontal merger.

Research on information diffusion has often privileged the central actors within any given context (Cook & Emerson, 1978; Tsai, 2001), because centrality provides more opportunities for new and valuable information to flow to the actor. These studies often consider the entire time period during which the innovation diffused. Practically speaking, however, especially in an organizational change context, central actors may already have processed and absorbed the new information — in this case, the new organizational culture — while the information is still spreading to actors that are more distant. This leads me to my first testable assertion:

> *A-1: Network centrality in Time-1 will have a negative association with language change between Time-1 and Time-2.*

Where Assertion-1 focuses on network centrality, assertion-2 focuses on the actor's network position. In this work, an actor's network position is a characterization of their local environment, and it indicates whether the actor interacts entirely within a given network cluster, operates between clusters, or has relatively few ties at all. Prior research indicates that structural positions, regardless of position in the larger network, are likely to inform how a given actor encounters the diffusion of new information. Individuals whose positions change as part of a merger are likely to experience more changes in their language as they adjust their language to present the best possible face to the people they work with.

> *A-2: A change in social network position is likely to lead to more change in language than no change in position.*

Adapting to an organization is often described as both stressful and cognitively demanding (D. L. Nelson, 1987). As an actor learns to replace old symbols with new ones, one might expect that that these cognitive demands would cause actors to become more negative in attitude and in the symbols they use day to day (the latter shift being more important here, because I can measure it). On the other hand, the very stress of continuously adapting to a new organizational context may cause symbol change to occur rapidly if individuals are invested in their relationships with others, à la the Goffman perspective. This investment — and the performance of it — may lead these individuals to use more positive tokens instead. In this perspective, individuals who are no longer invested in the organization and their relationships with others will both change less and be less careful of how they present themselves to others in the organization. I think Goffman's perspective is more likely to be the reality in a large multinational organization and therefore raise Assertion 3:

> *A-3: Change in language between Time-1 and Time-2 is positively associated with language sentiment in Time-2.*

If individuals who are more negative in Time-1 tend to be less often present in Time-2, then I have additional evidence that the latter interpretation above — that individuals who become less interested in managing their personas become more negative — has some utility.

> *A-4: Individuals who left by Time-2 are more likely to use negative language in Time-1.*

Finally, if language sentiment has to do with one's investment in the organization and the merger, I would hope that language sentiment in Time-2 would have some correlation with responses to survey data from Time-2. However, this relationship will be muddled by the demand characteristics of surveys (Orne, 1962) among other factors.

> *A-5: Language sentiment in Time-2 will have a weak positive correlation with self-reported commitment to the merger and belief in the organization's need for change in Time-2.*

Data

The data for this effort comes from a large multinational organization, which I call LuxuryCo. LuxuryCo performed a horizontal merger with another large multinational organization, StandardCo, to form MergedCo. MergedCo provided access to email records and annual survey records. I use the survey records to inform the functional and locational grouping of individual employees, as well as their perceptions of the need for change and their commitment to the merger.

Email Data

The email data includes both metadata (time, sender, and recipients) and text data (subject and body). For textual analysis, I considered an email's subject and body as one document. As a note for further research, MergedCo underwent significant effort to provide these email records, which were stored offsite. The research team manually aggregated these records. Based on the concentration of emails shown in Figure 7, I aggregated these emails into three periods: Time-0, Time-1, and Time-2.

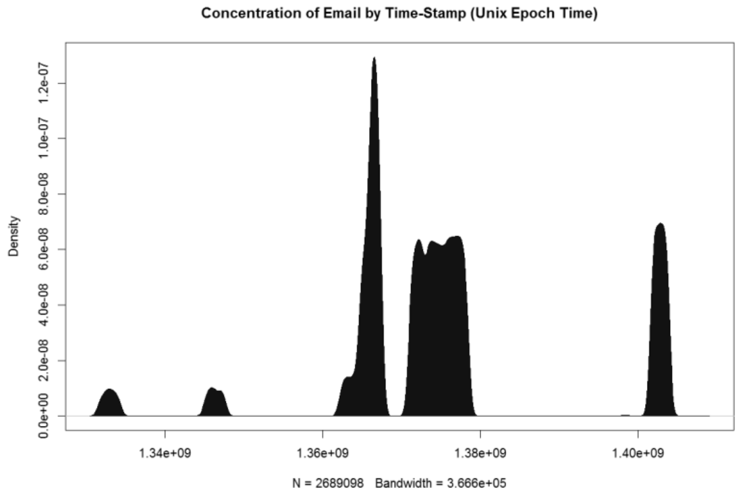


Figure 7. The distribution of email timestamps

Not all emails are created equal, so I wanted to limit the data set to emails where personal interaction was likely to be significant. I established three criteria:

- That the email be primarily in English, as identified by Apache Tika (Charron, Mattmann, & Zitting, 2014)
- That the number of recipients be fewer than seven, none of which are MergedCo's mailing lists
- That the sender of the email is an individual, not a mailing list

In Figure 8, I focus on Time-1 and Time-2 to examine the distribution of emails across time. In Late 2013, I see thirteen (13) spikes, which correspond to thirteen work weeks. There are two deeper-than-typical dips: one between the 4^{th} and 5^{th} work weeks, and one between the 12^{th} and 13^{th} work weeks. These correspond to United States holidays. In 2014, there are four work weeks. The regular cadence of work weeks shown in Figure 8 reassures me that, despite the sampling procedure, I have not introduced problematic bias in the data. I use Time-1 and Time-2 data for the remainder of this work.

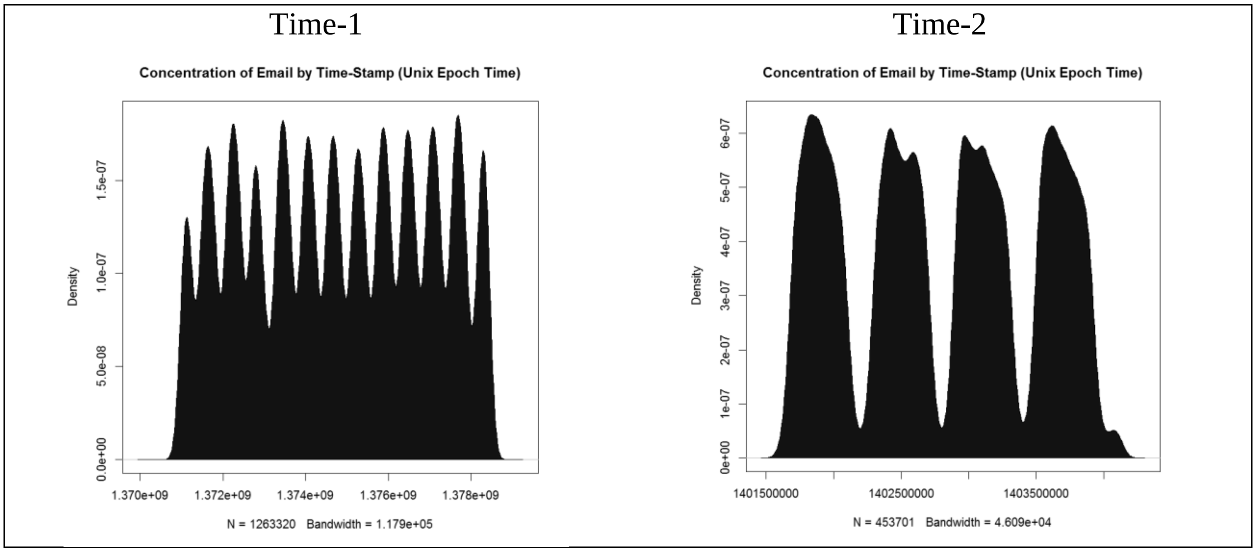


Figure 8. Zoomed in distributions of email time-stamps

MergedCo has significant assets in countries where English is not a primary language. Because of data-protection measures, which included systematically anonymizing all nouns, I removed emails that were not primarily in English. I used Apache Tika (Charron et al., 2014) to identify languages.[7]

Survey Data

MergedCo also performed annual employee surveys. I use survey responses from years that map well to the periods I have identified as Time-1 and Time-2. While response rates are very high for questions about functional role and location, they drop off significantly for questions about what it is like to work at the organization. A summary of response rates and email data availability is in Table 12.

Table 12. Response rates to questions from Time-1 and Time-2. Location was not in the survey at Time-1.

Time-Period	# Individual Respondents	Functional Role Response Rate	Locational Response Rate	Substantive Question Response Rate	Time-2 Email Data Available
Time-1	1,660	100%	--	34.2%	93.4%
Time-2	1,693	100%	99.8%	33.5%	96.9%

I use survey results to assign respondents to functional and locational groups in order to inform the analysis.

Merging survey results with Time-2 email data, I end up with 2,122 participants for whom both email and survey data is available. These participants are visualized in Figure 9 based on their first reported legacy affiliation. Figure 9 shows that by Time-2, MergedCo has come to dominate

[7] I examined the language distribution of emails in the data set, and the vast majority are in English. I removed all emails not identified as written in English.

the organization. LuxuryCo actors appear on the right and StandardCo actors appear on the left margins, but both are well integrated into the central mass of MergedCo.

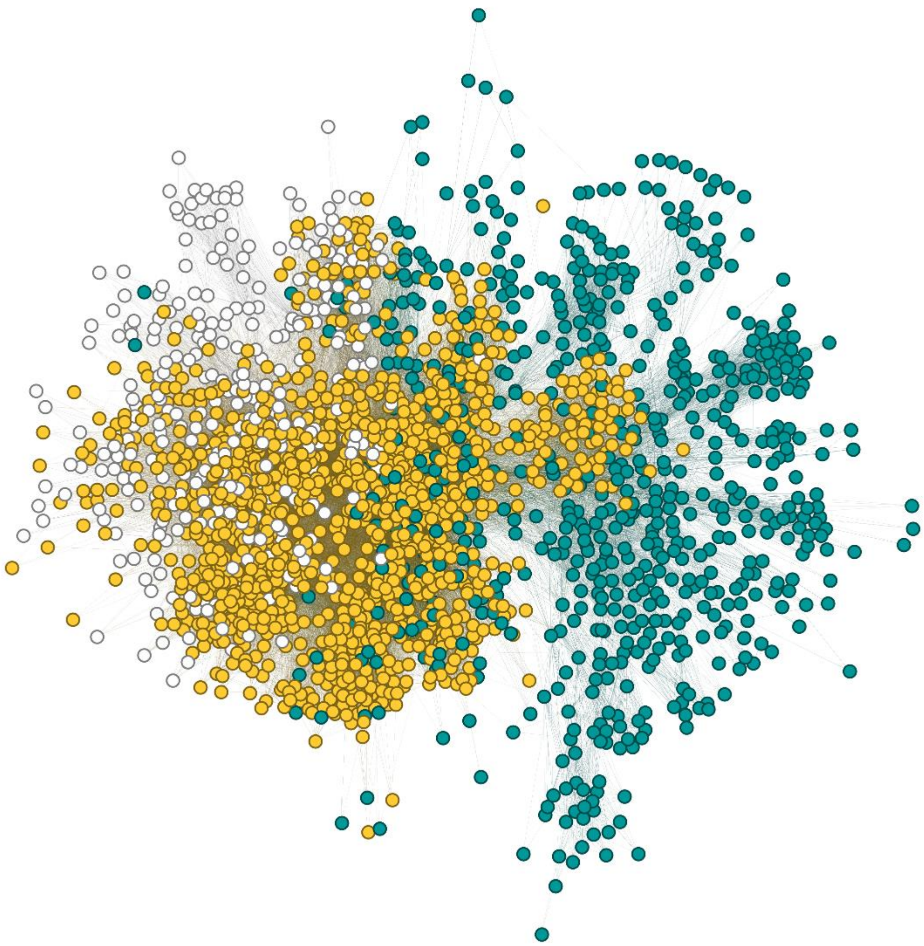


Figure 9. MergedCo in Time-2, colored by legacy organization (LuxuryCo = Green, MergedCo = Gold, StandardCo = White)

Data Preparation and Treatment

In this section, I describe how I processed emails to support the statistical modeling and results. First, I describe my Corpora Comparison method in some detail and provide the equations required to reproduce it. I also compare my Corpora Comparison technique to prior art, both categorically and quantitatively. I then discuss how I used the Corpora Comparison method to develop a language similarity network. Finally, I describe how I used the high-reciprocity social network ties described in Chapter 1 to generate network positions, and how I used language ties to identify network clusters and positions with the Louvain method (De Meo, Ferrara, Fiumara, & Provetti, 2011).

Corpora Comparison

Each token, t, is a unique symbol among all tokens, T. Group membership and document authorship are used together to generate two distinct corpora — A and G — which I then compare. For example, by using legacy affiliation and document authorship, I can create a unified list of documents that were created by members of LuxuryCo, StandardCo, the new

MergedCo, or all members of the organization. Using this approach, I can then compare documents from LuxuryCo to all documents from the unified organization, or I can compare documents from StandardCo to the documents of people who first indicated they were members of MergedCo.

Equation 3 describes the normalized odds ratio of a given term, *t*, being affiliated with either *A* or *G*. The normalized odds ratio is the occurrence of term *t* in Corpus *A*, normalized by the occurrence of all terms in Corpus *A*, divided by the number of times term *t* appears in Corpus *G*, divided by the occurrence of all terms in Corpus *G*. Scores range from -0.5 to 0.5. Scores near 0 indicate that a particular token does not offer a meaningful distinction between the two corpora. Scores below 0 indicate that the token is more common in Corpus *A* than in Corpus *G*; scores above 0 indicate that the token is more common in Corpus *G* than in Corpus *A*.

Equation 3. The normalized odds ratio of term t *between Corpora* A *and* G

$$odds(t, AG) = \left(1 - \left(\frac{1}{\left(\frac{|t_A|}{|T_A|} \Big/ \frac{|t_G|}{|T_G|}\right)}\right)\right) - .5$$

As described in Equation 4, I use a cutoff value, *c*, to avoid assigning importance to marginal cases. If the absolute value of *odds(t)* in Equation 3 is less than *c*, then *odds(t)* is reassigned to 0. The cutoff ratio *c* is typically 0.1 — meaning I ignore tokens that are tilted less than 60/40 in favor of one corpus or the other. I do this to avoid assigning importance to terms that may appear frequently but show minimal evidence of being distinctive.

Equation 4. The flattened odds ratio cancels out marginal cases.

$$fOdds(t, AG, c) = abs\big(odds(t, AG)\big) > \ c,\ odds(t, AG)$$
$$else,\ 0$$

After identifying whether a term is more strongly associated with *A* or with *G*, I calculate, as shown in Equation 5, a contextualized frequency term that compares the prominence of the term in a particular corpus against the prominence of that term in an appropriate prior, *P*. In my case, the prior is all emails from periods other than Time-1 or Time-2.

Equation 5. I also calculate a frequency term, which compares the term against a prior. Thus, terms have to both be unusually associated with A *or* G*, and be significantly novel compared to the prior* P.

$$\begin{aligned} freq(t, AGP, c) = {} & fOdds(t, AG, c) > 0,\ max(odds(t, AP,\ 0) \\ & fOdds(t, AG, c) < 0,\ max(odds(t, GP),\ 0) \end{aligned}$$

Equation 6 combines the frequency term calculated in Equation 5 with the flattened odds of Equation 4 to score the token's overall impact. Flattened terms are canceled out no matter how prominent they are compared to the prior. As a result, high scoring terms:

- Distinguish *A* and *G*
- Are more prominent in either *A* or *G* than in *P*

Equation 6. A term's score is a function of both the flattened odds ratio (EQ3) and the frequency score compared to the prior (EQ4)

$$S(t, AGP, c) = fOdds(t, AG, c) * freq(t, AGP, c)$$

Equation 7 summarizes the scores for all terms to produce a relative comparison of Corpora *A* and *G* against Prior *P* with cutoff value *c* for a given set of tokens *T*. Unless noted otherwise, all tokens found in *A* and *G* are used to inform *T*.

Equation 7. I can score a group in comparison to another group by defining A *and* G *(the comparison groups), the prior* P, *the cutoff value* c, *and the term list,* T.

$$Score(T, AGP, c) = \sum_{t} abs(s(t, AGP, c))$$

In this section, I have identified how I score tokens and corpora. I use this technique to a) to score language similarity between individuals, b) to score language difference between two points in time in a single individual, and c) to identify specific tokens that should be considered distinctive to an individual compared to the organization norm.

Comparing Morgan Corpora Comparison, TF-IDF, and Jaccard Similarity

Because the Morgan Corpora Comparison (MCC) is a novel method for calculating the difference between two textual corpora, I compare it to two existing techniques: 1) using Term Frequency, Inverse Document Frequency, or TF-IDF (Sparck Jones, 1972), as a Euclidean distance metric, and 2) using Jaccard Distance (Jaccard, 1912) on normalized term frequency.

To help differentiate my approach from TF-IDF and Jaccard, Table 13, below, compares the features of the three methods.

Table 13. Feature comparison between TF-IDF, Jaccard Distance, and Corpora Comparison

Method Characteristic	TF-IDF	Jaccard Distance	Morgan Corpora Comparison (MCC)
Scores "stop words" low	Yes	No	Yes
Identifies valuable words	Yes ("most substantive terms")	No	Yes ("most distinctive terms")
Uses term document count to contextualize term counts	Yes	No	No
Requires a second corpus to be used	No	Yes	Yes
Identifies both positive and negative relationship of tokens to a given corpus	No	No	Yes
Robust to high variance in document size	No	Yes	Yes
Robust to high variance in individual corpus size	No	If normalized	Yes

Both TF-IDF and MCC penalize "stop words" (Wilbur & Sirotkin, 1992) — words without inherent meaning — while Jaccard, which does not generate token-level values, does not. TF-IDF and MCC each identify valuable words, but their definitions of value are different. TF-IDF scores terms highly if they are used often in relatively few documents, while MCC gives high scores to the terms that are most distinctive between two corpora. Both Jaccard and MCC rely on comparison between corpora, while TF-IDF can be used on a single corpus. Both Jaccard and MCC are robust to high variance in document size, while TF-IDF is not. Jaccard may or may not be robust to corpus size, depending on whether term frequencies are normalized or not. Because I am interested in comparing corpora, and I want to know what terms drive differences between corpora, the MCC approach is most useful to me in this work.

However, I can also quantitatively assess how these methods compare. To do so, I will first introduce the calculations of TF-IDF and Jaccard Similarity. I then show the correlation between these two techniques and Corpora Comparison by using each technique to track changes in an actor's language between Time-1 and Time-2.

TF-IDF as a Euclidean distance metric is calculated by taking the square root of the square of different TF-IDF values, g_t and b_t, for a given token t, where g_t is a specific TF-IDF value for token t in the documents of the Corpus G, and b_t is a specific TF-IDF value for token t in the documents of Corpus B. Corpus B and Corpus G may or may not have overlap in documents. T represents all tokens across both G and B.

Equation 8. TF-IDF as a Euclidean distance metric

$$d_{GB} = \sqrt{\sum_t (g_t - b_t)^2}, g \in G, b \in B, t \in T$$

To calculate Jaccard Similarity for a set of all real numbers, sum the minimum normalized term frequency and divide it by the sum of maximum normalized term frequency for a given token t. Here, g_t is the term frequency of token t in Corpus G, and b_t is the normalized term frequency of token t in Corpus B.

Equation 9. Calculating Jaccard Similarity

$$J_{GB} = \frac{\sum_t min(g_t, b_t)}{\sum_t \max(g_t, b_t)}, g \in G, b \in B, t \in T$$

I calculate Jaccard Distance by subtracting the calculated Jaccard Similarity from one.

Figure 10 shows pairwise scatterplots comparing the three distance metrics. On the diagonal, I have the histogram of each distance metric. Each point on a scatterplot represents the calculated change in language, using two different methods, between a single person's Time-1 corpus and their Time-2 corpus. To make comparison and interpretation easier, all methods have been Z-scored so that their means are at 0, and plus or minus 1 indicates 1 standard deviation above or 1 standard deviation below the mean. While all three methods are strongly correlated to each other, the correlation is not perfect.

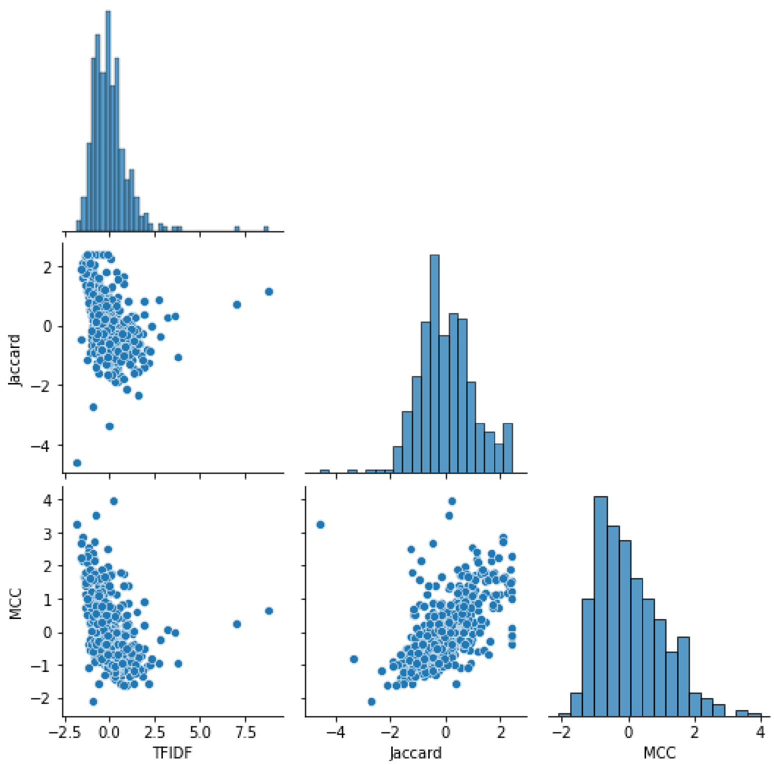


Figure 10. Pairwise scatter plots of the three distance metrics

I can use the Spearman Ranked Correlation to describe the ranked correlation between these different distance metrics. Ranked correlations compare the ordering of values within a dataset. Table 14 tells me that there are correlations between all three methods, but the strongest correlation is between Jaccard and MCC. The MCC's incorporation of the term's weight within its corpus provides me more information than the Jaccard comparison would alone, making it more valuable for my work.

Table 14. The Spearman Ranked Correlation scores of all of three comparison methods. All of these methods are correlated, but the highest correlation is between Jaccard and MCC.

Method One	Method Two	Spearman Correlation
TFIDF	Jaccard	-0.452
TFIDF	MCC	-0.537
Jaccard	MCC	0.638

Differences between LuxuryCo and StandardCo revealed by Morgan Corpora Comparison

I can use the Morgan Corpora Comparison method not only to compare changes in entities over time, as I did by comparing individuals at Time-1 to themselves at Time-2, but also to compare two or more entities to each other at the same point in time. In this section, I apply the MCC technique to draw out some tokens that distinguish LuxuryCo from StandardCo. In addition to examining key tokens, I use existing thesauri from NetMapper (Malloy & Carley, 2021) to consider eight topic groups which differ between LuxuryCo and StandardCo. Table 16 shows the most distinctive terms between LuxuryCo and StandardCo at Time-1.

Table 15. Top terms for LuxuryCo and StandardCo in Time-1, after accounting for Prior

Top Terms	
LuxuryCo	StandardCo
Best	Returning
Free	Pricing
Call	Information
Dear	Credit
Us	Status

I applied NetMapper (Malloy & Carley, 2021), which supports the conversion of tokens in over 40 languages to general topic groups. This lets me consider how the most distinctive tokens relate to a variety of topic groups that are relevant to organizational performance, horizontal mergers, and cultural differences. Table 16 has the list of topic groups and distinctive terms.

Table 16. NetMapper topic groups and key terms from LuxuryCo and StandardCo members

Topic Group	LuxuryCo	StandardCo	Topic Group	LuxuryCo	StandardCo
Management	Team	Reports	Financial	Sales	Credit
	Advised	Status		Customer	Pricing
	Submit	Assigned		Business	Charge
	Representative	Scheduled		Stock	Sale
	Inspection	Employee		Deal	Expenses
Topic Group	LuxuryCo	StandardCo	Topic Group	LuxuryCo	StandardCo
Trust	Kind	Assign	Encouragement	Happy	
	Rhapsody	Charge		Fantastic	
	Recommendation			Cheers	
				Excellent	
				Supreme	
Topic Group	LuxuryCo	StandardCo	Topic Group	LuxuryCo	StandardCo
Exclusivity	Difference	Specifically	Explanations	Training	Reminder
	Private			Option	
	Specifications			Alternative	
	Exception			Improve	
	Selected			Preparation	
Topic Group	LuxuryCo	StandardCo	Topic Group	LuxuryCo	StandardCo
Positive	Best	Reports	Negative	Problems	Busy
	Call	Credit		Unable	Reminder
	Sales	Production		Emergency	Label
	Free	Between		Difference	Pending
	Open	Care		Lean	Both

In Table 16, we have a set of eight topic groups from NetMapper, which I chose based on the larger context of horizontal mergers in large multinational organizations. These eight topic groups are: 1) management, 2) financial, 3) trust, 4) encouragement, 5) exclusivity, 6) explanations, 7) positive words, and 8) negative words. Each topic group represents a concept — validated by previous research teams for cross-cultural compatibility — and words that are associated with that concept. I identify the words associated with each topic group that are found within my corpora. I list a term under "LuxuryCo" if MCC identified the word as being distinctive to LuxuryCo, and I list it under "StandardCo" if MCC identified it as being distinctive to StandardCo. For each legacy organization, I list terms in order of their frequency, with higher-frequency words appearing first. All listed terms were used at least 30 times in the text.

The topic group "Management" covers the concept of assigning, delegating, and providing feedback on work to one or more individuals. LuxuryCo's top words in this category were *team*, *advised*, *submit*, *representative*, and *inspection*. In LuxuryCo's culture, work is assigned to teams, and representatives report on the status of that work and submit it to higher authority. StandardCo's top words, meanwhile, were *reports*, *status*, *assigned*, *scheduled*, and *employee*. In StandardCo's culture, management assigns work to individual employees. These employees report their status routinely and on a regular schedule.

The topic group "Financial" covers the concept of fiscal transactions. LuxuryCo's top words in this category were *sales*, *customer*, *business*, *stock*, and *deal*. In LuxuryCo's culture, the primary financial concern was ensuring that enough stock (inventory) was available to meet sales demand. StandardCo's top words in this category were *credit*, *pricing*, *charge*, *sale*, and *expenses*. In StandardCo's culture, financial concerns are focused on the tabulation of the account books — making sure the organization stays financially secure.

The topic group "Trust" covers the concept of extending support to others without certainty that they will support you in kind. Trust is essential to successful work relationships, particularly in virtual organizations where very few other people have the knowledge to second-guess your expertise. LuxuryCo's top words related to trust were *kind*, *rhapsody*, and *recommendation*. StandardCo's top words were *assign* and *charge*. Neither organization uses many trust words, but LuxuryCo's words imply a culture where individuals are able to provide connections to others. StandardCo's words, by contrast, are both about individual trust.

The topic group "Encouragement" covers the concept of supporting others as they do their work. LuxuryCo uses many encouragement words, its top five being *happy*, *fantastic*, *cheers*, *excellent*, and *supreme*. LuxuryCo is a place where individuals encourage each other with their words. StandardCo does not use any encouragement words.

The topic group "Exclusivity" covers the concept of creating distinction or difference between things. LuxuryCo uses many exclusivity words, its top five being *difference*, *private*, *specifications*, *exception*, and *selected*. Part of LuxuryCo's brand is how it distinguishes itself from anything else in the world, so it is not surprising this language also appears when LuxuryCo employees talk to each other. StandardCo, meanwhile, uses the word *specifically*. StandardCo's brand focuses on producing items very efficiently, reliably, and at an attractive profit margin.

The topic group “Explanations” covers the concept of providing instructions or guidance. LuxuryCo uses many explanation words, its top five being *training*, *option*, *alternative*, *improve*, and *preparation*. StandardCo uses the word *reminder,* a negative sentiment word focused on making sure people remember to do their tasks. In LuxuryCo’s culture, individuals were encouraged to know more about the “why” of their work.

The topic group “Positive” covers words that have a positive sentiment in many cultures. Both legacy organizations use many positive words. LuxuryCo’s top words are *best*, *call*, *sales*, *free,* and *open*. LuxuryCo’s culture focuses on the opportunity to be the best. StandardCo’s top words are *reports*, *credit*, *production*, *between*, and *care*. StandardCo’s culture is much more focused on the day-to-day operations of the organization.

The topic group “Negative” covers words that have a negative sentiment in many cultures. Both groups use many negative words. LuxuryCo’s top words are *problems*, *unable*, *emergency*, *difference*, and *lean*. LuxuryCo’s negative words focus on exception management and an inability to meet needs. StandardCo’s top words are *busy*, *reminder*, *label*, *pending*, and *both*. StandardCo’s negative words focus on keeping people on task.

Taking these eight topic groups together, we can see two starkly different organizational cultures. LuxuryCo’s team-focused culture revolves around making sure they can meet demand for their exclusive product and giving each other encouragement as they go. StandardCo’s culture, in contrast, focuses on individuals and their responsibilities to the organization.

These insights were drawn from the raw text of emails, without further processing or cleaning of terms before applying NetMapper. Future work with Morgan Corpora Comparison would benefit from sharpening the analysis by applying more cleaning — and potentially using thesauri specific to an organizational context — before topic mapping takes place.

Using the Morgan Corpora Comparison Method to Create High-Similarity Language Ties

Because the MCC method is robust to wide variances in corpus size, it can be used to characterize differences in token use between groups of any size — including “groups” that consist of a single individual. I used the method on every pairwise combination of agents in each time period for which I had at least 10 documents for each agent. This created 5.6 million MCC scores in Time-1 and 2 million scores in Time-2.

Such dense, weighted networks can be difficult to work with, however, so I developed a simple link-reduction process that retains only the top *k* edges of each node that have the highest similarity scores. Ties in similarity score were broken in favor of links that best retained local structure.[8] While the source graph was symmetric by definition, this top-*k* transformation adds directionality to the resulting network, because an edge may be retained by Agent A but not by Agent B. In the resulting language similarity network, an edge should be interpreted to mean “most similar written language.”

[8] This link-reduction utility script is available on GitHub and can be used with any network in a standard link-list csv format. I do not believe this capability exists off-the-shelf in any commonly used network software.

After I apply the link-reduction process, the Time-1 and Time-2 language similarity networks become computationally tractable and generate modular clusters.

Aggregating and Processing Networks to Identify Network Clusters and Network Positions
In this work, I have two source networks of interest:

- An *agent x agent reciprocity network*, calculated using the method defined in Chapter 1. I know this network has useful properties for predicting individual perceptions of an organization. The reciprocity network is weighted and directed.
- An *agent x agent language similarity network*, with tie strengths identified using the MCC method described in this chapter. The language-similarity network is weighted and directed.

When Agent A has a tie with Agent B in the reciprocity network, that means that A frequently responds to B within 24 hours when B communicates with A. When Agent A has a tie with Agent B in the language-similarity network, that mean's that B's written language is among the most similar to A's of all other nodes.[9]

I then applied the Louvain community detection algorithm to the reciprocity and language-similarity networks and explored multiple resolution values to identify an optimal cluster solution. To do this, I iterated from 0.5 to 1.8 in increments of 0.05. A good solution is one where the number of communities that result is stable (i.e., not in a canyon or ridge), and the convex hulls of the modularity scores converge. Generally, optimal solutions I found were near the standard resolution value of 1.0.

After identifying Louvain communities in each network, I characterized the intra-community and inter-community behavior of each node to identify its network position. To do this, I used:

- Total Degree: The number of edges this node has in this network
- Weighted Total Degree: The sum of the weight of all edges for this node in this network
- Total Degree – InGroup: The number of edges this node has with alters in the same Louvain community for this network
- Weighted Total Degree – InGroup: The sum of the weight of all edges for this node with alters in the same Louvain community for this network
- Total Degree – Bridge: The number of edges this node has with alters in a different Louvain community for this network
- Weighted Total Degree – Bridge: The sum of the weight of all edges for this node with alters in a different Louvain community for this network

After calculating these measures, I used Ward's Minimum Variance to generate a set of positions for each network. Ward's Minimum Variance is a hierarchical clustering approach that attempts to minimize variance within reported clusters. The method proceeds by iteratively merging clusters, always opting to merge the two clusters that will generate the smallest marginal increase

[9] I explored the idea that perhaps these networks should be combined, but I found that the network positions in the combined network did not offer any additional clarity that the text or structure networks alone did not.

in variance. After identifying these clusters, I used the distribution of each measure within the clusters to generate descriptive and meaningful labels. These are shown in Figure 11.

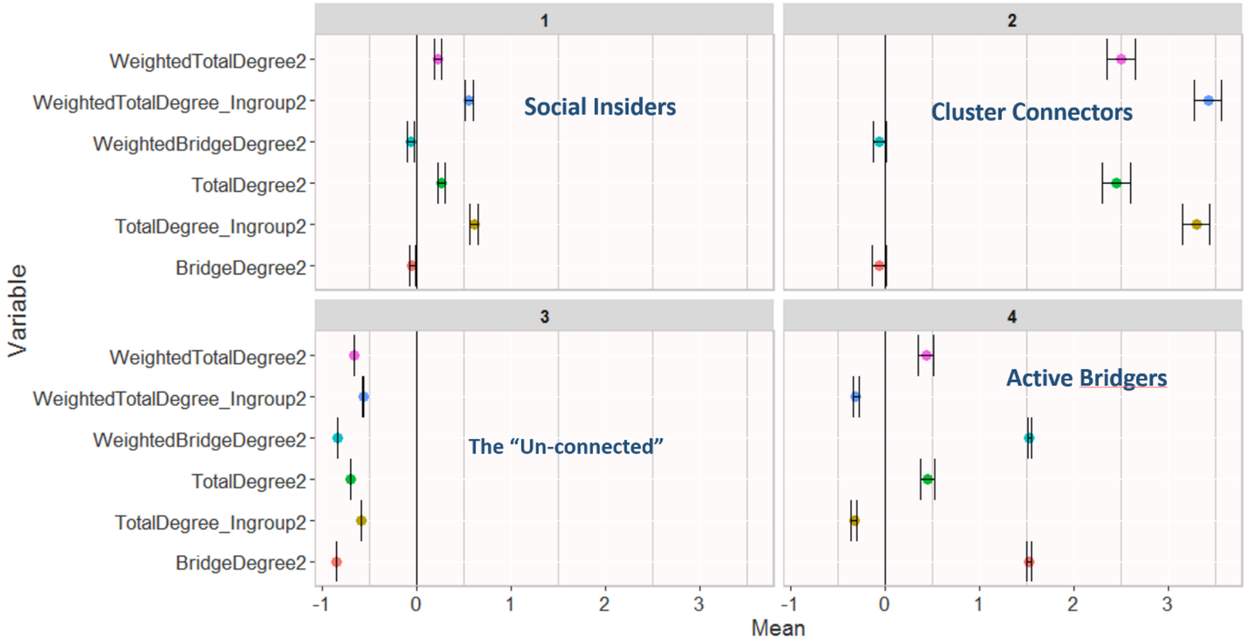


Figure 11. Time-1 and Time-2 structural positions

The four positions I identified show distinct distributions of the six contributory variables, which focus on in-group and out-group links. All scores are normalized so that 0 is the mean, and the *x*-axis values represent standard deviations. Nodes in Position 1, the Social Insiders, have high in-group degree and slightly below-average bridge (or out-group) degree. Cluster Connectors, in Position 2, have substantially higher in-group degree and average bridge degree. The Unconnected, in Position 3, have well below-average in-group degree and bridge degree. Active Bridgers, in Position 4, have much higher bridge ties and below-average in-group ties. I include labels on Figure 11 to help others understand what these statistical distributions mean in an organizational context.

Those with less tight connections within the organization often have more insight into what is happening externally. I reviewed the top most distinctive of tokens of the Unconnected. Their most distinctive tokens are focused on pricing, marketing, and consultants. Table 17 has the first five ranked tokens of each position.

Table 17. Top distinctive tokens of each position

Rank	Position 1	Position 2	Position 3	Position 4
1	Supports	Warranty	Catalogs	Free
2	Sincerely	Advisor	Processed	Return
3	Whatever	Cook	Price	Reader
4	Fantastic	Product	Tracking	Downloaded
5	Holds	Read	Pricing	install

After identifying node positions in each network, I can also consider how each node changes in position between the two time periods. Each combination of positions identifies a unique path an individual node may experience during this merger, such as moving from a peripheral network position to a central one. While the work with these paths is relatively early, I consider them a promising way to describe how an individual experiences a merger, with each network acting as a proxy for an important aspect of professional life.

Data Summary

In the previous sections, I discussed how I took the unstructured text data and used it to generate quantitative values representing language difference. I combined that information with email metadata, which I used to generate networks, and with survey data, to make it possible to evaluate my five assertions.

Data supporting testing Assertions 1, 2, 3, and 5

To be included in the data for testing Assertions 1 and 2, which evaluated the effects of network centrality and change in network position on language change between Time-1 and Time-2, individuals needed to:

1) Have at least 10 documents in both Time-1 and Time-2
2) Have taken the survey at least once
3) Have at least one reciprocal relationship in Time-1
4) Have used one or more tokens with positive or negative sentiment

Note that for Assertion 5, I am correlating against Merger Commitment and Need for Change survey responses, so the *n* of those variables strongly influences the analysis. Both Merger Commitment and Need for Change are averages drawn from several questions with 7-item Likert scales. Higher scores mean stronger commitment to the merger and more recognition of the need to change, respectively. Individuals included in the data were willing to give their legacy affiliation, but many chose not to respond to questions of how committed they were to the merger and whether MergedCo needed to change. Table 18 summarizes the available data.

Table 18. Summary of the data for testing Assertions 1, 2, 3, and 5.

Quantitative Variables	N	Mean	Median	StDev	Skew	Min	Max
Total Degree Centrality	552	15.56	14	10.607	1.090	1	59
Closeness Centrality	552	0.000355	0.000355	0.000013	0.046	0.000318	0.000392
Difference Score	552	-0.051	-0.213	0.950	0.731	-1.715	3.978
Total Token Sentiment	552	0.002	0.02	0.03	2.389	-0.006	0.025
Merger Commitment	137	5.265	5.250	1.011	-0.345	2.500	7.000
Need for Change	344	4.091	4.000	1.510	0.250	1.000	7.000
Categorical Variables	**N**	**# of Classes**		**# of Each Class (Top 5)**			
Legacy Affiliation	552	4		DNT-Survey: 123 LuxuryCo: 197 MergedCo: 174 StandardCo: 58			
Structural Positions, Time-1	552	4		1: 248 2: 37 3: 47 4: 220			
Unconnected in Time-1	552	2		True: 47 False: 505			
Disconnected by Time-2	552	2		True: 74 False: 478			
Structural Role Changed Time-1 to Time-2	552	2		True: 173 False: 379			
Language Positions, Time-1	552	4		1: 316 2: 200 3: 34 4: 2			

Because these variables differ widely in implicit scale, I Z-scored the values of all quantitative variables here. I Z-scored the values by recentering the distribution on 0 and dividing the recentered values by the standard deviation. For example, the three values 1, 3, and 5, from a larger distribution with an average of 3 and a standard deviation of 2, would now be reported as -1, 0, and 1, respectively. I otherwise did not transform these variables.

Data supporting testing Assertion 4

To test Assertion 4, I started with the 6,000 individuals for whom I have metadata in Time-1. From this set, I extracted 1,587 individuals who used ten or more tokens with either positive or negative sentiment in Time-1, excluding *thanks*, *thank*, and *please*. For this analysis, I determined whether a term had sentiment based on its VADER (Hutto & Gilbert, 2014) compound-sentiment score. I then weighted that token's sentiment based on how often the person used that term. Note that an individual is marked as having used a token only if they used it more than expected compared to everyone in the organization. A person may use terms with positive, negative, or neutral sentiment. Total token sentiment may range from -1 to 1, but will generally be near 0. Table 19 summarizes the available data.

Table 19. Summary of data used to test Assertion 4

Quantitative Variables	N	Mean	Median	StDev	Skew	Min	Max
Number of Sentiment Tokens	1,587	38.76	32	25.41	1.41	10	188
Total Token Sentiment	1,587	0.0022	0.0016	0.0027	2.86	-0.0055	0.0322
Categorical Variables	**N**	**# of Classes**		**# of Each Class**			
In Time-2	1,587	2		0: 1,244 (Not in Time-2) 1: 343 (In Time-2)			

Method

In the final theory section, I identified five assertions I wished to test through statistical modeling. I also gave an overview of the data and data preparation techniques, discussing how I converted email data — which has both unstructured text and structured metadata — into representations of change over time, as well as into reciprocity and language similarity networks. I concluded by explaining how I created network clusters and then used Ward's Minimum Variance to find network positions within both the reciprocity and language similarity networks. With the data preparations complete, I am now ready to evaluate the assertions.

The first two assertions focus on change in language between Time-1 and Time-2:

> *A-1: Network centrality in Time-1 will have a negative association with language change between Time-1 and Time-2.*
>
> *A-2: A change in social network position is likely to lead to more change in language than no change in position.*

I grouped these two assertions together because they are focused on the same outcome variable: change in language between Time-1 and Time-2. In addition to these two assertions, there are other factors that could potentially affect language change. For example:

- Language change may also be a function of language position in Time-1. Each language position indicates a different cultural role within the organization. Some individuals may be connected entirely within a given language community, while others may have ties between multiple language communities.
- Language change may be a factor of legacy affiliation (identification with LuxuryCo, StandardCo, or MergedCo), suggesting that individual change effectively took place because of organizational membership rather than an individual's particular position.

I test Assertion 1 and 2, along with the two confounding factors, using a robust model selection procedure. Using Bayesian Information Criterion (BIC), I generate and then evaluate models to determine which of the independent variables best explain the variance I see in language change between Time-1 and Time-2. For comparing structural positions, the control case is Position 3, The Unconnected. I use the largest group — Position 1, Social Insiders — as the control group for the language positions. Because evaluating Assertion 2 generates many paths in the data, I used Van Waerden post-hoc tests with a Bon Ferroni correction to mitigate the possibility of type-1 error.

The next assertion focused on the predominance of negative sentiment tokens among actors in Time-2. Because I believe strongly in the Goffman narrative — that individuals invested in the organization will be more likely to adapt to changes as needed and self-edit their language to present a professional persona — I believe that change in language between Time-1 and Time-2 will be positively correlated with language sentiment in Time-2.

> *A-3: Change in language between Time-1 and Time-2 is positively associated with language sentiment in Time-2.*

Assertion 3 presents a specific idea of the relationship between language change and language sentiment, but evidence is available for alternative theories and additional factors. For example:

- The relationship between language change and negative sentiment could be reversed, if the use of negative sentiment tokens was primarily influenced by the additional cognitive burden and stress that active change places on an individual (D. L. Nelson, 1987).
- The effect of language change could be entirely overcome by how embedded in the organization the individual is at Time-1, as measured by the total number of connections the individual has in the Time-1 reciprocity network. Highly embedded people, who interact with more people and therefore have a broader stage on which to present themselves, may use positive sentiment terms routinely even if their language doesn't change over time.
- Sentiment in Time-2 could be entirely related to the legacy affiliation (LuxuryCo, StandardCo, or MergedCo), which would suggest that each legacy organization experienced the merger differently.

As I did earlier, I explore Assertion 3 through a model-building exercise where I generate and evaluate models based on BIC.

Assertion 4 is intended to provide further support to the idea that individuals who use more negative tokens are less invested in maintaining a professional persona within the organization.

> *A-4: Individuals who left by Time-2 are more likely to use negative language in Time-1.*

To test Assertion 4, I compare two populations of individuals, all of whom have sentiment data available in Time-1. The first population consists of individuals who appear in Time-1 but not in Time-2, and the second consists of people who appear in both Time-1 and Time-2. I expect that individuals who do not appear in Time-2 are more likely to be negative. However, it is possible that the data will not support this for any of several reasons. It could be that the organization's culture values "truth-telling" through negative language. It could be individuals not appearing in Time-2 represent a data sampling issue, rather than employees who left between Time-1 and Time-2. It could also be that the token sentiment scoring approach is not refined enough to support this ancillary analysis.

I will test this assertion by evaluating whether presence in Time-2 can explain variance in sentiment, and whether those who remain in Time-2 tend to have higher sentiment. Even with a validated statistically supported finding, I will use this result only to suggest support for the Goffman perspective on the organization.

Assertion 5 connects language sentiment in Time-2 to survey-related outcomes. I would expect that individuals who are invested in the organization, and therefore willing to self-edit their use of negative tokens, would also be more likely to report that they are committed to the merger and believe in the organization’s need to change.

> *A-5: Language sentiment in Time-2 will have a weak positive correlation with self-reported commitment to the merger and belief in the organization’s need for change in Time-2.*

Due to response bias and inherent demand characteristics (Orne, 1962) — that is, the context in which the survey is done and the fact that respondents likely presuppose what the correct answers should be — I expect that this relationship will be weak. Most individuals are unlikely to report honestly that they are not committed to the merger or do not believe the organization needs to change when the organization is in the midst of the merger effort. I will assess this assertion by comparing individual language sentiment in Time-2 to survey outcomes in Time-2, then reporting the Spearman correlation score and its associated p-value.

Results

In this section, I walk through each of the five assertions and evaluate the results of the statistical modeling.

Predictive Models of Language Change between Time-1 and Time-2

The first two assertions involve predicting change in language between Time-1 and Time-2. I explore each of these assertions in separate subsections.

Network Position and Language Change: Assessing Assertion 1

The creation of a new organizational culture during a merger begins with leadership, which occupies a central position within the organization. However, with empirical studies like this one, the timeframe of the captured data may not match precisely with the onset of the merger and the associated behavior change. In addition, leadership may have chosen to create a new organizational identity and culture that they themselves would find relatively easy to adopt. In other words, we may not see much change in these individuals because they designed an organizational culture that already fits their current worldview.

As such, even though many previous works have focused on the diffusion of new information to the central core — e.g., Cook and Emerson (1978); Tsai, (2001) — I see instead what could be described as that effect's shadow: a central core that changes very little, while the periphery rushes to catch up.

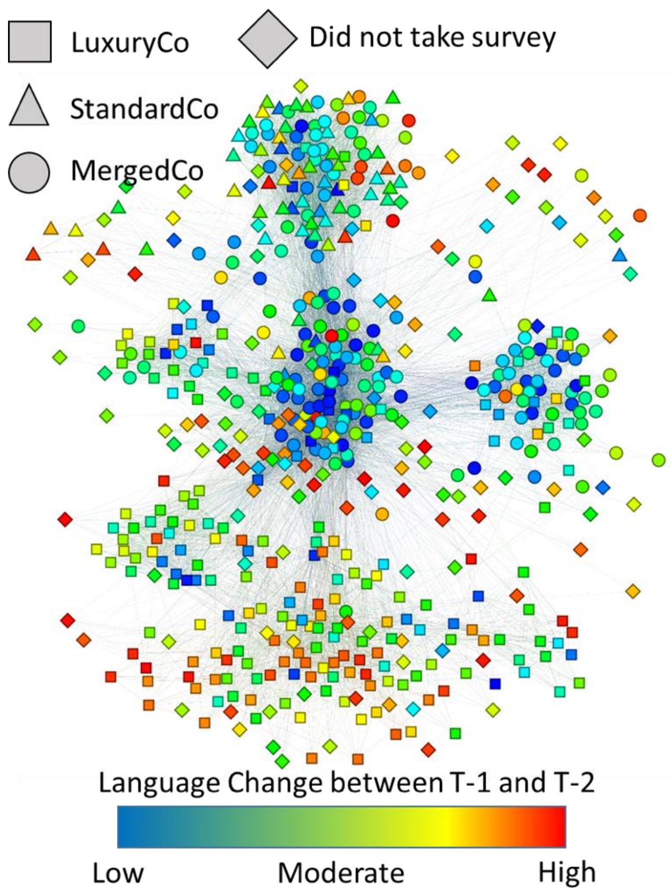


Figure 12. Language change between Time-1 and Time-2.

In Figure 12, I display the limited set of nodes for which I have data in both Time-1 and Time-2. These nodes are shaped according to their legacy affiliation and colored according to the amount of language change they experienced. I can see a knot of cool blues and greens in the center, with brighter greens on the edge of each cluster and bright yellows and reds on the periphery of the network. This visual analysis gives me additional confidence that nodes that are more central in the overall network are less likely to change their language. In network science, this centrality in the overall graph is called closeness centrality. Assertion 1 states the assertion below.

> *A-1: Closeness centrality in Time-1 will have a negative association with language change between Time-1 and Time-2.*

To evaluate this assertion statistically, I use a model selection procedure where I compare the factor of interest — closeness centrality — to other possible factors, including structural role, language role, and legacy organization affiliation. I build independent models of each of these factors, then evaluate them by comparing their BIC scores. The best models — those with the lowest scores — are those that better explain the variance in language change. I require an improvement (a decrease in score) by two or more points to justify switching to a model that adds complexity (e.g., adds an interaction factor). All continuous variables were Z-scored but otherwise unmodified. I evaluated seven models of language change using different network and position variables. These models were:

- Model 1: Closeness Centrality in the Structure Network
- Model 2: Structural Role Position in Time-1
- Model 3: Language Role Position in Time-1
- Model 4: Legacy Organization Affiliation
- Model 5: Model 2 + Model 1
- Model 6: Model 4 + Model 1
- Model 7: Model 6 + Interactions between Legacy and Closeness Centrality

Results are shown in Table 20, below. Each column represents one of the previously listed models. Each row represents a variable that may or may not be part of a model. Variables not in a model are blank in the column for that model. Categorical variables have the specific categorical value specified after a colon. The values in a cell represent the factors coefficient and standard error. Significance for a given row-cell is indicated by + for p-values of 0.1, * for p-values < .05, ** for p-values < 0.01, or *** for p-values < 0.001.

Table 20. Predicting Language Change via Network and Position Variables - Model Result Summary. Selected models in bold.

Predicting Language Change	**Models – Network and Position Variables**						
Variables	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Closeness Centrality	-0.405*** (0.036)				-0.416*** (0.036)	-0.341*** (0.036)	-0.445*** (0.078)
Structural Position T1:1		0.051 (0.150)			0.195 (0.134)		
Structural Position T1:2		0.007 (0.208)			0.081 (0.185)		
Structural Position T1:4		0.032 (0.152)			0.222 (0.136)		
Language Position T1:2			-0.077 (0.086)				
Language Position T1:3			0.015 (0.171)				
Language Position T1:4			1.268+ (0.672)				
Legacy: LuxuryCo				-0.254* (0.103)		-0.242** (0.092)	-0.225* (0.094)
Legacy: MergedCo				-0.914*** (0.105)		-0.667*** (0.098)	-0.666*** (0.099)
Legacy: StandardCo				-0.509*** (0.142)		-0.556*** (0.128)	-0.497*** (0.136)
Closeness * LuxuryCo							0.094 (0.099)
Closeness * MergedCo							0.137 (0.098)
Closeness * StandardCo							0.220 (0.152)
Intercept	-0.122** (0.036)	-0.125 (0.138)	-0.095 (0.053)	0.318** (0.081)	-0.306 (0.123)	0.236** (0.073)	0.217 (0.074)
BIC	**1430.07**	1543.51	1539.97	1463.10	1446.83	**1396.47**	1413.99

The best model is Model 6, which incorporates network centrality and legacy affiliation. I explored an interaction in Model 7, but the interaction factor did not improve on the independent factors of Model 6.

As documented in Table 20, the two variables of closeness centrality and legacy affiliation were each statistically significant predictors of language change when taken independently. When combined in Model 6 (closeness centrality and legacy affiliation), the two combine to create a superior model.

Plotting the interaction between closeness centrality and legacy affiliation, as I do in Figure 13, helps me see why the best model of these variables includes both factors.

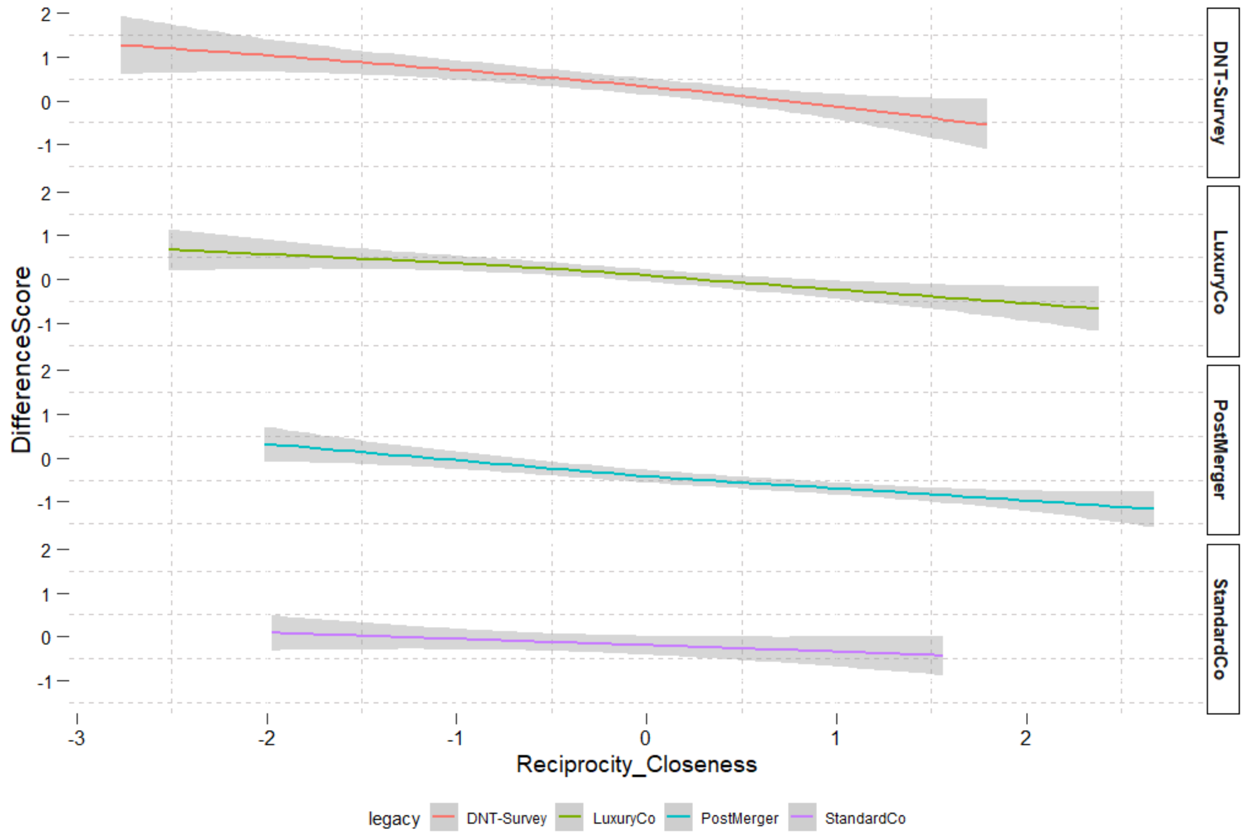


Figure 13. Closeness centrality vs. language change vs. legacy affiliation

In Figure 13, I can see that the relationship between closeness centrality and language change depends on legacy affiliation, with legacy affiliation changing the intercept.

At this point, I see strong evidence supporting Assertion 1, that closeness centrality will have a negative correlation with language change. In the next two subsections about modeling language change, I will assess Assertion 2, then conclude by bringing together both network position variables and structural position variables.

Changing Structural Positions: Assessing Assertion 2

In the previous section, I evaluated whether starting at a particular position in Time-1 affected language change. However, I also wanted to explore the idea of changing network positions more broadly. A node's network position indicates the role it fills within its local network — indicated in this case by the number and proportions of its out-cluster and in-cluster ties. Because network ties, and therefore network position, will inform the information a given node is exposed to, I thought there might be evidence that a change in structural position would influence change in language. This led me to Assertion 2:

> *A-2: A change in social network position is likely to lead to more change in language than no change in position.*

To visualize change in structural position over time in relation to legacy affiliation, I used a Sankey diagram, shown in Figure 14. The diagram includes data for all individuals in Time-1, regardless of whether I was able to observe them in Time-2. In the Sankey diagram, positions are ordered from bottom to top based on the overall number of individuals occupying that position in that time period. Each flow represents a unique path that one or more individuals took from the same starting position over the two time periods. Paths are colored according to legacy organization affiliation.

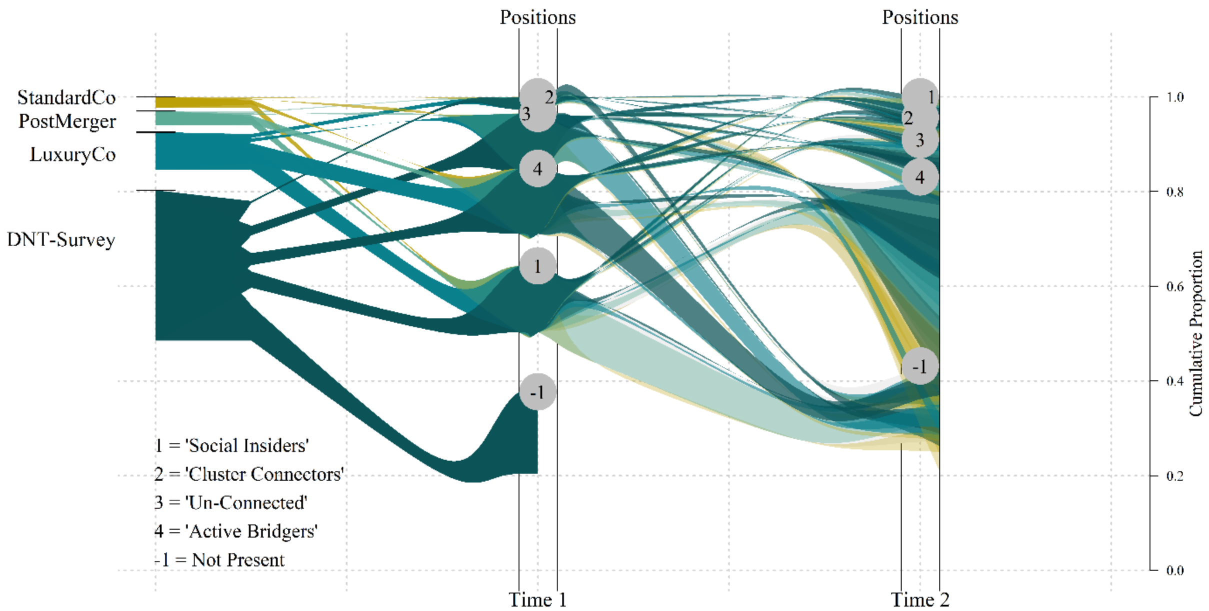


Figure 14. Sankey diagram of legacy affiliation, Time-1 positions, and Time-2 structural positions.

In Figure 14, I can see some interesting dynamics. Position 1, Social Insiders, which is quite common in Time-1, is quite rare in Time-2, when most people who began in Position 1 have left the organization. Instead, the new most common position in MergedCo by Time-2, after reorganization, is Position 4 — Active Bridgers, or people who are connecting both inside their cluster and outside of it. This suggests that the reciprocal network has experienced significant change, and value creation processes now need to occur across group clusters rather than being confined within them. This insight corroborates analysis performed in Chapter 1 that many functional organizations are now reaching across department bounds to get work done.

Position 3, The Unconnected, was quite interesting to me, because it includes individuals with much less attachment to the organization as a whole. I thought that either being or becoming disconnected might have a strong effect on language change, especially if I consider the Goffman perspective. Those who are already disconnected may have little interest in changing their language, while those who are becoming disconnected may want to change more to show that they fit in and are eligible for more integration into the organization. Note that by "disconnected," I do not mean having no ties at all, but rather having substantially fewer in-cluster and out-cluster ties than are typical for an individual within this organization.

To evaluate these ideas and explore Assertion 2, I again use a model selection procedure where I compare the factor of interest — role position change — to other possible factors, including being disconnected at Time-1 and being disconnected at Time-2. I build independent models of each of these factors and evaluate them by comparing their BIC scores, with lower BIC scores indicating superior models. The outcome variable, language change, was Z-scored but otherwise unmodified. In total, I evaluated four models:

- Model 1: Changed Role Position
- Model 2: Disconnected at Time-1
- Model 3: Disconnected at Time-2
- Model 4: Model 3 + Model 1

Results are shown in Table 21. Each column represents one of the previously listed models. Each row represents a variable that may or may not be part of the model; variables not in a model are blank in the column for that model. A cell represents the factors coefficient and standard error. Significance for a given row-cell is indicated by + for p-values of 0.1, * for p-values < .05, ** for p-values < 0.01, or *** for p-values < 0.001.

Table 21. Predicting language change using path variables - model result summary. Selected models in bold.

Predicting Language Change	**Models – Path Variables**			
Variables	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Changed Structure Position, Time-1 to Time-2	-0.097 (0.087)			0.009 (0.086)
Unconnected at Time-1		-0.013 (0.145)		
Unconnected at Time-2			0.800*** (0.114)	0.803*** (0.117)
Intercept	-0.084 (0.049)		-0.214 *** (0.042)	-0.217** (0.052)
BIC	1529.81	1531.28	**1487.76**	1494.10

Based on the path variables, the best model is Model 3, which suggests that being unconnected at Time-2 is the best of the path variables at predicting language change.

In Table 21, I can see that the most generic variable — change in role position between Time-1 and Time-2 — was not significantly correlated to language change. However, being unconnected at Time-1 or at Time-2 both indicated more change in language, the latter being by far the stronger of the two effects. I tried models where both factors were incorporated, but being unconnected at Time-2 remained the best model. I can see the relationship between being unconnected in Time-2 and language change in Figure 15.

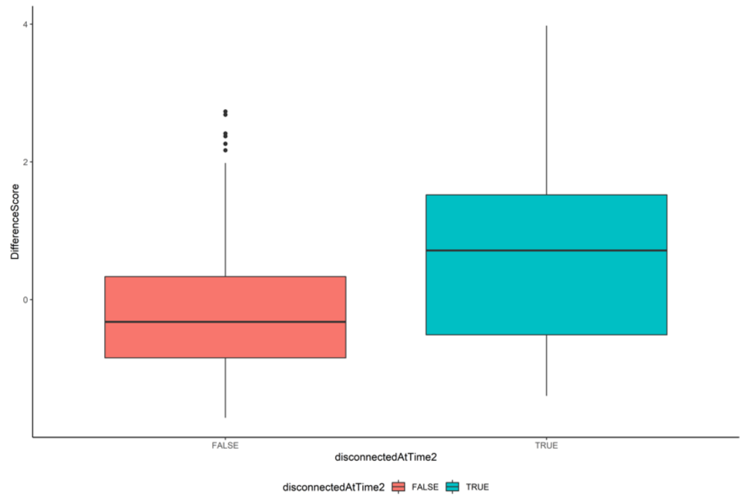


Figure 15. Box plot comparing language change by Position 3 (disconnected) at Time-2

Individuals who were unconnected at Time-2 changed their language more than those who were not unconnected at Time-2. This is consistent with the Goffman perspective.

In general, I do not find evidence for Assertion 2, and therefore it must remain unsupported. Relatively few individuals avoided a change in structural position between Time-1 and Time-2, so perhaps this treatment was too broad.[10] However, I do find that becoming disconnected by Time-2 did indicate substantial change in language.

After reviewing these models, I wanted to bring together the network, position, and path variables and identify the most explanatory model of language change. I do this in the final subsection.

Modeling Language Change with Network, Position, and Path Variables

After reviewing the network and position variables in isolation, and the path variables in isolation, I wanted to bring these models together into a unified set. The models I evaluated are:

- Model 1: Closeness Centrality + Legacy Affiliation
- Model 2: Disconnected at Time-1
- Model 3: Model 1 + Model 3

I show the modeling results in Table 22. Each column represents one of the previously listed models. Each row represents a variable that may or may not be part of the model; variables not in a model are blank in the column for that model. A cell represents the factors coefficient and standard error. Significance for a given row-cell is indicated by + for p-values of 0.1, * for p-values < .05, ** for p-values < 0.01, or *** for p-values < 0.001.

[10] I did evaluate, using Posthoc Tests, the various paths an individual could take and did find specific paths that indicated statistically significant language change compared to other paths, but the sample sizes were too small to be more than a discussion item for this work.

Table 22. Predicting language change using all variables - model result summary. Selected models in bold.

Predicting Language Change	**Models – All Variables**		
Variables	**Model 1**	**Model 2**	**Model 3**
Closeness Centrality	-0.341*** (0.036)		-0.341*** (0.035)
Legacy: LuxuryCo	-0.242** (0.092)		-0.173+ (0.089)
Legacy: MergedCo	-0.667*** (0.098)		-0.589*** (0.095)
Legacy: StandardCo	-0.556*** (0.128)		-0.503*** (0.124)
Disconnected at Time-2		0.800*** (0.114)	0.682*** (0.098)
Intercept	0.236** (0.073)	-0.214 *** (0.042)	0.100 (0.074)
BIC	**1396.47**	1487.76	**1357.64**

The best model is Model 3, which suggests that centrality, legacy affiliation, and being disconnected by Time-2 are together the best predictors of language change.

In Table 22, I combined the best model from the Assertion 1 modeling — which included closeness centrality and legacy affiliation — with the best model from the Assertion 2 modeling, which included being unconnected by Time-2. When I combine these factors, I get a substantially superior model for predicting language change. I visualize the interaction of these variables in Figure 16.

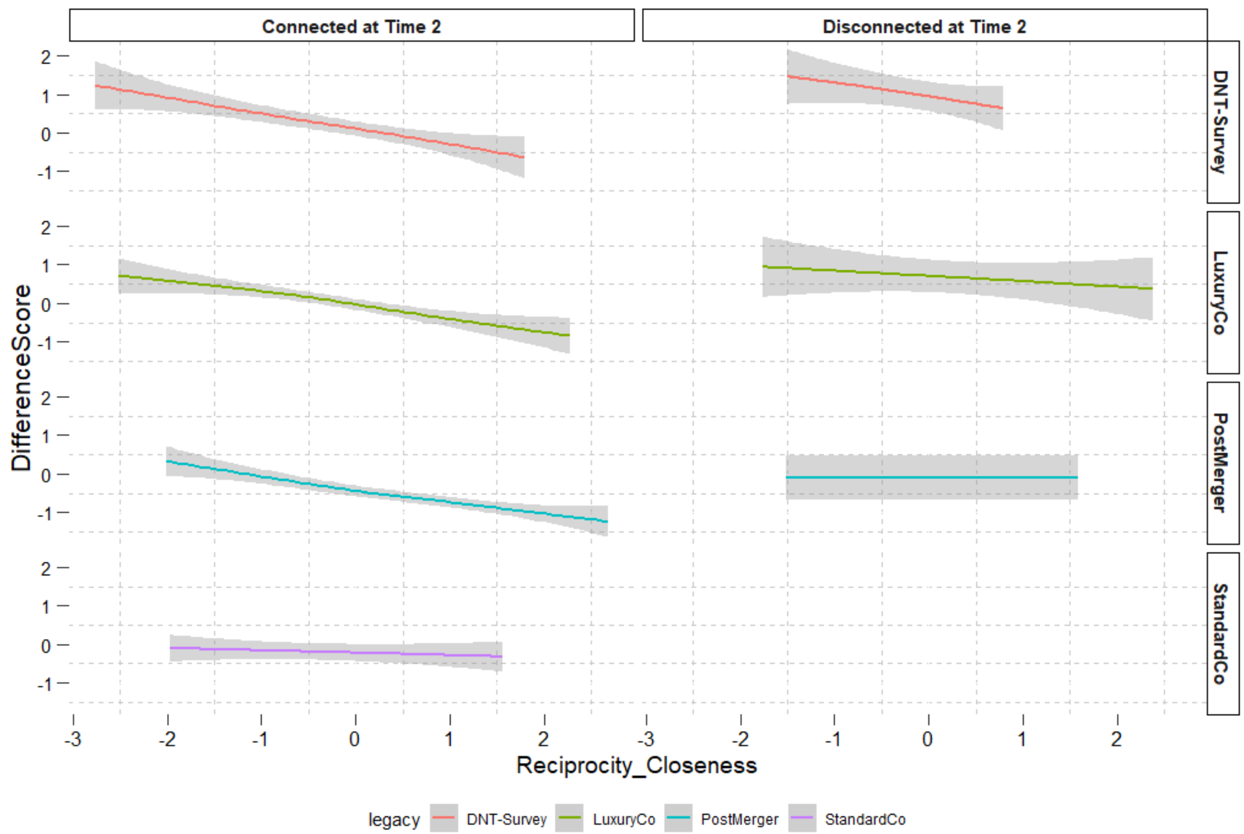


Figure 16. Language change as a factor of closeness centrality, legacy affiliation, and Time-2 disconnection

In Figure 16, I can see the interplay of closeness centrality, legacy affiliation, and disconnection in Time-2. As stated previously, "disconnected" means the individual has substantially fewer cross-cluster and in-cluster ties than are typical, rather than that the individual is completely isolated. For individuals who did not take the survey and are disconnected, there is no relationship between Time-1 closeness centrality and language change. For those who did take the survey but are not disconnected, however, there is a negative correlation. A similar dynamic is true for those who joined the organization after the merger. By contrast, individuals in LuxuryCo who are disconnected in Time-2 show a much stronger negative relationship between closeness centrality and language change than do those who are not disconnected.

Overall, I can see support for Assertion 1 but not for Assertion 2, although there are positions in the network that do have a relationship with language change.

Predictive Models of Time-2 Sentiment – Exploring Assertion 3

In this section, test my third assertion by building models that attempt to predict actor sentiment in Time-2.

> *A-3: Change in language between Time-1 and Time-2 is positively associated with language sentiment in Time-2.*

The third assertion might seem controversial. Why should language change between Time-1 and Time-2 positively correlate with language sentiment when I know adapting to an organization is stressful and cognitively demanding? I believe that the individuals who are changing are the ones most trying to show their investment in the organization, and that they are therefore the most likely to present a positive tone in their communications with others. This belief is based on the idea that language selection on an email is driven by both a need to communicate and a need to perform. However, there could be other variables that better predict language sentiment in Time-2. It is possible that having many ties in Time-1 indicates you are a person who uses language skillfully in order to maintain relationships within the organization. It could be that people at different legacy organizations are more or less positive in their use of language — either as a reflection of their fundamental wellbeing or because of a cultural perspective that indicates more or less use of positive language.

In the model selection procedure for Assertion 3, I combined each of these factors to identify the model which best explained the changes in language between Time-1 and Time-2.[11] After running each factor in isolation, I proceeded with the best of those models and added the other factors iteratively. As I did before, I evaluate each model using its BIC score, with lower scores indicating superior models. All continuous variables were Z-scored but otherwise unmodified. In total, I evaluated eight models:

- Model 1: Total Degree in the Structure Network
- Model 2: Language Change between Time-1 and Time-2
- Model 3: Legacy Organization Affiliation
- Model 4: Model 1 + Model 2
- Model 5: Model 2 + Model 3
- Model 6: Model 1 + Model 2 + Model 3
- Model 7: Model 2 + IsLuxury Binary Variable
- Model 8: Model 7 + Interaction Factor of IsLuxury and Language Change

Table 23 has the results. Each column represents one of the previously listed models. Each row represents a variable that may or may not be part of the model; variables not in a model are blank in the column for that model. A cell represents the factors coefficient and standard error. Significance for a given row-cell is indicated by + for p-values of 0.1, * for p-values < .05, ** for p-values < 0.01, or *** for p-values < 0.001.

[11] Although closeness centrality appears to impact language change, when I evaluated closeness centrality using the model selection procedure for Time-2 sentiment, it was never a competitive with a similar model using language change.

Table 23. Predicting Time-2 sentiment - model result summary. Selected models in bold.

Predicting Time-2 Sentiment	**Models**							
Variables	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Total Degree Centrality	-0.068* (0.028)			-0.031 (0.028)		-0.058+ (0.029)		
Language Change		0.143*** (0.027)		0.132*** (0.028)	0.135*** (0.029)	.119*** (0.030)	0.1259*** (0.027)	0.061+ (0.033)
Legacy: LuxuryCo			0.374*** (0.073)		0.424*** (0.073)	0.457*** (0.075)		
Legacy: StandardCo			0.112 (0.102)		0.189+ (0.103)	0.210* (.103)		
Legacy: MergedCo			-0.055 (0.074)		0.064 (0.080)	0.102 (0.082)		
Is Luxury Co							0.364*** (0.057)	0.396*** (0.058)
Interaction (Luxury * Change)								0.254*** (0.059)
Intercept	-0.162*** (0.028)	-0.146*** (0.027)	-0.283*** (0.056)	-0.148*** (0.027)	-0.326*** (0.058)	-0.355*** (0.059)	-0.266*** (0.034)	-0.281*** (0.034)
BIC	1635.88	**1612.61**	1612.83	1619.48	**1596.46**	1602.51	**1586.27**	**1579.89**

The best model is Model 8, which incorporates three variables: language change, a binary variable indicating affiliation with LuxuryCo, and an interaction factor between these two independent factors.

Because Model 8 is the best of these models, I know that language change and affiliation with LuxuryCo best explain the sentiment of an actor's tokens in Time-2. Membership in LuxuryCo, regardless of language change, is a positive factor affecting language sentiment. Language change is only a weak predictor of positive sentiment for individuals not affiliated with LuxuryCo, but it is a very strong predictor of positive sentiment for individuals who are part of LuxuryCo. Note that the best models dispensed with the general legacy categorical — which had LuxuryCo, StandardCo, and MergedCo as modeled values and "Did Not Take Survey" as a control — in favor of a simple LuxuryCo binary. As I can see in Figure 17, what mattered most in the prediction of language sentiment in Time-2 was your change in language between Time-1 and Time-2 and whether you were or were not a member of LuxuryCo.

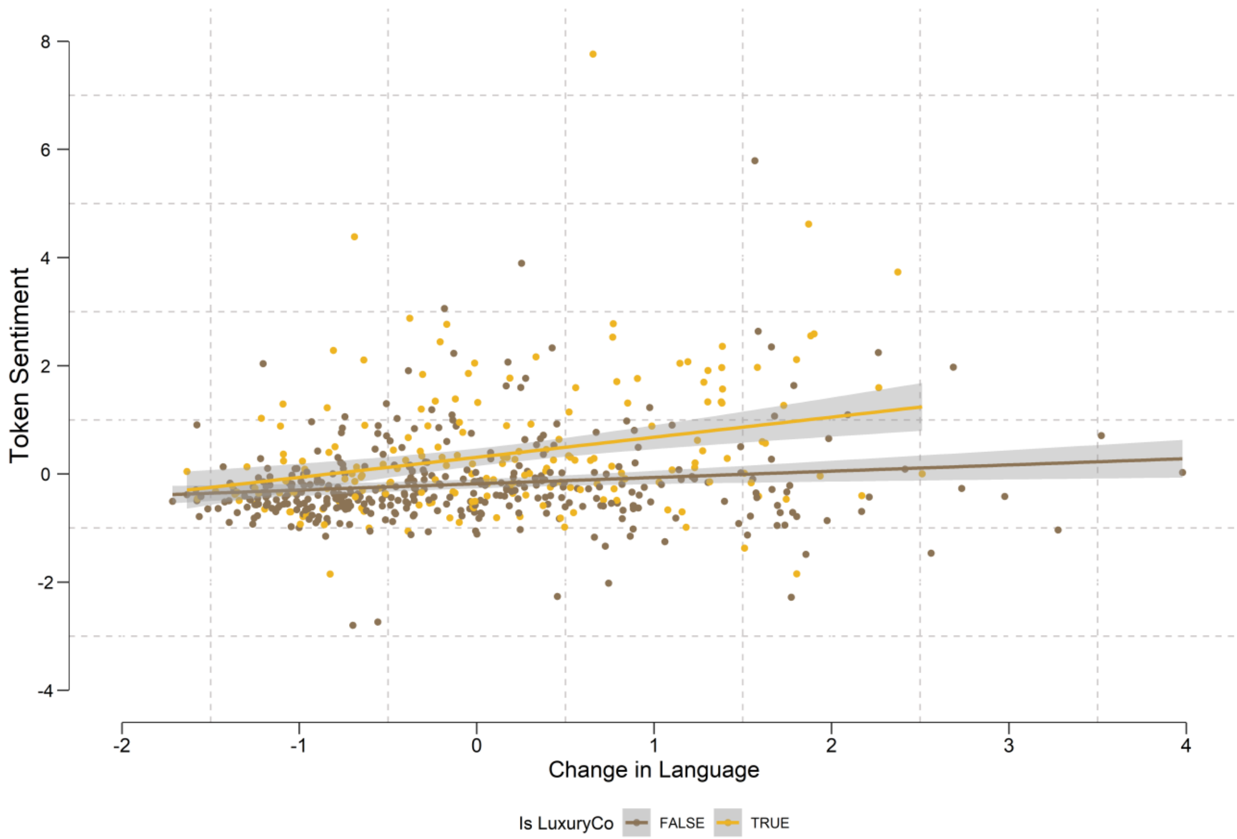


Figure 17. Token sentiment correlated with language change, with LuxuryCo as a grouping variable.

This general positivity in LuxuryCo tokens could be explained by multiple factors, including an organizational culture that privileges expressions of positive sentiment. It could also mean that LuxuryCo employees see the merger in a positive light. The interaction factor suggests that LuxuryCo employees who changed their language saw themselves as becoming more embedded in the new organization, and that the work required to change their language was not a burden, but rather part of the project of joining a grand new enterprise.

Retention to Time-2 and Sentiment – Assertion 4

An overall theme of these results so far is support for the Goffman perspective — that individuals in an organization are not only actively engaged in doing their work as best they can, but also in performing that work in ways that reflect themselves in a positive light. I wanted to further test for evidence to support that perspective by evaluating whether people who were no longer present by Time-2 were more or less positive in Time-1 than those who remained in the Time-2 data.

A-4: Individuals who left by Time-2 are more likely to use negative language in Time-1.

If there is a relationship, it suggests that the Goffman perspective holds additional merit. However, there are many reasons why the assertion may not hold. The data draws for both Time-1 and Time-2 are samples from the overall organization, and they are therefore not complete data sets. It could be that the sample is too biased or too nonrepresentative to show the relationship. It could also be true that language sentiment has nothing to do with whether an individual remains at the organization.

I performed a Mann-Whitney U-Test, a non-parametric test, which determines whether the sample mean between two group samples is statistically significant. Using that test, I found a statistically significant ($p < 0.001$) relationship between language sentiment and retention to Time-2. I can see a comparison of language sentiment and retention to Time-2 in Figure 18.

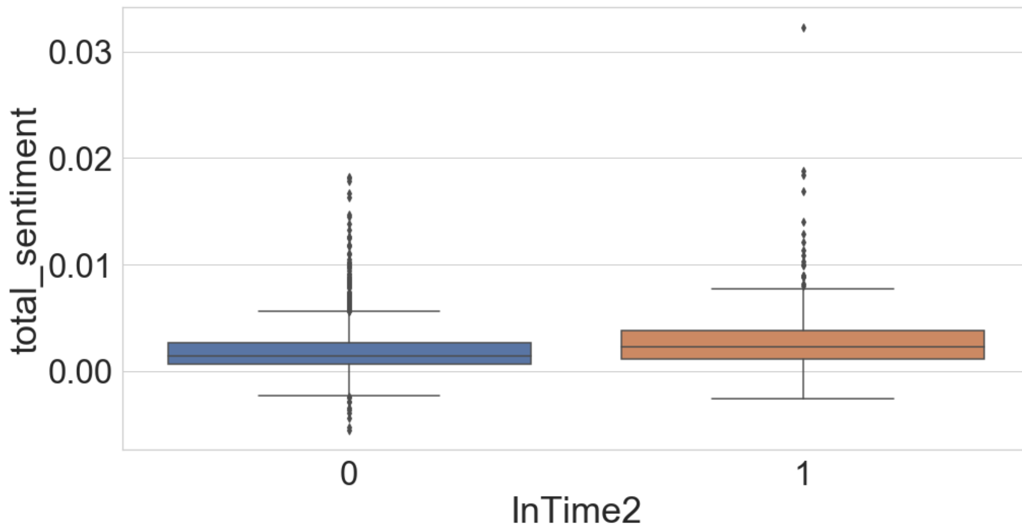


Figure 18. Box plot of language sentiment in Time-1, grouped by presence in Time-2

As I can see in Figure 18, individuals who are present in Time-2 have more positive language sentiment in their emails than those who are absent by Time-2. This outcome is supported by a comparison of means and lends further evidence to the idea that individuals who are at risk of departing tend to have lower language sentiment in their emails. In short, language sentiment has important implications for turnover and other relevant organizational outcomes.

Correlation between Sentiment and Survey Factors – Assertion 5

In the final result subsection, I want to return to the idea that an individual's language may reveal important things about how they feel about the organization — that I can use empirical evaluations of language to tell me things about individuals that I would otherwise have to use survey instruments to learn. In testing for Assertion 4, I have already established that language sentiment has some relationship with organizational turnover, or how likely an individual is to remain at the organization. Can language sentiment also give me insight into individuals' feelings about the organization and its need for change?

> *A-5: Language sentiment in Time-2 will have a weak positive correlation with self-reported commitment to the merger and belief in the organization's need for change in Time-2.*

I used Spearman correlation to test the relationships between language sentiment, merger commitment, and belief in the need for change. Spearman is a ranked order correlation that is robust to non-normal distributions. The results are in Table 24.

Table 24. Spearman ranked order correlations between Merger Commitment, Need For Change, and Token Sentiment

Merger Commitment	1.0		
Need For Change	0.253** (p = 0.003)	1.0	
Token Sentiment	-0.124 (p = 0.149)	-0.357*** (p < 0.001)	1.0
	Merger Commitment	Need For Change	Token Sentiment

As reported in Table 24, there is a moderate and statistically significant negative correlation between token sentiment and individual's self-reported perceptions of the organization's need for change. In other words, the more positive your language, the less likely you were to indicate that the organization needed to change. This is interesting because it is not what I expected. The Goffman-esque performance of positivity apparently does not reflect the individual's actual belief in the need for change. There is also a weak and not statistically significant relationship between token sentiment and merger commitment. I only had 138 reported values for merger commitment in this set, so more data may have made this relationship statistically significant as well, but it would probably always be weaker than the relationship between perceived need for change and token sentiment. Note also that both survey outcomes are subject to demand characteristics, which tend to concentrate the responses of everyone who chooses to respond into a narrow band of possible responses. The correlation between Need for Change and Token Sentiment is stronger than the relationship between Need for Change and Merger Commitment, albeit in the opposite direction. I can see the relationship between the Z-scored elements in Figure 19.

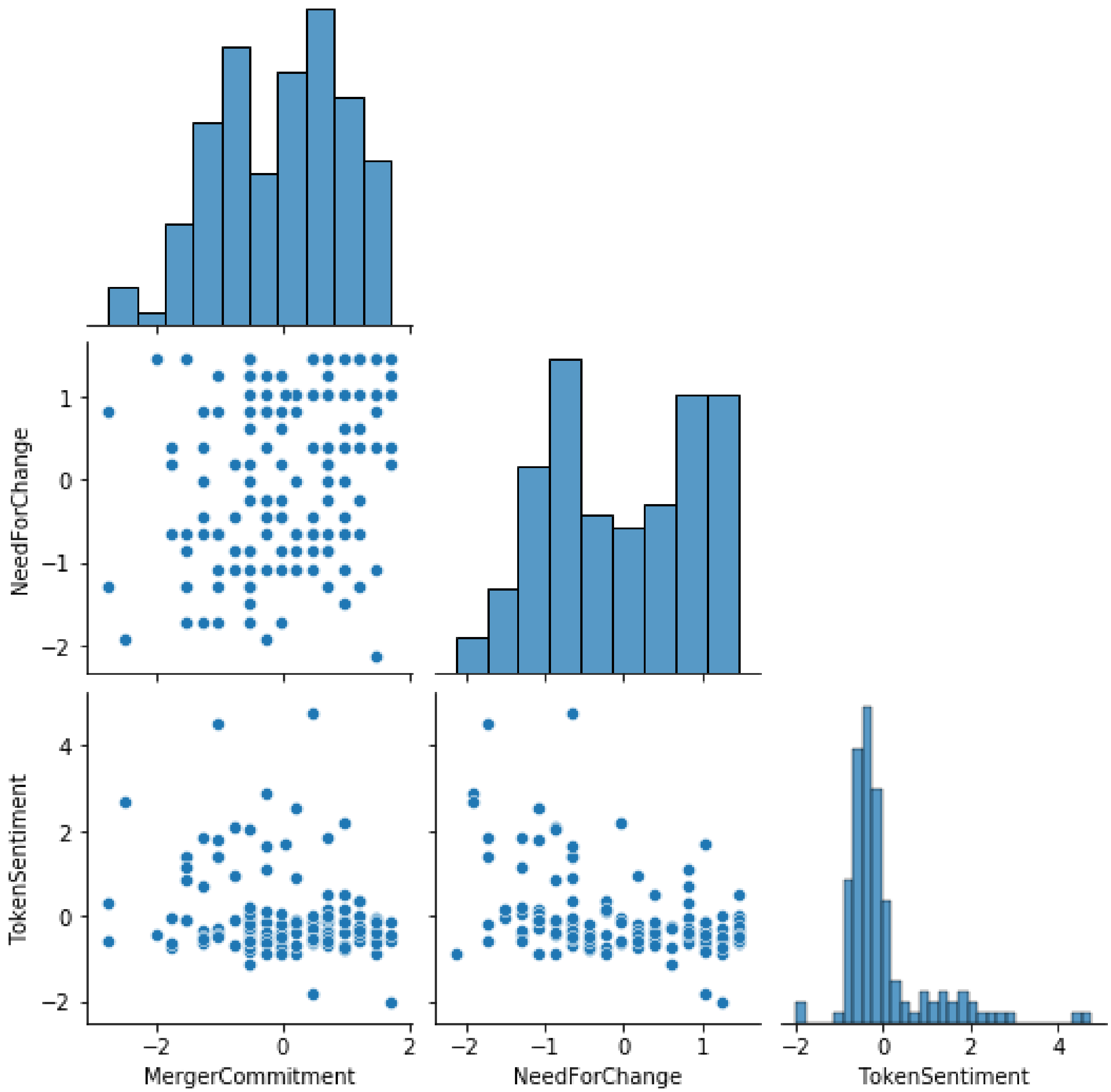


Figure 19. Correlation plots of merger commitment, perceived need for change, and token sentiment in Time-2.

Figure 19 helps me see that those who report the least need for the organization to change are also those whose token sentiment is most positive. Those who report that the organization does need to change tend to have lower or more neutral sentiment in their email language. Given the relationship between language change and token sentiment, one possible interpretation is that those who believe the organization least needs to change are those who have already changed the most.

Discussion and Conclusion

In this paper, I have empirically explored a horizontal merger and the related individual and organizational changes that occur over time. Conceptualizing the new united organizational identity as an innovation that must diffuse through the organization, I have examined, in the context of this diffusion problem, the organization as an information processor; how the culture of the organization reflects the organization itself; and how individuals adapt to organizations.

In so doing, I identified two research lacunae I wanted to explore: 1) the difficulty of and lack of existing work measuring cultural change at the individual level, and 2) the rarity of studying organizational socialization in a changing context.

To explore the first of these lacunae — the difficulty of measuring cultural change at the individual level — I made the statement that the language of the organization reflects its culture. This was in fact central to the definition of organizational culture when the concept was first developed, but it was later de-emphasized. With this statement in mind, I argued that a change in language reflects a change in culture, and that I can use modern natural language processing techniques to empirically quantify differences in language between any two entities. Because I needed a technique that was robust to variance in both corpus size and document size and that generated per-token scores, I created a novel technique, the Morgan Corpora Comparison method. I used this approach to quantify similarities in language between different people in a single time period and to quantify the difference in language for the same person between two time periods.

To make a start on the second of these lacunae — that organizational socialization in a changing context is rarely studied — I examined changes in organizational culture over time during a horizontal merger and made five assertions. I want to use these assertions to better empirically understand the merger and to identify necessary elements for appropriate COT modeling and simulation of horizontal mergers. These assertions were:

- A-1: Network centrality in Time-1 will have a negative association with language change between Time-1 and Time-2.
- A-2: A change in social network position is likely to lead to more change in language than no change in position.
- A-3: Change in language between Time-1 and Time-2 is positively associated with language sentiment in Time-2.
- A-4: Individuals who left by Time-2 are more likely to use negative language in Time-1.
- A-5: Language sentiment in Time-2 will have a weak positive correlation with self-reported commitment to the merger and belief in the organization’s need for change in Time-2.

In this work, I found strong evidence for three of the five assertions: A-1, A-3, and A-4. I found partial support for A-5, in that belief in the need for change had a strong correlation with language change, but commitment to the merger had only a weak relationship that was not statistically significant. For the remaining assertion, A-2, I did not find evidence. Too many individuals changed positions between Time-1 and Time-2 — although I do have some secondary analysis that identified particular paths as statistically different from others even with post-hoc correction. Instead, what mattered about these positions was whether someone ended up in Position 3 — functionally disconnected. This does not imply the individual was a network isolate, but instead that their reciprocal ties to other clusters and within their own cluster number far below average.

There were significant surprises along the way in doing this work. Contrary to typical expectation, I found that being at the center of the organization meant you — or at least your language — changed the least over time, rather than the most. However, I interpret this to mean either 1) that the individuals at the center had *already* changed their language before data was captured, or 2) that the organizational changes were made with the preferences and existing language of these central individuals already in mind. If pressed, I would assume that the first possibility is the more likely of the two. Future work could benefit from evaluating the preferences and communication patterns of organizational leaders who are involved in horizontal mergers and examining how those preferences and patterns relate to the language of the new organizational culture. The importance of position within the larger network suggests that, when modeling mergers for COT simulation, the structure or overall typology of the network must be considered in order to build good models.

I also found it striking that change in language between Time-1 and Time-2 tended to lead to higher sentiment in Time-2. It suggests that the words required to adapt to the cultural change were not new to the individuals — they did not need to be learned, so there was relatively little cognitive burden. Instead, the individuals who changed the most were those who were most engaged in and attached to the organization and therefore wished to appear the most positive. This, along with validation of Assertion 4, suggests that Goffman's perspective — that individuals are performing while they work — is valuable one for understanding the dynamics of a merger. This validation of Goffman's perspective further suggests that COT models of merger dynamics must take into account the local neighborhood of any given individual — the audience of that individual's performance. This, in turn, suggests that the distribution of knowledge to individuals within an organization is important for accurately modeling individual behavior, because the audience is almost as important to individuals' performances as the individuals themselves.

Throughout these explorations, legacy organizational membership regularly arose as a weighty and important factor. This suggests that legacy organization affiliation affects how individuals perceive and interact with the changing organization, which future efforts to model horizontal mergers should consider. As I discussed in the literature review, I chose to keep individuals' organizational identity constant through this work based on the first affiliation they provided in surveys. An alternative treatment that still relied on the survey data would have been to allow each individual's legacy affiliation to change over time if they gave different answers on subsequent surveys. This alternative treatment may have had several benefits. Are those who chose to retain their original identities more strongly rooted in old narratives and myths than those who later began identifying with MergedCo? Comparing alternative treatments could have allowed me to consider this and related questions.

Another possible approach would have been to determine an identity for each organization in Time-1 based on the words used in that organization. Then I could have identified each individual's "true" membership over time by assessing the words they choose to use and which organizational identity they most closely matched. This treatment would have the benefit of being empirically driven rather than reliant on self-reported survey data. However, it supposes that the organization’s identity at Time-1 is static enough that it can be used as an appropriate yardstick at Time-2, nine months later.

The partial validation of Assertion 5 helped give credence to this entire enterprise, which rests on the idea that quantitative models of language can provide insight into important organizational outcomes. Surveys, despite their limitations, remain the gold standard for understanding organizational dynamics, and I am still some time away from being able to predict how an individual feels about an organization from a randomly selected sample of 25 of their emails. However, the strong correlation of language sentiment to perceived need for change suggests that such work could ultimately be beneficial to the organizational change practitioner. It also suggests that, for future COT merger modeling efforts, individuals’ language choices may be able to serve as a proxy for what they know and feel about the organization.

I had expected language clusters to play a larger role in language change over time, but perhaps within the context of large organization, language is more malleable. It would be interesting to compare the language change dynamics of professionals in accounting and engineering — both fields with robust national accreditation standards and associated technical terminology — with those of professionals in other functional areas.

This work and its antecedents could benefit from considering literature on resistance to change, although I do not have the data in this work to explore how individuals interact directly with change initiatives over time, as is done in Isabella (1990) and Löwstedt (1993). For future work, COT models that explore the diffusion dynamics of a horizontal merger should take into account individual perceptions of the organizational change as a whole. In the simulation in Chapter 3, the resistance to change mechanism is activated in individuals who believe the organization’s performance is going down.

In this work, I have demonstrated that the language of an organization, as revealed by emails, can be used to understand a horizontal merger in progress. It can be used to evaluate who is changing, and it can also signal which areas of the business have successfully concluded their re-identification process. Furthermore, this work demonstrates that the language individuals choose to use has strong correlations to important attitudinal and organizational outcomes. It defines a novel method of comparing text corpora that is appropriate in the context of organizational change, and it uses that method to quantify language change over time in ways that have strong correlations to network effects, language sentiment, and ultimately survey outcomes. Finally, its exploration of its five assertions reveals factors that are important in understanding organizational mergers, which will benefit future empirical and COT modeling work.

References

Adolph, G., Buchanan, I., Horney, J., Jackson, B., Jones, J., Kihlstedt, T., . . . Quarls, H. (2001). *Merger Integration: Delivering on the Promise*. Retrieved from

Aguilera, R. V., Dencker, J., & Yalabik, Z. Y. (2006). Institutions and Organizational Socialization: Integrating Employees in Cross-Border Mergers and Acquisitions. In D. A. Griffith, S. T. Cavusgil, G. T. M. Hult, & A. Y. Lewin (Eds.), *Emerging Research Frontiers in International Business Studies* (Vol. 2, pp. 55): Academy of International Business.

Ahuja, M. K., & Carley, K. M. (1999). Network Structure in Virtual Organizations. *Organization Science, 10*(6), 741-757. doi:10.1287/orsc.10.6.741

Ahuja, M. K., & Galvin, J. E. (2003). Socialization in Virtual Groups. *Journal of Management, 29*(2), 161-185. doi:10.1177/014920630302900203

Argote, L., Aven, B. L., & Kush, J. (2018). The Effects of Communication Networks and Turnover on Transactive Memory and Group Performance. *Organization Science, 29*(2), 191-206. doi:10.1287/orsc.2017.1176

Argote, L., Beckman, S. L., & Epple, D. (1990). The Persistence and Transfer of Learning in Industrial Settings. *Management Science, 36*(2), 140-154. doi:10.1287/mnsc.36.2.140

Argote, L., & Hora, M. (2017). Organizational Learning and Management of Technology. *Production and Operations Management, 26*(4), 579-590. doi:10.1111/poms.12667

Argote, L., & Ingram, P. (2000). Knowledge Transfer: A Basis for Competitive Advantage in Firms. *Organizational Behavior and Human Decision Processes, 82*(1), 150-169. doi:10.1006/obhd.2000.2893

Argote, L., McEvily, B., & Reagans, R. (2003). Managing Knowledge in Organizations: An Integrative Framework and Review of Emerging Themes. *Management Science, 49*(4), 571-582. doi:10.1287/mnsc.49.4.571.14424

Bechky, B. A. (2011). Making Organizational Theory Work: Institutions, Occupations, and Negotiated Orders. *Organization Science, 22*(5), 1157-1167. doi:10.1287/orsc.1100.0603

Brown, J. S., & Duguid, P. (2001). Knowledge and Organization: A Social-Practice Perspective. *Organization Science, 12*(2), 198-213. doi:10.1287/orsc.12.2.198.10116

Burt, R. S. (2000). The network structure of social capital. *Research in Organizational Behavior, 22*, 345-423.

Burton, R. M., Obel, B., & DeSanctis, G. (2011). *Organizational design: A step-by-step approach*: Cambridge University Press.

Carley, K. M. (1990). Group stability: A socio-cognitive approach. *Advances in Group Processes, 7*(1), 44.

Carley, K. M. (1991). A theory of group stability. *American Sociological Review*, 331-354.

Carley, K. M. (1992). Organizational learning and personnel turnover. *Organization Science, 3*(1), 20-46.

Carley, K. M. (2002). Smart agents and organizations of the future. In L. A. Lievrouw & S. Livingstone (Eds.), *Handbook of New Media: Social Shaping and Consequences of ICTs* (pp. 206-220). Thousand Oaks, CA: SAGE Publications, Ltd.

Carley, K. M., Martin, M. K., & Hirshman, B. R. (2009). The Etiology of Social Change. *Topics in Cognitive Science, 1*(4), 621-650. doi:10.1111/j.1756-8765.2009.01037.x

Charron, J., Mattmann, C., & Zitting, J. (2014). Apache Tika. tika.apache.org: Apache Software Foundation.

Chatterjee, S., Lubatkin, M. H., Schweiger, D. M., & Weber, Y. (1992). Cultural differences and shareholder value in related mergers: Linking equity and human capital. *Strategic management journal, 13*(5), 319-334. doi:10.1002/smj.4250130502

Comer, D. R. (1991). Organizational Newcomers' Acquisition of Information from Peers. *Management Communication Quarterly, 5*(1), 64-89. doi:10.1177/0893318991005001004

Cook, K. S., & Emerson, R. M. (1978). Power, Equity and Commitment in Exchange Networks. *American Sociological Review, 43*(5), 721. doi:10.2307/2094546

Cooke, R. A., & Rousseau, D. M. (1988). Behavioral norms and expectations: A quantitative approach to the assessment of organizational culture. *Group Organization Studies, 13*(3), 245-273.
Cummings, J. N. (2004). Work Groups, Structural Diversity, and Knowledge Sharing in a Global Organization. *Management Science, 50*(3), 352-364. doi:10.1287/mnsc.1030.0134
Cyert, R. M., & March, J. G. (1963). *A Behavioral Theory of the Firm* (Vol. 2). Englewood Cliffs, NJ: Prentice-Hall.
Darr, E. D., Argote, L., & Epple, D. (1995). The Acquisition, Transfer, and Depreciation of Knowledge in Service Organizations: Productivity in Franchises. *Management Science, 41*(11), 1750-1762. doi:10.1287/mnsc.41.11.1750
Davies, H. T. O., Nutley, S. M., & Mannion, R. (2000). Organisational culture and quality of health care. *Quality in Health Care, 9*(2), 111-119. doi:10.1136/qhc.9.2.111
De Meo, P., Ferrara, E., Fiumara, G., & Provetti, A. (2011, 2011). *Generalized Louvain method for community detection in large networks.* Paper presented at the 11th International Conference on Intelligent Systems Design and Application, Cordoba, Spain.
Denrell, J., Fang, C., & Levinthal, D. A. (2004). From T-Mazes to Labyrinths: Learning from Model-Based Feedback. *Management Science, 50*(10), 1366-1378. doi:10.1287/mnsc.1040.0271
Drucker, P. F. (1988). The coming of the new organization. *Harvard Business Review*, 3-12.
Feldman, D. C. (1976). A Contingency Theory of Socialization. *Administrative Science Quarterly, 21*(3), 433. doi:10.2307/2391853
Feldman, D. C. (1981). The Multiple Socialization Of Organization Members. *Academy of Management Review, 6*(2), 309-318. doi:10.5465/amr.1981.4287859
Finkelstein, S., & Haleblian, J. (2002). Understanding Acquisition Performance: The Role of Transfer Effects. *Organization Science, 13*(1), 36-47. doi:10.1287/orsc.13.1.36.539
Goffman, E. (1961). *Asylums: Essays on the Social Situation of Mental Patients and Other Inmates*. New York, NY: Doubleday Anchor.
Hansen, M. T. (1999). The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits. *Administrative Science Quarterly, 44*(1), 82. doi:10.2307/2667032
Hansen, M. T. (2002). Knowledge Networks: Explaining Effective Knowledge Sharing in Multiunit Companies. *Organization Science, 13*(3), 232-248. doi:10.1287/orsc.13.3.232.2771
Hutto, C., & Gilbert, E. (2014). *Vader: A parsimonious rule-based model for sentiment analysis of social media text.* Paper presented at the Proceedings of the International AAAI Conference on Web and Social Media.
Isabella, L. A. (1990). Evolving interpretations as a Change Unfolds: How Managers Construe Key Organizational Events. *Academy of Management Journal, 33*, 7-41. doi:https://doi.org/10.5465/256350
Jaccard, P. (1912). THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.1. *New Phytologist, 11*(2), 37-50. doi:10.1111/j.1469-8137.1912.tb05611.x
Kane, A. A., Argote, L., & Levine, J. M. (2005). Knowledge transfer between groups via personnel rotation: Effects of social identity and knowledge quality. *Organizational Behavior and Human Decision Processes, 96*(1), 56-71. doi:10.1016/j.obhdp.2004.09.002
Klindzic, M., Braje, I. N., & Galetic, L. (2015). Exploring the Processes behind the New Organizational Identity Formation Following a Postmerger Integration. In (pp. 168-178). Zagreb: Centar za istrazivanje i razvoj upravljanja d.o.o.
Koehler, C. (2009). *The Process of Post-Merger & Acquisition Integration: An Application of Archetype Theory*.
Krackhardt, D., & Stern, R. N. (1988). Informal networks and organizational crises: An experimental simulation. *Social Psychology Quarterly*, 123-140.
Levin, D. Z., & Cross, R. (2004). The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management Science, 50*(11), 1477-1490.

Liang, D. W., Moreland, R., & Argote, L. (1995). Group Versus Individual Training and Group Performance: The Mediating Role of Transactive Memory. *Personality and Social Psychology Bulletin, 21*(4), 384-393. doi:10.1177/0146167295214009
Louis, M. R., Posner, B. Z., & Powell, G. N. (1983). The availability and helpfulness of socialization practices. *Personnel psychology, 36*(4), 857-866. doi:10.1111/j.1744-6570.1983.tb00515.x
Löwstedt, J. (1993). Organizing Frameworks in Emerging Organizations: A Cognitive Approach to the Analysis of Change. *Human Relations, 46*(4), 501-526. doi:10.1177/001872679304600405
Malloy, E., & Carley, K. M. (2021). *NetMapper's User Guide, v12*. Retrieved from https://netanomics.com/netmapper/
Morgan, G. P., & Carley, K. M. (2015). Computational Organizational Theory. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (Vol. 4, pp. 485-492). Oxford: Elsevier.
Morrison, E. W. (1993a). Longitudinal study of the effects of information seeking on newcomer socialization. *Journal of Applied Psychology, 78*(2), 173-183.
Morrison, E. W. (1993b). Newcomer Information Seeking: Exploring Types, Modes, Sources, and Outcomes. *The Academy of Management Journal, 36*(3), 557-589.
Nelson, D. L. (1987). Organizational socialization: A stress perspective. *Journal of Organizational Behavior, 8*(4), 311-324. doi:10.1002/job.4030080404
Nelson, R. E. (1989). The strength of strong ties: Social networks and intergroup conflict in organizations. *Academy of Management Journal, 32*(2), 377-401.
Oh, H., Chung, M.-H., & Labianca, G. (2004). Group Social Capital and Group Effectiveness: The Role of Informal Socializing Ties. *Academy of Management Journal, 47*(6), 860-875.
Oh, H., Labianca, G., & Chung, M.-H. (2006). A Multilevel Model of Group Social Capital. *The Academy of Management Review, 31*(3), 569-582.
Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American psychologist, 17*(11), 776.
Ostroff, C., & Kozlowski, S. W. J. (1992). ORGANIZATIONAL SOCIALIZATION AS A LEARNING PROCESS: THE ROLE OF INFORMATION ACQUISITION. *Personnel psychology, 45*(4), 849-874. doi:10.1111/j.1744-6570.1992.tb00971.x
Pablo, A. L. (1994). Determinants of acquisition integration level: A decision-making perspective. *Academy of Management Journal, 37*(4), 803-836.
Pettigrew, A. M. (1979). On Studying Organizational Cultures. *Administrative Science Quarterly, 24*(4), 570. doi:10.2307/2392363
Porter, M. E. (1987). From Competitive Advantage to Corporate Strategy. *Harvard Business Review*(May-June), 43-59.
Reagans, R., & McEvily, B. (2003). Network structure and knowledge transfer: The effects of cohesion and range. *Administrative Science Quarterly, 48*(2), 240-267.
Rodan, S., & Galunic, C. (2004). More than network structure: How knowledge heterogeneity influences managerial performance and innovativeness. *Strategic management journal, 25*(6), 541-562.
Rogers, E. M. (2010). *Diffusion of innovations*: Simon and Schuster.
Schein, E. H. (1988). Organizational Socialization And The Profession Of Management. *Sloan Management Review, 30*(1), 53. Retrieved from https://search.proquest.com/scholarly-journals/organizational-socialization-profession/docview/224963008/se-2?accountid=9902

https://cmu.primo.exlibrisgroup.com/openurl/01CMU_INST/01CMU_INST:01CMU?genre=article&atitle=Organizational+Socialization+And+The+Profession+Of+Management&author=Schein%2C+Edgar+H&volume=30&issue=1&spage=53&date=Fall+1988&rft.btitle=&rft.jtitle=Sloan+Management+Review&issn=0019-848X&isbn=&sid=ProQ%3Aabiglobal_00424075
Schweiger, D. M., & Goulet, P. K. (2005). Facilitating Acquisition Integration Through Deep-Level Cultural Learning Interventions: A Longitudinal Field Experiment. *Organization Studies, 26*(10), 1477-1499. doi:10.1177/0170840605057070

Simon, H. A. (1991). Bounded rationality and organizational learning. *Organization Science*, *2*(1), 125-134.
Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, *28*(1), 11-21.
Stryker, S. (2008). From Mead to a Structural Symbolic Interactionism and Beyond. *Annual Review of Sociology*, *34*(1), 15-31. doi:10.1146/annurev.soc.34.040507.134649
Sung, W., Woehler, M. L., Fagan, J. M., Grosser, T. J., Floyd, T. M., & Labianca, G. (2017). Employees' Responses to an Organizational Merger: Intaindividual Change in Organizational Identification, Attachment, and Turnover. *Journal of Applied Psychology, 102*(6), 910-934.
Thom-Santelli, J., Millen, D. R., & Gergle, D. (2011). *Organizational acculturation and social networking*.
Tsai, W. (2001). Knowledge transfer in intraorganizational networks: Effects of network position and absorptive capacity on business unit innovation and performance. *Academy of Management Journal*, *44*(5), 996-1004.
Van Maanen, J. (1978). People processing: Strategies of organizational socialization. *Organizational Dynamics*, *7*(1), 19-36. doi:10.1016/0090-2616(78)90032-3
Van Maanen, J., & Schein, E. H. (1979). Toward a theory of organizational socialization. *Research in Organizational Behavior*.
Wilbur, W. J., & Sirotkin, K. (1992). The automatic identifiation of stop words. *Journal of Information Science*, *18*, 45-55.

Chapter 3: Simulating a horizontal merger with data-driven multilevel simulation

Abstract
In this chapter, I describe a new multilevel model of organizational change based in computational organizational theory (G. P. Morgan & Carley, 2015) and use that model to simulate the horizontal merger of two large multinational companies. The model is able to replicate prior findings of similar models and to emulate stylized facts relevant to organizational performance and mergers. In a case study using data derived from the horizontal merger of a large multinational, the model can both predict an important future outcome and correlate to the merger's initial performance as represented by stock prices. Such a model could be useful to organizational theorists interested in the performance of stylized organizational forms as well as to organizational leaders, managers, and change management professionals interested in gathering insight about potential organizational futures.

Introduction
Even with the best efforts at due diligence by the acquiring firm and its financial backers, the success of a horizontal merger is not guaranteed. In practice, a slight majority of horizontal mergers fail (Adolph et al., 2001; Porter, 1987). Because the acquiring firm makes substantial investments into understanding the market position and debt of the acquired firm before the merger, post-mortem analyses often attribute failed mergers to "integration issues" (Cartwright & Cooper, 1994; Epstein, 2005), also called the "people factor" (Giessner, Viki, Otten, Terry, & Täuber, 2006; Kansal & Chandani, 2014). Integration problems include difficulty retaining key staff, an inability to deduplicate common internal competencies, and an inability to achieve the united strategy of the combined firm. One can imagine the merger advocate shouting from the sidelines, "Get on with the work already!"

In fact, these challenges are entirely predictable in modern times. As more and more organizational capabilities have evolved from rote standard practice into dynamic, expert-driven processes, more and more work is done by specialists who must coordinate their efforts and capabilities in order to create value. This results in what Drucker called the "virtual organization" (Drucker, 1988). Because these specialists must communicate and collaborate to create a unified work product, their patterns of interaction are themselves valuable organizational knowledge (Argote & Hora, 2017) — knowledge that is rarely taken into account by the leaders making decisions about the merger. As a result, high-performing individuals are often retained, but shorn entirely of the complex web of other specialists who helped them achieve their results. Without their collaborators and partners in creating value, there is no work to "get on with."

The merging organizations, however, retain their power as symbols and identities for the employees of the unified company. If members of these legacy organizations are not actively integrated into the new merger identity, their previous identity will remain extremely relevant and important to them. This may ultimately influence strategic decisions of the merged entity;

for example, decisionmakers who identify with their pre-merger company may retain legacy corporate IT systems that are not ideal for the new organization (Vieru & Rivard, 2014).

Furthermore, the market environment does not stand still while the organization takes it time to change and evolve. Established competitors and potential disruptors are constantly seeking to change the rules of the game to their own advantage. What was once a product differentiator becomes a capability everyone has; what was once a standard capability gets outsourced to a smaller firm. The organization must continue to learn and grow if it wishes to retain its ability to attract customers, and therefore survive, even as it wrestles with its internal identity.

To understand the possible futures an organization may experience, I can turn to Computational Organizational Theory (COT) simulation (G. P. Morgan & Carley, 2015). I use COT simulation techniques to understand possible futures emergent from the interactions of thousands of individuals over time. Classic COT simulation techniques, such as the Garbage Can Model (M. D. Cohen, March, & Olsen, 1972) and the Mutual Learning Model (March, 1991), were used to understand and consider highly stylized organizations. However, modern simulations can take advantage of existing data to instantiate a specific organization from the best data available, a technique called data-driven modeling (Lanham, Morgan, & Carley, 2011).

In this paper, I introduce a new COT simulation, the Unified Network Model, which extends previous personal work to create an agent-based multilevel organizational simulation that can be used to model horizontal mergers. This model embodies a multilevel theory of individuals and multiple overlapping organizations, with both individuals and organizations represented as cognitively bounded actors (Simon, 1991). Cognitively bounded actors have natural limits to their ability to recall, think, act, and make decisions. Classical economic modeling, by contrast, focused on fully rational actors, who were presupposed to have all available information and to therefore be able to take any necessary action. A multilevel theory of how cognitively bounded actors interact to create organizational outcomes would be useful for the study of horizontal mergers and other organizational contexts (Bechky, 2011).

Related Work

This model is a network-based, multilevel adaptation of the Unified Hierarchical Model (G. P. Morgan & Carley, 2012, 2014). The Unified Hierarchical Model described the activities and information processing of an organization by integrating two different COT models (G. P. Morgan & Carley, 2015): the Hierarchical Garbage Can Model (Carley, 1986b) and the Mutual Learning Model (March, 1991). In addition to integrating these two models, the Unified Heirarchical Model accounted for actor bias by adaptating mechanisms from the Participation Model (J. H. Morgan, Morgan, & Ritter, 2010). Before discussing the extensions of the current model, I will introduce the general mechanisms of the Unified Hierarchical Model and how it integrated the Mutual Learning Model and Hierarchical Garbage Can Model.

In the Unified Hierarchical Model, I had to integrate organizational processes of the Mutual Learning Model and Hierarchical Garbage Can Model. I begin with a short review of these models.

The Mutual Learning Model presents a turn-based model of how organizational knowledge develops from the aggregate of individuals' knowledge, and how organizations use this knowledge to maintain their performance in a turbulent external environment, which can be thought of as the market in which these organizations operate. In each turn of the simulation, elements of the external environment may change due to unlisted exogenous factors, described with a uniform probability of change per bit. The organization does not interact with the environment directly, but instead is able to learn about it from high-performing individuals. The organization learns from the consensus of these high performers and socializes individuals over time to the views it has developed. However, because all individuals eventually hold the same views, the organization is no longer able to learn, and its performance degrades quickly. March presents turnover, which brings in new actors who are not yet socialized to the views of the organization, as a solution to this problem and demonstrates that turnover, as a mechanism, can help an organization maintain its performance in a changing world.

The Mutual Learning Model presents an organization as a structureless collection of individuals and the organization's understanding of reality — its code — as a product of these individuals' views as they develop over time. In this view, the organization it aggregates and processes the views of those individuals over time in order to best match its environment.

The Hierarchical Garbage Can Model (Carley, 1986b), implemented in a simulation model called GARCORG (Carley, 1986a), presents a turn-based simulation similar to that described by Padgett (1980). Padgett described an extension of the original Garbage Can Model (M. D. Cohen et al., 1972) where the organization has an explicit formal structure — a hierarchy with multiple levels. This formal structure, which is hierarchical but can be described with network formalism, has directional reporting ties from each lower level to each upper level. On each turn, information comes into the organization at a specific randomly-chosen point, and individuals report that information through their hierarchical reporting ties. Higher-level individuals within the organization may or may not be able to effectively process and take advantage of the information. Using a wide variety of stylized hierarchical forms, the model demonstrated that the structure of an organization may create "hot-spots" — areas within the organization where even the most talented individuals will fail. These hot spots occur because an individual's only reporting line runs through an already overburdened manager. As such, whenever that individual attempts to update the organization, the update is missed because the manager fails to process it and pass it along. After enough of these misses, the individual may be let go because the organization is performing poorly in the area for which that individual is responsible. This can happen regardless of the actual capabilities of the individual to perform their job duties!

The Hierarchical Garbage Can presents the organization as a structured collection of individuals who must effectively process information. The structure of the organization — that is, the network of ties between individuals — determines how successful it will be at processing and disseminating this information.

The Unified Hierarchical Model (G. P. Morgan & Carley, 2012) incorporates both the Mutual Learning Model and the Hierarchical Garbage Can Model into a single model. It presents an organization with a hierarchical structure, with information passing along reporting ties. Each turn, one or more of the organization’s members receives new information about changes in the environment and passes that information along to others. The organization learns from its high performers and informally socializes everyone in the organization to its views. In this way, the Unified Hierarchical Model allows for both explicit structure — which may be or more less useful to the organization as an information processor (Cyert & March, 1963) — and additional implicit communication between individuals, implemented through the socialization mechanism. To survive, the organization must function simultaneously as a processor of information and as an effective aggregator of it.

The Unified Hierarchical Model also added multiple hiring mechanisms that could be used to explore the impact of similarity bias in hiring. These hiring mechanisms ranged, explicitly, in that deliberative capacity the organization exercises in its hiring process. By deliberative capacity, I mean how much the organization can deliberate before deciding to hire a new person. Hiring is expensive for an organization, so some organizations want to a more thorough process to vet candidates. Three representative mechanisms were built into the code: hiring at random, hiring the first “good enough” person, and picking the candidate considered “best” — or most similar to the committee — out of a large pool. The decision to vote for a candidate was adapted from the Participation Model (J. H. Morgan et al., 2010), but the Unified Hierarchical Model formalized the action as a discrete choice rooted in perceived similarity. The Unified Hierarchical Model found, as March predicted, that the more random the selection mechanism, the more successfully turnover functioned as a mechanism to combat environmental turbulence. (G. P. Morgan & Carley, 2014).

Although the Unified Heirarchical Model was a step forward in modeling the organization and the multiple roles a healthy organization must perform simultaneously, it was designed explicitly with organizational hierarchies in mind. As organizations become more virtual (Drucker, 1988), a trend which has only accelerated in recent years, strongly hierarchical organizations are less and less common. Instead, work is typically accomplished through the interaction of many individuals, each of whom possesses unique knowledge and capabilities.

The Unified Network Model

In this work, I make a series of notable changes to the Unified Hierarchical Model. These changes include: 1) generalizing the social and knowledge network formalism, 2) making the spread of information more boundedley rational (Simon, 1991), 3) adding a similarity bias to the determination of whether an individual accepts messages from others, 4) adding a more sophisticated assessment of organizational performance adapted from Fang, Lee, and Schilling (2010), and 5) allowing multiple organizations to coexist within a single organizational envelope at the same time, thereby extending the new model into an active multilevel model that can support examination of the possible futures of merging organizations.

First, I replace the specific notion of hierarchical design with a more general network formalism; modeling hierarchies is still possible, if desired, but it is no longer the only option. As such, the simulation now requires the specification of two distinct networks: the social network, which connects simulated people to each other, and the knowledge network, which connects simulated people to knowledge bits. For the social network, the system now supports reading in-network files as link lists as well as several network generators, including a hierarchy generator, the Fang et al modified Connected-caveman generator (Fang et al., 2010), and an Erdos-Renyi (Erdős & Rényi, 1960) random tie generator. For the knowledge network, the tool supports random and hierarchical knowledge generators. It can also read link-list format network files.

In the Unified Hierarchical Model, information only flowed upward in the organization, so it was not necessary to explicitly bound the number of messages a given agent could send in a turn — that would naturally be constrained by the number of messages that agent could receive. In a network that allows for symmetric links and an unlimited number of ties per agent, however, individuals should not be able to message everyone they are connected to if I think of them as both cognitively and socially bounded (Simon, 1991). To account for this, I set a base threshold for how far messages can travel from their originators. Most information spreads to a neighborhood no more than three to four (3–4) jumps from its origin.

In the Unified Network Model, recipients of information may not accept the information from someone who is too different from them. In reality, humans often exhibit a homophily bias (McPherson, Smith-Lovin, & Cook, 2001): They prefer others who are like them, whether in surface characteristics (sometimes called shallow homophily) or in goals, beliefs, and knowledge (sometimes called deep homophily). In code, the agent perceives both themselves and others around them, assessing the difference between themselves and each of their alters. Every time an individuals shares information with the agent, the agent decides probabilistically whether to accept the information. The agent is more likely to accept the information the more similar the other person is to them. Because the strength of the homophily in work contexts can be debated (Fu, Nowak, Christakis, & Fowler, 2012), I allow this mechanism to be turned on or off through experimental controls.

When calculating the performance of organizations in the Unified Hierarchical Model (G. P. Morgan & Carley, 2012), I assessed the organization's accuracy as compared to the external environment with a calculation as similar as possible to that described for the Mutual Learning Model (March, 1991). Accuracy, in this context, refers to the idea that being right about the market in which you operate means you should be able to perform well. In code, the environment is characterized by a long series of bits (1s and -1s). If an organization was correct in its perception of every external environment bit, its performance was rated at 100%, while if the organization was wrong on every bit, its performance was rated at 0%.

The problem with this straightforward approach is that, as reality changes, every organization is likely to achieve performance close to 50% as even views that may have ossified long ago randomly become correct again. This phenomenon is often called the Drunkard's Walk. To address this problem, I adapted the payoff function described in Fang et al. (2010). This function specifies a knowledge interdependence constant, *s*, which must be an integer equal to or greater than 1 and as high as *c*, the total number of bits in the external environment. The Mutual Learning Model effectively had a knowledge interdependence value of 1. The value of *s* sets the baseline performance of an organization at 0.5^s — that is, no better than guessing randomly.

At settings above 1, this knowledge interdependence constant, *s*, also has implications for how knowledge assignments should be distributed to the agents in code. Rather than distributing knowledge to agents independently of previous knowledge assignments, clusters of knowledge are assigned only if the individual agent has available capacity. If the agent has only partial capacity available for an assigned cluster of interdependent knowledge, then available bits are assigned randomly within that cluster.

Finally, I extend the concept of the organization and organizational membership to allow multiple organizations to be present and operating simultaneously. In the real world, individuals are often members of multiple organizations, whether those are multiple independent organizations or multiple suborganizations within a larger organization. Each organization, meanwhile, is learning from its high performers and socializing its members to its beliefs. Because information and membership are not likely to be equally distributed across a population, separate organizations may decide on contradictory views of reality, and suborganization members may be socialized into beliefs with which the larger organization disagrees.

Because individuals can now be members of multiple organizations, I also added a resistance to change mechanism that operates based on group membership. When a recipient evaluates information from a sender, the recipient considers group membership and may reject messages from an out-group individual if the recipient's own group is doing poorly. In the real world, resistance to working with perceived out-group individuals is common (Sidle, 2006).

Several of these mechanisms were added not only to support the change from modeling organizational hierarchies to modeling networks, but also to better support the modeling of organizations going through horizontal mergers. These include the homophily bias in accepting information from others, resistance to change, and the cooperation of multiple organizations at the same time.

The resulting new model, the Unified Network Model, is a turn-based active multilevel model that is suitable for modeling horizontal mergers and requires the organization to effectively learn from its members and process new information. Individuals may be members of multiple organizations, and these organizations may hold contradictory beliefs about the environment in which all of them operate simultaneously.

Initialization

The Unified Network Model allows a modeler to specify characteristics of both the environment and the organization. Characteristics of the environment include its complexity and its turbulence. For the organization, the modeler may specify the number of employees, how those employees are connected, how knowledge is distributed to employees within the organization, and how hiring, if any, takes place. Network generators can be used to create highly-stylized organizations, such as strict hierarchies or cellular structures, or network characteristics can be read from inputted data. This flexibility allows the model to be used both intellectively, to explore organizational forms writ large, or in an emulative fashion, drawing on the data of a real organization to consider its possible futures. In this work, I will take advantage of both capabilities.

The model parameters also allow for a warmup period. During this warmup period, all turns occur as they do during the normal run of the simulation (described in the following subsection), except that information spread cannot occur across inter-organizational ties. This allows each organization that exists in the model to have a period of self-development before the start of the full simulation.

Operation

Once initialized, the simulation proceeds through a number of turns. During each turn, these steps take place:

1. Reality change
2. Network information spread
3. Organizational inference
4. Implicit socialization
5. Turnover

Phase 1, *reality change*, addresses the changes in the environment that may occur. The turbulence variable, a value between 0 (inclusive) and 1 (inclusive), indicates the likelihood that any particular bit of the environment flips between its two possible states. Higher values indicate a more turbulent environment. Typically, turbulence should be quite low, because the environment as a whole is usually perceived to be mostly static from one week to the next. One could imagine, and indeed the GARCORG model (Carley, 1986a) supported, multiple ways of implementing turbulence. These additional models of turbulence include clumpy change — where change, if it occurs, occurs in larger quantities — as well as change stickiness, where a bit that has flipped recently is more likely to flip again. While these capabilities are interesting for modeling various change scenarios, I have left them for future work.

In Phase 2, *network information spread*, information about changes in reality travels along coworker ties. First, agents who are responsible for the bit that changed may or may not notice the change. Multiple agents may be responsible for the same bit, in which case each has a chance to notice the change. If an agent does not notice the change, then nothing happens. If an agent with responsibility for that bit does notice the change, then that individual attempts to inform a number of their coworker ties. The number of individuals they attempt to inform is determined randomly, but they always attempt to inform at least 1 other person. The alter to be informed is selected randomly from their existing ties.

However, this alter may not be able to accept the update. This may occur for one of three reasons: 1) the alter's inherent capacity for updates (an implementation of bounded rationality, Simon (1991)), 2) an individual level homophily bias (McPherson et al., 2001), or 3) a group-level resistance to change mechanism.

When a message arrives, an individual with spare available capacity can decide whether to act on the message. In this case, acting on the message means accepting the update and potentially passing it on to others. When an update arrives in this simulation, the recipient considers the sender and calculates the perceived distance in beliefs between the sender and themselves, as described in Equation 10. The perceived distance is a Euclidean distance calculated across c

dimensions and will range in value from 0 to the square root of *c*. This distance may be more or less noisy depending on the accuracy of perception.

Equation 10. The social distance, d_{ij}, between one actor, i, and another actor, j

$$d_{ij} = \sqrt{\sum_{b=1}^{c} (i_b - j_b)^2}$$

This distance value, along with the average social distance of everyone within the receiver's local neighborhood, is used to inform a McFadden (1980) discrete-choice equation. The output of this logit-transform equation, p_{ij}, informs a probability that the recipient will accept the message. The logit-transform compares the distance to the specific sender *j* against the average distance between the *n* agents in the recipient *i*'s local neighborhood .

Equation 11. The probability of acceptance of an update by agent i from agent j

$$p_{ij} = 1 - \frac{1}{e^{\left(\frac{\sum_{k=1}^{n} d_{ik}}{|n|} \Big/ d_{ij} \right)}}$$

The inclusion of homophily bias implies that bridging ties — ties that connect discrete groups of people — will often be less successful at transmitting information, because the bridge actor's updates may be accepted less often. This code mechanism also produces an emergent social pressure effect: Agents will be more accepting of updates from strangers if many people they interact with hold different views from their own, an effect we also observe in reality. If almost everyone an agent knows is quite similar to themselves, on the other hand, they will often reject updates from people outside that group. This emergent social pressure effect limits the utility of turnover and suggests that a real-world organization may choose to be selective in its hiring not because they do not value a diversity of opinions, but because those who are too different will be ignored within the organization.

As part of my focus on the challenges of merging organizations, I have also incorporated a resistance to change mechanism. Individuals within groups going through a change-management process often resist change from those they consider "out-group" (Sidle, 2006) — even when the change in fact benefits them (Giessner et al., 2006). In code, this mechanism works at the group level. When a message arrives from a given sender, the recipient compares their own group memberships to those of the sender. If they do not share groups in common, a group at random is selected from what they do not share, and if that group's performance is currently on a downward trend, the sender's message is not accepted and is effectively ignored.

This mechanism is likely to inhibit the successful flow of information between elements of the organization and negatively affect overall performance, especially when the organization's structure relies on successful interoperation between the merging units.

In Phase 3, *organizational inference*, the organization attempts to learn from its high performers. Each organization may have its own view on what constitutes a high-performer, because each organization may value different things. As such, each organization has its own set of high performers and may therefore learn different and contradictory things each turn. In the simulation, high performers are those who are doing better than the organization as a whole at understanding the environment, as weighted by the organization's view of what matters. The model uses the consensus of these high performers to evaluate whether the organization should update its understanding of reality, called its organizational code. If most high performers agree that a certain bit has changed , then the organization will probably change. If the high performers are split on whether that bit has changed, then the organization will almost certainly not change. The equations used to identify consensus are listed in Appendix C and come from G. P. Morgan and Carley (2012). These equations have not changed since that work.

In Phase 4, *implicit socialization*, the organization socializes individual members of the organization to agree with its perspective on the environment. This perspective may be at odds with the actual nature of the environment and may even disagree with information the same individuals have recently shared with others. Some amount of socialization is helpful to each organization, because it allows information to be distributed outside formal ties and ensures that individuals within the group perceive others within the group as being similar. There is no prescriptive guidance for real-world organizations on socialization. Generally, real-world organizations that emphasize socialization focus on making sure that new hires quickly become useful members of the staff whose views are accepted by others. Other real-world organizations place less emphasis on socialization and allow a diversity of opinions to exist, which can be helpful when the environment in which the organization operates is changing rapidly. In code, our simulated organizations may differ on the nature of the environment, so actors may be socialized into conflicting views, even within one turn. When there are multiple organizations in the simulations, the order of socialization is set randomly each turn to avoid primacy effects. By default, larger organizations socialize their team members less frequently. I use the floored cube root of the organization's size to determine socialization cadence, or how often socialization occurs in the turn-based simulation, but the modeler may set any value they prefer.

In Phase 5, *turnover*, simulated individuals leave the organization. In the real world, individuals may choose to leave or organizations may choose to let people go for underperformance. In code, individuals may choose to leave due to many exogenous factors that are not modeled in the current simulation. Organizations may also choose to replace individuals due to failure to perform. When the latter mechanism is active, individuals are replaced if more than 50% of the bits for which they are responsible in the organizational code are incorrect for a consecutive period longer than a defined grace period.

Staff members may be replaced in several ways. The Unified Network Model retains the hiring mechanisms from the Unified Hierarchical Model, including random hiring and hiring based on similarity to a selection committee. Because the Unified Network Model is adapted for networks, hiring committees, when used, are drawn from the departing employee's current ties — the reasoning being that any replacement employee will be expected to interface with the former coworkers of the departing individual.

Outcomes

The Unified Network Model includes several organizational outcome metrics that are recorded each turn and reported per represented organization. These are:

- ScorePercent: This is the percentage accuracy of the simulated organization's code as compared to its external environment. Following the example of Fang et al. (2010), I add a knowledge interdependence value. Essentially, there is no partial credit for having most of the elements of a knowledge chunk, or collection of bits. When this interdependence value is set to 1 (meaning knowledge is completely independent), ScorePercent is the number of bits that are accurate against the total number of bits. By default, the interdependence value is set to 1 if it is not provided. ScorePercent ranges from 0 to 100, with 100 indicating perfect performance and 0 indicating complete inaccuracy.
- BitsSet: This is the number of organizational bits that the organization has set through its group inference mechanism. By default, all organizations start with their bits in a neutral "0" position. As member consensus emerges, as discussed in the previous subsection, these bits are set to "-1" or "1." Over time, this BitsSet value will rise from 0 to the number of bits in the organization's code.
- MaxBitWeightDifference: Each organization may weigh the value of different pieces of knowledge differently. This is implemented as a double vector the same length as the organizational code. This measure, which takes the absolute value of the difference in weights for all elements of the organizational code, only rarely changes during the simulation.
- MaxCodeDistance: How different is this organizational code from all other organizational codes in the simulation? This is a measure of inter-organizational disagreement. This is the difference in vector between this organizational code and other codes at this point in time.
- DisagreementScore: How unified are the organization's high performers? This is a measure of intra-organizational disagreement. I calculate DisagreementScore by taking the number of high performers who disagreed with the majority position each turn and dividing it by the total number of high performers that turn. Internal disagreement in moderation is valuable to the organization.

In addition to these organizational outcome metrics, the simulation also reports individual level outcomes, including:

- Updates Sent: How many times did this agent send an update to another agent over the course of the entire simulation?
- Updates Ignored: When this agent sent an update to another agent, how often did the receiving agent have no capacity to handle the update? This is distinct from update refusal, below.
- Updates Refused: When this agent sent an update to another agent, how often did the receiving agent choose to ignore the update due to bias? (This value is always 0 if the homophily bias and resistance to change mechanisms are turned off.)
- Updates Accepted: This is the number of updates sent by this agent that were successfully received. This number is inferred by subtracting Updates Ignored and Updates Refused from Updates Sent.

Docking to the Unified Hierarchical Model

Because I am extending the Unified Hierarchical Model, I must establish that the findings of that model still hold in this work. I believe the Unified Network Model should be able to replicate the following results of the Unified Hierarchical Model:

- That socialization has a nonlinear relationship with organizational performance
- That turnover is an effective mitigator of environmental turbulence
- That the more deliberative capacity an organization brings to bear on hiring replacements for lost members, the less effective the organization

In the Unified Hierarchical Model, I found that some amount of organizational socialization led to superior organizational performance. It seems likely that moderate amounts of informal socialization acted to fill voids in inefficient or problematic organizational structures, while higher levels of socialization limited the ultimate potential of the organization.

The second result to replicate is that turnover is an effective mitigator of environmental turbulence — and, by extension, that organizations that do not experience turnover end up stagnating. This was a finding of the original Mutual Learning Model as well. Because network information spread can also be an effective mitigator of turbulence, I must demonstrate that an organization stagnates without turnover or network information spread.

The third result to replicate is that the more deliberative capacity an organization brings to bear on hiring replacements for lost employees, the less effective the resulting turnover is as mitigation for a changing environment. March argued, but not did simulate, that truly random turnover allowed the organization to bring in a diversity of thought, thereby maintaining its stability and high performance. The more deliberative capacity the organization brought to the question of whom to hire, according to March, the less effective turnover would be as a mechanism to combat organizational decay. The Unified Hierarchical Model was able to demonstrate this result. However, because the Unified Network Model adds a homophily bias related to accepting information from alters, it’s possible that new hires, not yet socialized, may frequently reject new information from their coworkers and have their own updates rejected in turn. For this reason, I include the use of this homophily bias as an additional experimental factor.

Experiment Design

To perform this docking exercise, I use a virtual experiment modeled after the original experiment in G. P. Morgan and Carley (2012), which featured a range of socialization parameters. The virtual experiment is described in Table 25.

Table 25. Virtual Experiment Table for the Docking to the Unified Hierarchical Model

Variable	Values	#
Org Socialization	0.0, 0.10, 0.30, 0.50, 0.70, 0.90	6
Knowledge Responsibility Structure	Hierarchical, None	2
Use Homophily Bias	True, False	2
Turnover and Hiring Strategies	0, 0.02:Random, 0.02:Aggregated Minimum Distance	3
	Combinations	72
	Replications	50
	Total Runs	3600
Outcome Variable	Definition	
Equilibrium Org Code Accuracy	The average performance (ScorePercent) of the organization in the last fifty (50) turns of the simulation, after all organizations have successfully reached equilibrium	
Constants		Setting
Turns		100
Number of Active Organizations		1
Resistance to Change		N/A
Number of Agents		63
Number of Knowledge Bits		50
Coworker Structure		Hierarchy
Knowledge Complement Size		1
Environment Turbulence		0.05
Hiring Committee Size		3
Update Capacity		2

The Unified Hierarchical Model's experiment involved hierarchies, so in this experiment I compared using hierarchies to not using them. Having a "None" setting for knowledge responsibility also allows me to turn off the use of network spread and compare the effectiveness of the organization with and without it.

These results should be robust regardless of whether homophily bias is active, so I include simulations where this variable is and is not in play. I did not use the resistance to change mechanism, because the original experiment focused on a single organization without multiple internal identities.

Because the Unified Hierarchical Model's experiment evaluated the impact of more or less deliberative hiring strategies, I needed to do the same, so I allowed for no turnover (the "0" setting), random (minimally deliberative) hiring with turnover at 2%, and hiring based on aggregated minimum distance (maximally deliberative), also with turnover at 2%.

Results

Using the experiment defined above, I wanted to test the three prior findings of the Unified Hiearchical Model using the Unified Network Model.

I tested the first of these findings — that socialization has a nonlinear relationship with organizational performance — by comparing the relationship between organizational socialization and performance in the Unified Network Model. I aggregated equilibrium performance in two ways:

- An average of the last fifty (50) turns of the simulation, reported as Average Performance
- The maximum performance reported in the last fifty (50) turns of the simulation, reported as Peak Performance

The results are below in Figure 20.

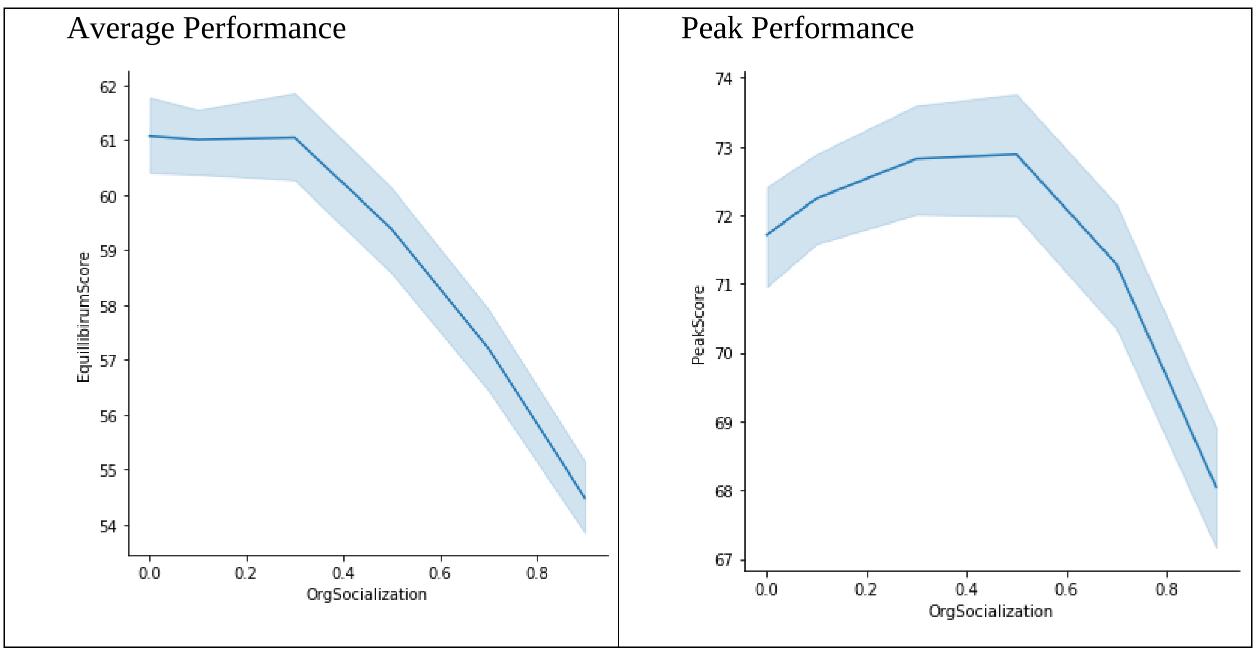


Figure 20. Average and peak equilibrium performance across socialization values

On the left, under Average Performance, organizational performance holds steady as socialization increases until it begins to drop rapidly. On the right, under Peak Performance, increased socialization drives improvement until, after about the halfway point, it leads to decreased performance. I consider this outcome validated by these results.

I examine the second finding of the previous modeling — that turnover is an effective mitigator of turbulence — by comparing the average performance of organizations in four conditions resulting from the interaction of network spread and turnover. These four conditions are:

- No Network Spread with .02 Turnover: Indicated in Figure 21 as "None_0.02," this is where the organization allows new people to enter the organization, but does not assign those individuals to pay attention to changes in the world. The organization will never completely stagnate in this condition.
- No Network Spread and No Turnover: Indicated in Figure 21 as "None_0.0," this is where the organization improves to match the environment simply by learning from its members and then, upon reaching complete agreement, stagnates.
- Network Spread with 0.02 Turnover: Indicated in Figure 21 as "Hierarchical-50_0.02," this is where the organization has people paying attention to the environment and also has turnover. I would expect this to be the highest performing organization, because it is less susceptible to groupthink and less likely to ignore new information from members.
- Network Spread and No Turnover: Indicated in Figure 21 as "Hierarchical-50_0.0," this is where the organization's members pay attention to the environment but the organization does not have turnover. The organization is subject to groupthink and may ignore people with new information that clashes with what it believes, but it should never completely stagnate.

Figure 21 has the results.

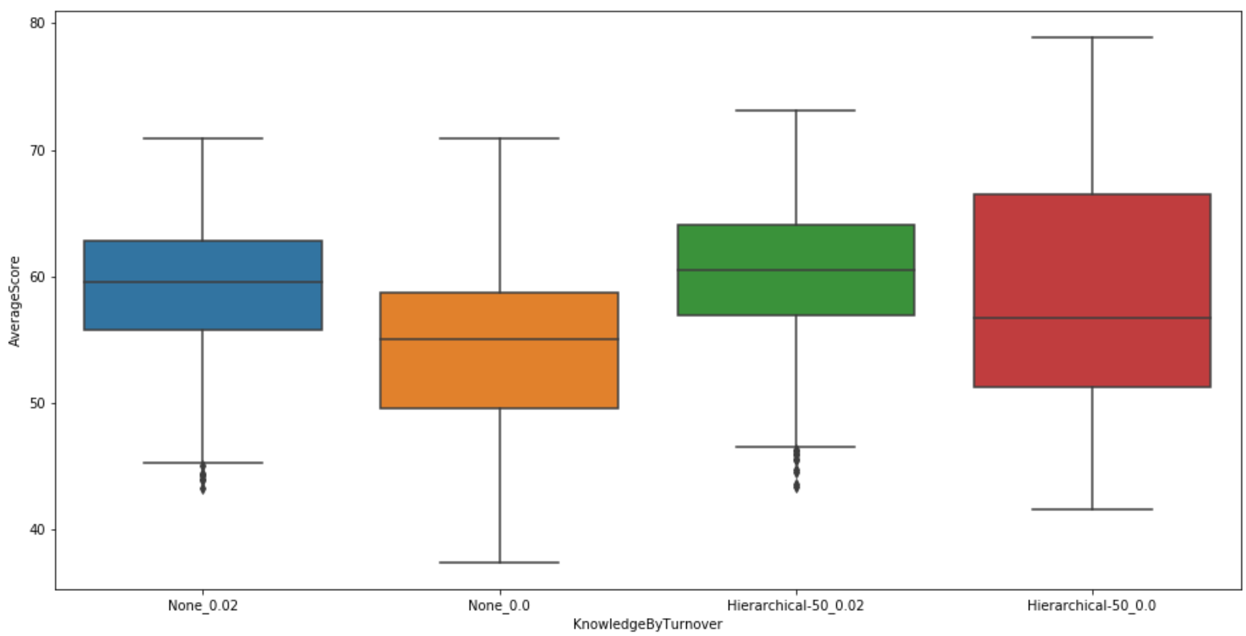


Figure 21. Average performance based on network spread and the use of turnover

I see in Figure 21 the pattern I would expect. Turnover without network spread is an improvement over network spread without turnover, and the former has a narrow band of possible outcomes. The worst performing condition is where the organization neither attends to changes in the environment nor experiences turnover. Network spread and turnover together lead to the highest average outcomes, with similar overall spread to the first condition. Network spread without turnover improves average performance but makes the span of possible outcomes much wider. This wide span of possible outcomes results from the emergent social pressure effect — some simulated organizations suffer from unfortunate placements of highly uniform groups that fall into groupthink. Based on these results, I consider the second prior finding well supported.

The final finding to replicate is that deliberative capacity for hiring harms organizational outcomes. In simulations with turnover, I compared the organization's average equilibrium performance when it used random hiring (SelectionModel "0") to when it hired based on aggregated minimum distance (SelectionModel "2"). Random hiring allows for a person with any views to join the organization. The Aggregated Minimum Distance model, as detailed in prior work, generates 100 possible individuals and selects the one who is most similar to the hiring committee. The membership of the hiring committee is informed by the position to be filled.

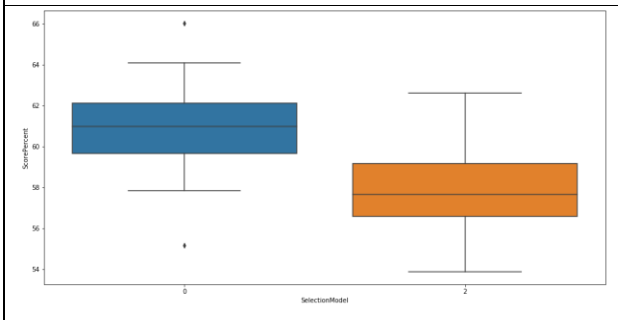
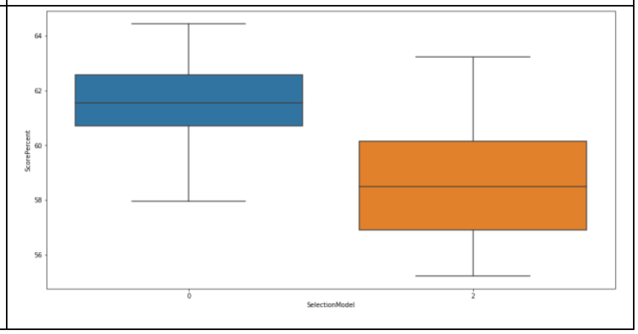
No network information spread	With information network spread
	

Figure 22. Comparison of hiring selection models when network spread is and is not available

As shown in Figure 22, the outcomes are similar whether or not network spread is active: deliberative hiring strategies harm the organization's ability to benefit from turnover, as March (1991) outlined. Network spread does not mitigate this outcome. The Unified Network Model's results again support the previous finding of the Unified Hierarchical Model.

All together, these results indicate that the Unified Network Model can generalize from hierarchies to networks but retain the ability to generate similar findings to those of the Unified Hierarchical Model. As such, the Unified Network Model is relationally docked (Axtell, Axelrod, Epstein, & Cohen, 1996) to the Unified Hierarchical Model (G. P. Morgan & Carley, 2012, 2014).

Factor Experiments

In the process of adapting the Unified Hierarchical Model to focus on the problem of horizontal mergers, I have added several mechanisms to the model. Additional mechanisms are valuable in a simulation only if they allow us to better replicate phenomena we see in the world. To that end, I will explore the necessity of these new mechanisms by evaluating whether they help the simulation better replicate well-understood phenomena related to organizations and organizational mergers. In the field of simulation, these well-understood phenomena are often referred to as "stylized facts," a term that is becoming more common in the social sciences as well (Hirschman, 2016).

Selected Stylized Facts

The stylized facts to replicate are:

- Active subgroup socialization generates more disagreement within the larger organization than no subgroup socialization.
- Disagreement between members of the organization has a complicated/U-shaped interaction with organization performance.
- Larger groups allow for more disagreement between members than smaller groups.
- Social networks with highly localized structures create more disagreement within the organization than networks with less local structure.
- Communications that cross organizational boundaries tend to be ignored more often than communications within the same organization.
- Mergers between organizations fail when there is cultural mismatch between the two merging organizations.

For each stylized fact listed above, I will go into some detail on the reasoning behind the fact as well as relevant research related to it. Finally, I will state how I plan to evaluate the fact in simulation.

Stylized Fact: Subgroup socialization generates more disagreement within the larger organization than no subgroup socialization. Subgroups naturally specialize within a larger organization and become responsible for that in which they specialize in. This, in turn, leads them to value different things than the organization as a whole. Individuals considered high performing based on their performance on tasks relevant to the subunit may not agree with the views of the organization as a whole, which will then cause the subgroup to disagree with the larger organizational consensus. In the merger context, each merging organization will continue to socialize its members to its own beliefs for some time — implicitly if not explicitly. To evaluate this in the Unified Network Model, I compare the amount of disagreement within the entire merged organization in cases where subgroup socialization is active and where it is not.

Stylized Fact: Disagreement between members of the organization has a complicated/U-shaped interaction with organizational performance. At the extremes, this fact is evident and demonstrable in simulation multiple times. When there is no disagreement between members, the organization is unable to learn and will eventually stagnate. When there is too much disagreement between members, the organization has no coherent identity and cannot determine or execute a strategy. However, most organizations exist somewhere in the middle, and scholarship in the area has had contradictory findings. Jehn (1995) found that the benefit of conflict depended on both the structure of the group and on conflict type — relationship conflict (disagreement between people) or task conflict (disagreement on how to do the work). Multiple other taxonomies of conflict have emerged over the years. Amason, Thompson, Hochwarter, and Harrison (1995) identified cognitive conflict (conflict of ideas) as positive and affective conflict (conflict of character) as negative. De Dreu (2008) argues that conflict is only rarely beneficial to the organization, and that effective conflict management is more about keeping conflict from damaging the organization than it is about harnessing conflict for the organization's benefit. In the merger context, conflict is inevitable, so it must be well managed through monetary rewards, careful balancing of power, and cultural negotiation (C. F. Cohen, Birkin, Cohen, Garfield, & Webb, 2006). I will test for this fact in the Unified Network Model by comparing statistical models of the relationship between conflict and performance. Performance may not be dependent on disagreement at all, performance and disagreement may relate linearly, or performance and disagreement may have a higher order relationship. If the best models are those where disagreement is a higher-order term, then the fact will be considered supported.

Stylized Fact: Larger groups allow for more disagreement between members than smaller groups. By allow, I mean that a larger organization is able to tolerate more dissent and still make decisions about the world. Mathematically, this is intuitive. A group of three people can make a group policy only if no more than one member disagrees, but a group of ninety-nine people can do so with far more defections. If we consider groups as boundedly-rational information processors (Cyert & March, 1963), there is only so much capacity for internally monitoring the beliefs of group members — as the group gets larger, it would become the group's only purpose. On the individual level, we are only able to retain a detailed model of the beliefs of a limited number of other people (Joseph, Morgan, Martin, & Carley, 2014). When there are people in a group whom we don't really know, we generally assume they are more like us than they actually are — a form of ethnocentrism (Hartshorn, Kaznatcheev, & Shultz, 2013). Therefore, when groups get large enough, individuals begin to assume that the group believes the same things they do, and the group itself is more able to tolerate disagreement. In the merger context, the number of people with whom a specific individual may need to interact has drastically increased, and they may be asked suddenly to begin working with others they do not know. I assess this fact in the Unified Network Model by comparing how disagreement affects performance levels across various group sizes.

Stylized Fact: Social networks with highly localized structures create more disagreement within the organization than networks with less local structure. In the network context, a highly localized structure means that many alters are shared between any two people. This is common in many real-world contexts, such as a primary school clique, an agile software development team, or a terrorist cell. These groups often share information internally. As a result, the individuals in the group tend to become more alike over time (Carley, 1990). As they become more similar, they will tend to be less interested in information from outsiders. Over time, therefore, the beliefs of a small group with highly localized structure will diverge organically from the beliefs of the rest of the organization, creating more disagreement within the organization. Furthermore, contained communities of practice often develop knowledge that can only be applied usefully within the context of those communities. That makes it difficult to share information outside of these communities even when desired (Brown & Duguid, 2001). I evaluate the impact of localized structures through three different network topologies: a cellular network (highly localized, long distances), a cellular small-world network (highly localized, short distances), and a random network with the same number of ties (not localized, short distances). I use the Kruskal-Wallis test, a non-parametric comparison of means, to evaluate whether the differences between these networks are significant in the Unified Network Model.

Stylized Fact: Communications that cross organizational boundaries tend to be ignored more often than communications within the same organization. Humans naturally classify others as in-group or out-group based on a variety of factors (Diehl, 1990), and they treat these groups differently. While homophily bias is a dyadic phenomenon between individuals, agents also pay attention to group affiliations and use them to make decisions. For example, existing teams that are asked to accept input from perceived strangers will frequently resist or ignore the information, even if it is objectively useful (Kane, Argote, & Levine, 2005). When internal groups are perceived to compete with each other for resources, even cross-group coordination ties fail to support knowledge transfer (Tsai, 2002). I test this in the Unified Network Model by using the Kruskal-Wallis test to compare the number of messages ignored between people who have cross-group ties and people who do not.

Stylized Fact: Mergers between organizations fail when there is cultural mismatch between the two merging organizations. Much has been written about why mergers fail, but many scholars agree that cultural mismatch is an important factor — even though the evidence has at times been ambiguous. Epstein (2005), who focused on domestic mergers, included cultural mismatch under "due diligence." For international mergers, concerns about the cultures of the merging firms are often subordinate to concerns about national cultural differences (Li & Guisinger, 1991; Xie, Reddy, & Liang, 2017). While national culture differences are indeed an important factor, they are distinct from the organizational cultures of interest here. In a large meta-analytic study, Stahl and Voigt (2008) found that organizational cultural differences had a small but statistically significant effect on the ultimate success of the merger. I evaluate this fact in my simulation by assessing the differences between the organizational codes of each suborganization, then comparing that with the mid-period performance of the resulting merged organization. After accounting for disagreement, I would expect a negative correlation between differences in each suborganization's organizational code and merged organizational performance.

Experiment Design

To evaluate whether my added mechanisms aid me in better replicating these stylized facts, I use a layered series of virtual experiments where each mechanism or factor is introduced individually. These experiments are:

- Subgroup Socialization: Subgroups socialize members to their beliefs. Each subgroup has its own view on what "high performance" means and uses that to develop its own organizational code. This code is then used to socialize members. Agent network ties are not enabled in this experiment.
- Communication Structure: Three different network topologies are used to construct the social networks of agents. These topologies are ER-Random (Erdős & Rényi, 1960), Small-World (Milgram, 1967; Watts & Strogatz, 1998), and Cellular (Fang et al., 2010; Frantz & Carley, 2005). Knowledge assignments are distributed randomly. Subgroup Socialization is not active in this or following experiments except Socialization and Spread.

- Grouped Knowledge: Similar to the Communication Structure experiment, except knowledge is grouped, where each suborganization has a set of bits they care about and agents are assigned bits at random within that group.
- Homophily Bias: Using all three defined network topologies, and assigning knowledge using both the grouped and random settings, the only difference between this and the previous two experiments is that the homophily bias mechanism is active.
- Resistance to Change: Using all three defined network topologies, and assigning knowledge using both the grouped and random settings, the only difference between this and the previous two experiments is that the resistance to change mechanism is active.
- Socialization and Spread: Both subgroup socialization and all network spread mechanisms are active. This is expected to be the most emulative simulation. I vary communication structures and knowledge assignment.

These experiments are described in Table 26.

Table 26. Six virtual experiments used for the six stylized facts

		Simulation Variable Settings					
Experiment	Total Combos	Merging Org Relative Size	Sub-Group Socialization Active	Communication Structures	Knowledge Assignment	Homophily Bias Active	Resistance to Change Active
Socialization Exp	8	50/50, 60/40, 75/25, 90/10	Yes (0.1), No	None	None	NA	NA
Communication Structure Exp	12	50/50, 60/40, 75/25, 90/10	No	Random, Small-World, Cellular	Random	No	No
Grouped Knowledge Exp	12	50/50, 60/40, 75/25, 90/10	No	Random, Small-World, Cellular	Grouped	No	No
Homophily Bias Exp	24	50/50, 60/40, 75/25, 90/10	No	Random, Small-World, Cellular	Random, Grouped	Yes	No
Resistance to Change Exp	24	50/50, 60/40, 75/25, 90/10	No	Random, Small-World, Cellular	Random, Grouped	No	Yes
Socialization and Spread Exp	24	50/50, 60/40, 75/25, 90/10	Yes (0.1)	Random, Small-World, Cellular	Random, Grouped	Yes	Yes

All other settings were held constant across simulation runs. Other value settings are documented in

Table 27.

Results

As discussed in the two preceding subsections, I used statistical tests to evaluate whether each stylized fact was validated when the simulation used the mechanisms described. Overall, the simulation became more emulative of these stylized facts as more mechanisms were added.

Because there are two distinct mechanisms by which information spreads through the simulated organization — socialization and network-based information spread — an interesting dynamic occurs where they interact, as they do in the final experiment, “Socialization and Spread.” For five of these six facts, the interaction between these two mechanisms accentuated the finding. Please see Figure 23 for an overview of the results. This graphical table identifies how well a given experiment supported a given stylized fact, as determined by statistical analysis. An empty (white) circle indicates the fact was not supported. A semicircle indicates a partially supported fact; this could indicate that the trend is moving in the correct direction, or that different statistical tests give different results. A filled-in circle indicates that the statistical evaluation fully supported the stylized fact.

Fact	Socialization	Communication Structure	Grouped Knowledge	Homophily Bias	Resistance to Change	Socialization + Spread
Sub-Group Socializations generates more disagreement between elements of the organization	●	NA	NA	NA	NA	●
The conflict of ideas has a complicated/U-shaped interaction with organizational performance	●	●	●	●	●	●
Larger groups allow for more disagreement between members	◐	●	●	●	●	◐
Social Networks with highly local structures create more disagreement within the organization	NA	●	●	●	●	●
Communications that cross organizational boundaries tend to be ignored more than communications within the same organization	NA	○	○	◐	●	●
Mergers between organizations fail when there is a cultural mismatch	●	◐	●	○	●	●

Figure 23. Stylized facts and experiment outcomes

I can see the power of the multilevel socialization mechanism, because it is emulative of most of the facts available. For simulations where I used only network spread, I need to add in more of the supplementary mechanisms in order to validate the selected stylized facts. When I combine both mechanisms, most facts are further accentuated by the interaction of the two structures, but the multilevel socialization mechanism reduces the amount of disagreement within the group.

Stylized Fact: Subgroup socialization generates more disagreement within the larger organization than no subgroup socialization. The effect is stronger when only multilevel socialization is used, as would be expected. However, subgroups are still able to maintain their distinction from each other even when individuals within these organizations interact routinely.

Stylized Fact: Disagreement between members of the organization has a complicated/U-shaped interaction with organizational performance. To evaluate this fact, I compared current internal disagreement to current performance using statistical models. In every case, the statistical model that used only the linear term for disagreement's relationship to performance was inferior in both Aikake Information Criterion (AIC) and Bayesian Information Criterion (BIC) to the statistical model with both a linear and a squared term. Often, the linear term alone did not usefully explain variation, while adding the squared term made both terms useful to the model.

Stylized Fact: Larger groups allow for more disagreement between members than smaller groups. In models that use only network spread mechanisms, the results fully support this stylized fact. Overall, internal disagreement increases as group size increases, while performance remains stable or slightly improves. When I add multilevel socialization, the relationship becomes much flatter. Only the smallest organizations tolerate significantly less disagreement than larger ones do. This perhaps suggests that the implemented multilevel socialization mechanism is too powerful. I will return to this idea in the discussion.

Stylized Fact: Social networks with highly localized structures create more disagreement within the organization than networks with less local structure. I evaluated this based on the difference in organizational code between the two merging organizations. Organizations that rely on cellular or semi-cellular structures are better able to maintain the differences in sub-organizational identities than organizations with other structures, and thus generate more disagreement. This stylized fact is validated.

Stylized Fact: Communications that cross organizational boundaries tend to be ignored more often than communications within the same organization. The implemented network spread mechanisms have a capacity constraint, because these agents are boundedly rational (Simon, 1991), as well as mechanisms to allow the agent to simply ignore another agent based on individual (homophily) or group (resistance to change) differentiation. Without homophily or resistance to change mechanisms activated, the agents tended be extremely receptive to cross-organization messages, because those messages tended to come when their local group was much less active. In other words, many messages from others within the group were ignored due to capacity limitations, but when messages from outside the group arrived, there was rarely a capacity constraint on taking in these new messages, since those agents were generally still available to receive new messages. When agents start to make choices about what information to accept, however, the outcomes change. Homophily drives the up the "ignored" rate for out-group messages while driving down the overall number of messages ignored for capacity reasons. The resistance to change mechanism, which is based on group identities, greatly increases the number of messages from out-group senders that are ignored compared to in-group messages. As we add these individual mechanisms, the simulation becomes able to replicate this fact.

Stylized Fact: Mergers between organizations fail when there is cultural mismatch between the two merging organizations. To determine cultural differences, I compare the weight vectors each merging organization uses to define high performers. I then compare the difference between weight vectors to the performance of the merged organization, after accounting for internal group differences with both a linear and squared term. For most of these experiments, there is a negative interaction between cultural differences and organizational performance. The homophily experiment stands out as unusual here, in that the relationship is not found. It is possible that the homophily mechanism, which works at the individual level, reduces the impact of group-level differences in such a way that these cultural mismatches are lost in the noise.

Now that I have reviewed the simulation's validation of stylized facts, I can instantiate the simulation using the data from the MergedCo case study that I analyzed in Chapters 1 and 2.

MergedCo Case Study

In the preceding sections, I showed that the new Unified Network Model could successfully replicate the demonstrated effects of its predecessor, the Unified Hierarchical Model. Then, I demonstrated that the full simulation was able to replicate six stylized facts considered common and important to mergers.

In this section, I use the best available data from a specific real-world merger to determine whether the model is able to provide useful insight into a merger in progress. In this case, I use social network and language data pulled from a large set of emails about 9 months after the merger announcement, when integration is already well underway. The reciprocity network developed in Chapter 1 of this dissertation, which shows who responds to whose messages within 24 hours, is used to inform the social network of the Unified Network Model. The token network developed in Chapter 2 of this dissertation, which shows who uses which language tokens, is used to inform the knowledge network. Survey data could also inform the model's social and knowledge networks if so desired. Figure 24 depicts the structure of MergedCo.

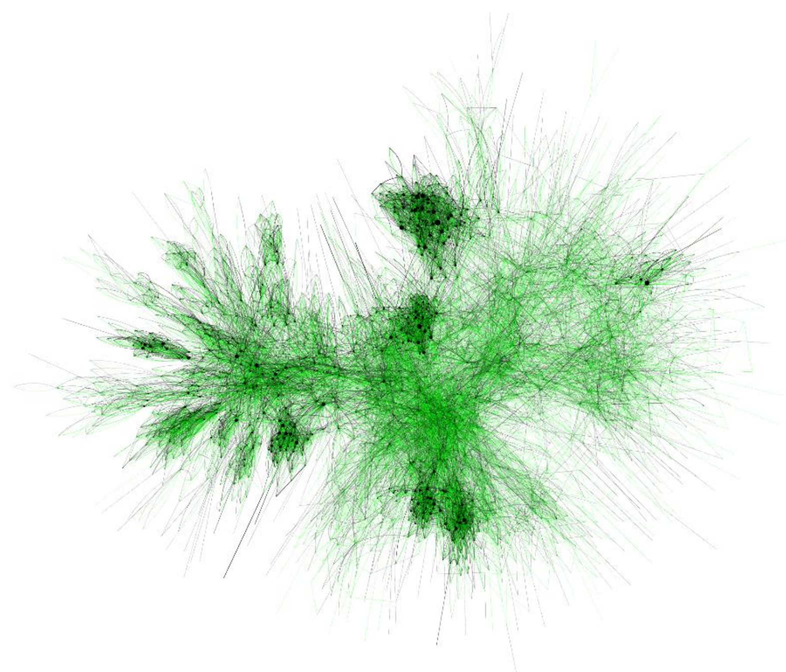


Figure 24. Social structure of MergedCo as revealed by an example run. Lighter (green) links indicate successful messaging and dark (black) links indicate unsuccessful messaging

In best practice, a model of this kind would be instantiated before the merger, based on planned unit integrations, and then updated continuously as the actual integration proceeded. The strength of socialization and its frequency would be assessed based on socialization expectations for new hires. For this case study, I assumed default values based on the size of the individual organizations.

Other settings were set based on defaults from the stylized fact validation experiments. I document these settings in

Table 27. Appendix D covers all these variables, their meanings, and their expected ranges.

Table 27. Constants set for the case study runs, with implications

Constant	Value	Meaning/Implication
MAX_TURNS	1000	The simulation will run for 1000 turns
REALITY_SHIFT_RATE	0.02	Every bit in the reality vector may flip at a 2% chance each turn
TURNOVER_RATE	0.01	Every individual has a 1% chance of leaving the organization each turn
GRACE_PERIOD	1001	How many turns an employee is given before they may be "fired for cause." In short, the "fired for cause" mechanism is turned off — this is a setting inherited from the Unified Hierarchical Model and not explored in this research
USE_HIRING_COMMITTEE	False	Replace people who leave with randomly selected candidates, rather than using a hiring committee
USE_HOMOPHILY_BIAS INFORMATION_SPREAD	True	Should the homophily bias mechanism affect likelihood of accepting a message from someone?
USE_RESISTANCE_BIAS INFORMATION_SPREAD	True	Should the resistance to change mechanism (based on group affiliation) affect likelihood of accepting a message from someone?
UPDATE_CAPACITY	2	How many messages can an individual accept in a turn? This is a strong mechanism to implement individual bounded rationality.
EXTRA_MESSAGE SEND_RATE	0.8	The likelihood of agents sending further messages, which decays based on this parameter. Sending the first message is 0.8^0, the second 0.8^1, the third 0.8^2, etc.
PERCEPTION_ACCURACY	0.9	How accurate is the perception of others at the individual level? This means that for each bit an alter has, the agent gets right about 90% of them.
STAFF_PERCEPTION ACCURACY	0.9	How accurate is the organization about understanding its staff? This means that for each bit an agent has, the organization gets right about 90% of them.
CODE_LEARNING_RATE	0.5	How risk-tolerant is the organization toward learning from its members? Is the organization risk-averse, wanting to learn only when there is a clear consensus from high performers? Or is the organization risk-tolerant, willing to take a position when only a slim margin of high performers think it is correct? 0.5 is only

		moderately risk-tolerant. High values indicate more risk tolerance.
SOCIALIZATION_RATE	0.05	How effective is an organization's socialization? This setting indicates that any individual bit of an agent has a 5% chance of changing to what the socializing organization believes on each turn when socialization takes place. Note every organization can have different values for this, but I do not change this value per organization in the simulation runs for this paper due to a lack of data.
USE_SUBGROUP SOCIALIZATION	True	Should each individual sub-organization socialize its members? This is a default setting for the run and can be set per organization.
KNOWLEDGE_COMPLEMENT SIZE	1	This is the "chunkiness" setting for how interdependent knowledge bits are to informing tasks. A value of 1 means each knowledge bit is independent of every other. Because I used a static knowledge network generated based on differentiation, I didn't want to imply an ordering of tokens I did not have evidence for in the data.
WARMUP_TURNS	50	How many turns are run with the sub-organizations interacting separately from each other before the full "merged" simulation begins?
WARMUP_SOCIALIZATION RATE	0.05	How effective is each organization's socialization? I keep this the same as during the main run.

Turnover and Messaging

One frequent concern is: Who will leave the organization? To investigate this, I identify individuals who are present in the Time-1 data used to instantiate the simulation but not present in the data available from a later time period, Time-2. As seen in Figure 25, the simulation presents a clear signal of who will not be present in Time-2: those who do not send messages.

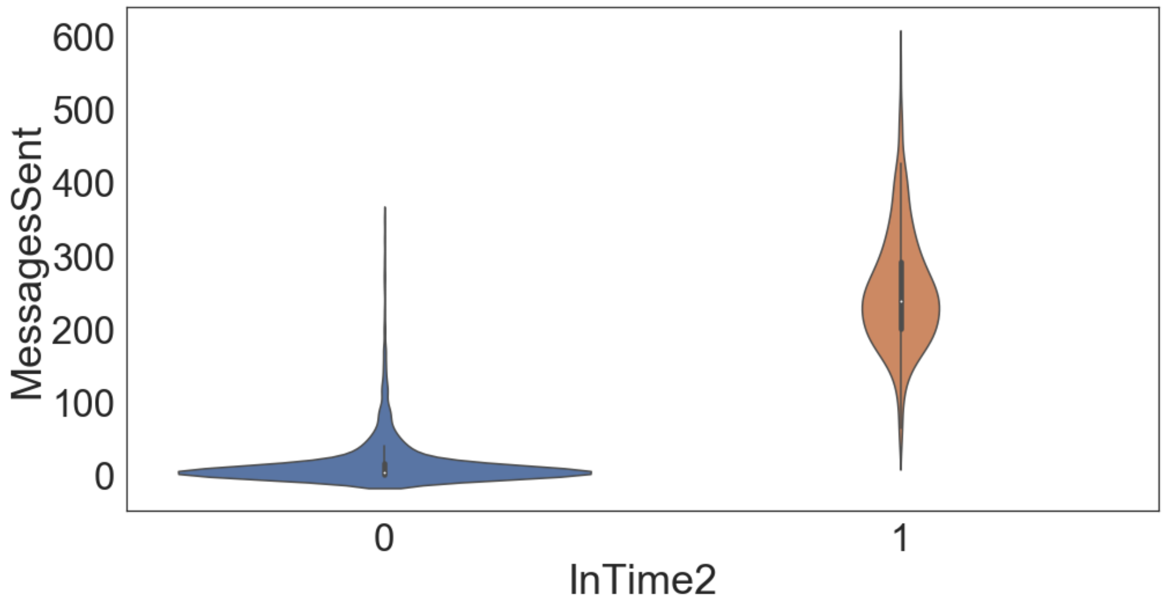


Figure 25. Individuals who do not send messages are very likely to be absent in Time-2 data, $p < 0.001$

These individuals may be routinely refusing messages from others, or they may be disconnected on the periphery. This result may also simply be an artifact of the data, but I believe it reflects the emergent outcomes of the simulation: Individuals without many reciprocal ties in Time-1 are left behind even as their respective organizations continue to socialize them.

Empirical research has connected lack of engagement with various negative outcomes; low job satisfaction and low job involvement can both predict absenteeism (Wegge, Schmidt, Parkes, & Dick, 2007), and job satisfaction has, separately, been connected to reported engagement levels at work (Sudibjo & Sutarji, 2020). Engagement itself has been found to correlate with turnover intentions, even after accounting for other traditional predictors of turnover (Alarcon & Edwards, 2011). If this finding is validated in future work, this is another important stylized fact the Unified Network Model is able to emulate.

Performance Over Time Compared to Stock Close Prices

In addition to using Time-1 data to predict departure by Time-2, I also wanted to see if the model outputs could be usefully correlated with organizational performance as indicated by historical stock close prices. To collect stock close prices, I used Alpha Vantage's API for the organization in question (AlphaVantage, 2021).

I use Time-1 data to instantiate the model. Specifically, I use the Time-1 reciprocity network from Chapter 1 as the agent-by-agent network, and I use the top 3 tokens for each actor from Chapter 2 to instantiate the agent-by-knowledge network. To consider these results appropriately, it is important to note that this data is from after the onset of the merger. Time-1 is about 3 months after the merger was completed, and the merger first became public knowledge 5 months before that.

To compare stock close prices against simulation performance, I needed a method for converting turns into time periods, and I needed a method for determining a starting point. I evaluated multiple options for converting turns to time periods. The best fit occurred when a turn was equal to two work-hours at the organization; in other words, 4 turns make up a workday. To set the starting point, I started at the midpoint of the Time-1 data, but used a sliding window to move the comparison forward or back.

With these two methods in place, I could compare the average per-turn performance predicted by the model to the actual stock prices for those periods (since 4 turns make up a day). The best models started roughly when the merger was announced and have an r-squared of around 0.8. Figure 26 shows the example fitted over-time correlation between the model output and the adjusted close price of the stock. I normalized both time and performance to simplify the visual and avoid identifying the company in question.

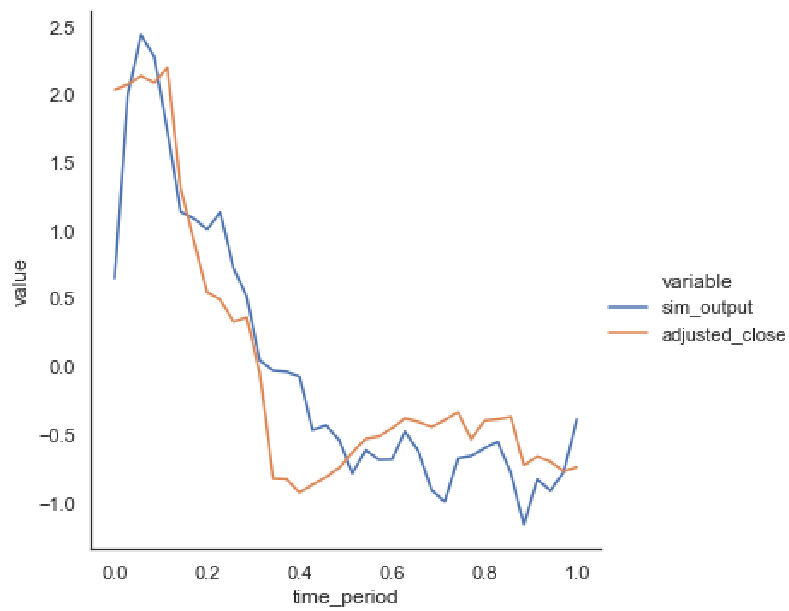


Figure 26. The normalized time trace of the simulation versus the adjusted close of the organization's stock price

The model predicts that performance will rise, then fall sharply as the differences in the two organizations' corporate cultures become evident. This model fit — which, again, uses data from farther in the future than the merger's onset — illustrates the potential of multilevel COT simulation to help organizations plan horizontal mergers and other large-scale changes and to predict problems that may occur.

Of course, while stock prices are relatively convenient to access, they are at the mercy of larger market forces. As such, future work should consider pulling quarterly earnings reports and using those as a lagging indicator to compare to simulation output. When working directly with an organization, weekly or monthly sales figures could also be used and provide more points of comparison.

Discussion and Conclusions

In this chapter, I introduced a multilevel agent-based simulation that extends prior work and can be used to model complex organizations going through a merger process. I call this simulation the Unified Network Model. I described the lineage of this model, which inherits traits from GARCORG (Carley, 1986b), Construct (Carley, 1990), March's Mutual Learning Model (March, 1991), and the Participation Model (J. H. Morgan et al., 2010), and is a generalization of the Unified Hierarchical Model (G. P. Morgan & Carley, 2012). Unlike any of its predecessors, it is an active multilevel model where organizational actors and individual agents both take actions, and the consequences of these actions combine to create the emergent outcomes of the simulation.

After describing its operation and further extensions from the Unified Hierarchical Model, I established the utility of the model through three distinct strategies. First, I docked (Axtell et al., 1996) the Unified Network Model against key outcomes of the Unified Hierarchical Model and was able to establish relational equivalence. Second, I identified six stylized facts important to organizational performance and salient to the specific issues of horizontal mergers and found that the model was able to emulate these stylized facts with all of its mechanisms engaged. Third, I instantiated the model using Time-1 empirical data from my other chapters, both of which focus on the horizontal merger of a large multinational company. The model is able to provide insight farther into the future than the case study by predicting who will not be present in Time-2 data, and it can generate suggestive correlations with performance at the onset of the merger. (On this latter element, I do not pretend that this evidence is persuasive, merely interesting.)

There is more work to be done to address weaknesses in this model, to make it more capable of modeling more contexts, and to make it more usable by practitioners.

Firstly, the socialization mechanism, which was inherited directly from the Mutual Learning Model, is too strong when it is used in the Unified Network Model, which is designed to better emulate real organizations. Individuals tend to resist organizational socialization (Feldman, 1981), and there are many studies on what makes socialization more or less effective — see Bauer, Erdogan, Bodner, Truxillo, and Tucker (2007) for a review. A socialization mechanism that is sensitive to trait-level or attitudinal measures, which are often gathered by surveys, may be helpful; a mechanism with an informed view of when socialization may fail, so that such cases can be implemented computationally, would be an excellent extension.

Market change is only cursorily modeled here, with each bit having an independent probability of changing based on a uniform distribution. In the real world, stable markets may exist for years only to abruptly crumble with the onset of an innovation. More models of change could be implemented, as was done in GARCORG (Carley, 1986), and these findings could be tested for robustness against alternative models of environmental change. More practically, stylized organizational forms could also be tested for agility by allowing the modeled organizations to reach stable performance, then abruptly changing the environment and seeing how long it takes them to adapt. This would be an interesting extension of the model for those interested in organizational agility.

I have an active multilevel model in that organizations act on their members and their members interact with each other. However, the organizations themselves do not act towards other organizations. This is appropriate in this context, where the legacy organizations primarily live on in the hearts and minds of their former members, but it may not be useful for modeling competition or other more explicit forms of inter-organizational conflict, such as competing for funding or racing to bring innovations to the market. I feel that such extensions would necessarily be domain-specific.

The homophily mechanism is used to determine whether or not an individual accepts an update from another individual based on similarity. The homophily mechanism could also be used to determine who is most likely to be messaged, as it is used in Construct (Carley, 1990). This could create interesting inter-agent dynamics when agents disagree about their perceived similarity because of differences in their local contexts.

Furthermore, the current operation of the model only allows messages to be passed to others with whom an agent has a network tie. This is an acceptable limitation when modeling stylized organizations, or very specific organizations at a specific points in time, but in reality, such ties tend to change. A tie creation and tie removal mechanism based on available literature would be a welcome extension of this model, particularly for modeling organizations over longer periods of time or at critical junctures.

The resistance to change mechanism implemented here doesn't take into account individual agents' attachments to their group identities. One possible extension would be to identify the distance between the group's conception of reality and the individual agent's. The more similar these conceptions are, the more likely the message recipient is to use group identity as the basis for rejecting messages from those "out-group" to the recipient when the recipient's group is doing poorly.

For the practitioner, it would be helpful to identify methods for specifying the various parameters of the models. Instruments for characterizing the environment, such as those included in Burton, Obel, and DeSanctis (2011), could be used to better understand the context in which the organization operates. Other instruments could be applied to characterize the strength of socialization and the interdependence of tasks at the organization.

Finally, the current model is implemented in Java and is instantiated through custom methods. The ability to specify the model through an instantiation script or, more ambitiously, through a GUI would make the model available for non-technical individuals to use.

In this chapter, I have delineated a new multilevel COT (G. P. Morgan & Carley, 2015) model of organizational change and used that model to simulate the horizontal merger of a large multinational company. The model is able to replicate prior findings of similar models and to emulate stylized facts relevant to organizational performance and mergers. Such a model could be useful to organizational theorists interested in the performance of stylized forms as well as to practitioners interested in gathering insight about potential organizational futures.

References

Adolph, G., Buchanan, I., Horney, J., Jackson, B., Jones, J., Kihlstedt, T., . . . Quarls, H. (2001). *Merger Integration: Delivering on the Promise*. Retrieved from

Alarcon, G. M., & Edwards, J. M. (2011). The relationship of engagement, job satisfaction and turnover intentions. *Stress and Health, 27*(3), e294-e298. doi:10.1002/smi.1365

AlphaVantage. (2021). Time Series Stock API (API). Retrieved from alphavantage.co. Retrieved 2021, from AlphaVantage alphavantage.co

Amason, A. C., Thompson, K. R., Hochwarter, W. A., & Harrison, A. W. (1995). Conflict: An important dimension in successful management teams. *Organizational Dynamics, 24*(2), 20-35. doi:10.1016/0090-2616(95)90069-1

Argote, L., & Hora, M. (2017). Organizational Learning and Management of Technology. *Production and Operations Management, 26*(4), 579-590. doi:10.1111/poms.12667

Axtell, R., Axelrod, R., Epstein, J. M., & Cohen, M. D. (1996). Aligning simulation models: A case study and results. *Computational Mathematical Organization Theory, 1*(2), 123-141.

Bauer, T. N., Erdogan, B., Bodner, T., Truxillo, D. M., & Tucker, J. S. (2007). Newcomer adjustment during organizational socialization: a meta-analytic review of antecedents, outcomes, and methods. *Journal of Applied Psychology, 92*(3), 707-721.

Bechky, B. A. (2011). Making Organizational Theory Work: Institutions, Occupations, and Negotiated Orders. *Organization Science, 22*(5), 1157-1167. doi:10.1287/orsc.1100.0603

Brown, J. S., & Duguid, P. (2001). Knowledge and Organization: A Social-Practice Perspective. *Organization Science, 12*(2), 198-213. doi:10.1287/orsc.12.2.198.10116

Burton, R. M., Obel, B., & DeSanctis, G. (2011). *Organizational design: A step-by-step approach*: Cambridge University Press.

Carley, K. M. (1986a). Efficiency in a garbage can: Implications for crisis management. In J. G. March & R. Weissinger-Baylon (Eds.), *Ambiguity and Command: Organizational Perspectives on Military Decision-Making* (pp. 195-231): Pitman Publishing Inc.

Carley, K. M. (1986b). Measuring efficiency in a garbage can hierarchy. In J. G. March & R. Weissinger-Baylon (Eds.), *Ambiguity and Command: Organizational Perspectives on Military Decision-Making* (pp. 165-194): Pitman Publishing Inc.

Carley, K. M. (1990). Group stability: A socio-cognitive approach. *Advances in Group Processes, 7*(1), 44.

Cartwright, S., & Cooper, C. L. (1994). The Human Effects of Mergers and Acquisitions: INTRODUCTION. *Journal of Organizational Behavior (1986-1998)*, 47. Retrieved from https://www.proquest.com/scholarly-journals/human-effects-mergers-acquisitions/docview/228923220/se-2?accountid=9902

https://cmu.primo.exlibrisgroup.com/openurl/01CMU_INST/01CMU_INST:01CMU?genre=article&atitle=The+Human+Effects+of+Mergers+and+Acquisitions%3A+INTRODUCTION&author=Cartwright%2C+Sue%3BCooper%2C+Cary+L&volume=&issue=&spage=47&date=1994&rft.btitle=&rft.jtitle=Journal+of+Organizational+Behavior+%281986-1998%29&issn=0894-3796&isbn=&sid=ProQ%3Aabiglobal_

Cohen, C. F., Birkin, S. J., Cohen, M. E., Garfield, M. J., & Webb, H. W. (2006). Managing conflict during an organizational acquisition. *Conflict Resolution Quarterly, 23*(3), 317-331. doi:10.1002/crq.140

Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A Garbage Can Model of Organizational Choice. *Administrative Science Quarterly*, 1-25.

Cyert, R. M., & March, J. G. (1963). *A Behavioral Theory of the Firm* (Vol. 2). Englewood Cliffs, NJ: Prentice-Hall.

De Dreu, C. K. W. (2008). The virtue and vice of workplace conflict: food for (pessimistic) thought. *Journal of Organizational Behavior, 29*(1), 5-18. doi:10.1002/job.474

Diehl, M. (1990). The minimal group paradigm: Theoretical explanations and empirical findings. *European review of social psychology, 1*(1), 263-292.

Drucker, P. F. (1988). The coming of the new organization. *Harvard Business Review,* 3-12.

Epstein, M. J. (2005). The determinants and evaluation of merger success. *Business Horizons, 48*(1), 37-46. doi:10.1016/j.bushor.2004.10.001

Erdős, P., & Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci, 5*(1), 17-60.

Fang, C., Lee, J., & Schilling, M. A. (2010). Balancing Exploration and Exploitation Through Structural Design: The Isolation of Subgroups and Organizational Learning. *Organization Science, 21*(3), 625-642. doi:10.1287/orsc.1090.0468

Feldman, D. C. (1981). The Multiple Socialization Of Organization Members. *Academy of Management Review, 6*(2), 309-318. doi:10.5465/amr.1981.4287859

Frantz, T. L., & Carley, K. M. (2005). A Formal Characterization of Cellular Networks. *SSRN Electronic Journal*. doi:10.2139/ssrn.2726808

Fu, F., Nowak, M. A., Christakis, N. A., & Fowler, J. H. (2012). The Evolution of Homophily. *Scientific Reports, 2*(1). doi:10.1038/srep00845

Giessner, S. R., Viki, G. T., Otten, S., Terry, D. J., & Täuber, S. (2006). The Challenge of Merging: Merger Patterns, Premerger Status, and Merger Support. *Personality and Social Psychology Bulletin, 32*(3), 339-352. doi:10.1177/0146167205282151

Hartshorn, M., Kaznatcheev, A., & Shultz, T. (2013). The evolutionary dominance of ethnocentric cooperation. *Journal of Artificial Societies and Social Simulation, 16*(3), 7.

Hirschman, D. (2016). Stylized Facts in the Social Sciences. *Sociological Science, 3*, 604-626. doi:10.15195/v3.a26

Jehn, K. A. (1995). A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative Science Quarterly*, 256-282.

Joseph, K., Morgan, G. P., Martin, M. K., & Carley, K. M. (2014). On the Coevolution of Stereotype, Culture, and Social Relationships. *Social Science Computer Review, 32*(3), 295-311. doi:10.1177/0894439313511388

Kane, A. A., Argote, L., & Levine, J. M. (2005). Knowledge transfer between groups via personnel rotation: Effects of social identity and knowledge quality. *Organizational Behavior and Human Decision Processes, 96*(1), 56-71. doi:10.1016/j.obhdp.2004.09.002

Kansal, S., & Chandani, A. (2014). Effective Management of Change During Merger and Acquisition. *Procedia Economics and Finance, 11*, 208-217. doi:10.1016/s2212-5671(14)00189-0

Lanham, M. J., Morgan, G. P., & Carley, K. M. (2011). *Data-driven diffusion modeling to examine deterrence.* Paper presented at the Network Science Workshop (NSW), 2011 IEEE.

Li, J., & Guisinger, S. (1991). Comparative Business Failures of Foreign-Controlled Firms in the United States. *Journal of International Business Studies, 22*(2), 209-224. doi:10.1057/palgrave.jibs.8490300

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*(1), 71-87.

McFadden, D. (1980). Econometric models for probabilistic choice among products. *Journal of Business, 53*(3), S13-S29. doi:10.2307/2352205

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology, 27*(1), 415-444. doi:10.1146/annurev.soc.27.1.415

Milgram, S. (1967). The small world problem. *Psychology today, 2*(1), 60-67.

Morgan, G. P., & Carley, K. M. (2012). Modeling formal and informal ties within an organization: a multiple model integration. In *The Garbage Can Model of Organizational Choice: Looking Forward at Forty* (pp. 253-292): Emerald Group Publishing Limited.

Morgan, G. P., & Carley, K. M. (2014). Comparing hiring strategies in a committee with similarity biases. *Computational Mathematical Organization Theory, 20*(1), 1-19.

Morgan, G. P., & Carley, K. M. (2015). Computational Organizational Theory. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (Vol. 4, pp. 485-492). Oxford: Elsevier.

Morgan, J. H., Morgan, G. P., & Ritter, F. E. (2010). A preliminary model of participation for small groups. *Computational Mathematical Organization Theory, 16*(3), 246-270.

Padgett, J. F. (1980). Managing garbage can hierarchies. *Administrative Science Quarterly*, 583-604.

Porter, M. E. (1987). From Competitive Advantage to Corporate Strategy. *Harvard Business Review*(May-June), 43-59.

Sidle, S. D. (2006). Resisting the urge to merge. *Academy of Management Perspectives, 20*(3), 115-118.

Simon, H. A. (1991). Bounded rationality and organizational learning. *Organization Science, 2*(1), 125-134.

Stahl, G. K., & Voigt, A. (2008). Do Cultural Differences Matter in Mergers and Acquisitions? A Tentative Model and Examination. *Organization Science, 19*(1), 160-176. doi:10.1287/orsc.1070.0270

Sudibjo, N., & Sutarji, T. (2020). The roles of job satisfaction, well-being, and emotional intelligence in enhancing teachers' employee engagements. *Management Science Letters, 10*, 2477-2482.

Tsai, W. (2002). Social structure of "coopetition" within a multiunit organization: Coordination, competition, and intraorganizational knowledge sharing. *Organization Science, 13*(2), 179-190.

Vieru, D., & Rivard, S. (2014). Organizational identity challenges in a post-merger context: A case study of an information system implementation project. *International Journal of Information Management, 34*(3), 381-386. doi:10.1016/j.ijinfomgt.2014.02.001

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440-442. doi:10.1038/30918

Wegge, J., Schmidt, K.-H., Parkes, C., & Dick, R. (2007). Taking a sickie: Job satisfaction and job involvement as interactive predictors of absenteeism in a public organization. *Journal of Occupational and Organizational Psychology, 80*(1), 77-89. doi:10.1348/096317906x99371

Xie, E., Reddy, K. S., & Liang, J. (2017). Country-specific determinants of cross-border mergers and acquisitions: A comprehensive review and future research directions. *Journal of World Business*, *52*(2), 127-183. doi:10.1016/j.jwb.2016.12.005

Conclusion

In this dissertation, I introduced the concept of the horizontal merger as the diffusion (Rogers, 2010) of a new organizational identity. Chapter 1 developed an approach for identifying the diffusion network by taking advantage of the nature of email to identify alters with whom individual senders have strong time-bounded reciprocal communication links. Chapter 2 focused on identifying change in individuals through computational textual analysis. Chapter 3 embodied the overarching theory in a multilevel agent-based simulation model. I believe that this analytical framework of applying diffusion processes to the horizontal merger is novel and helps explain the issues that are common in many mergers that have been well vetted but nonetheless fail.

Theoretical Contributions

In this dissertation, I believe I have contributed to the scholarship surrounding horizontal mergers. Specifically, **I argue that a root cause of the integration issues often discussed in the literature (Cartwright, 2002; Kansal & Chandani, 2014) is the unsuccessful diffusion of a new organizational identity in merging knowledge-economy firms.** Executive leadership, with or without the participation of the acquired firm (Giessner, Viki, Otten, Terry, & Täuber, 2006), develop this new identity, become its earliest adopters, and begin to spread it to the larger organization. As with any diffusion process, however, the adoption of the new innovation is uneven, and this creates uncertainty between individual contributors in the merged organization. This uncertainty, in turn, results in wasted effort and lower performance. Lower performance by individual contributors will eventually be reflected in the performance numbers of their organizational units, which will ultimately be communicated to the market as a disappointing quarterly report (at least for publicly traded organizations). Individual employees who were skeptical of the change in leadership will ascribe these failures to the merger, and employee resistance to the new organizational identity will harden further. In extreme cases, this vicious cycle of disagreement, inferior performance, and hardening resistance may eventually cause the merged organization to fail outright. More commonly, it delays the process by which everyone assimilates to the new organizational identity, finally allowing work to proceed with less uncertainty and improved performance.

Unlike most other contexts in which diffusion that has been studied, such as the adoption of smartphones or the spread of a social media service, the diffusion process of a merger must occur inside the organization while the organization continues operating within its market environment to produce value for customers. This constraint heavily informed the simulation design as described in Chapter 3. Individual agents within the simulated organization are sensitive to the performance of their organizations, both current and past, and this sensitivity — implemented in the simulation as the resistance to change mechanism — may have significant repercussions to the larger organization in which they are all embedded.

These repercussions are much more serious in the many organizations today which are virtual (Ahuja & Carley, 1999; Drucker, 1988), requiring experts in different domains to work together to produce value for customers. These experts rely on a common framework and understanding of the organization to collaborate successfully (Liang, Moreland, & Argote, 1995). This common framework is often lacking when workers from different legacy organizations are suddenly expected to work together, but this problem can also occur when some individuals from one legacy organization adapt to the new organizational identity sooner than others in the same organization. In the model in Chapter 3, this effect is implemented through the homophily bias mechanism, where each agent pays attention to the beliefs of the specific alter whose information they are considering, but also to the changing opinions of others with whom they work. As new beliefs shift from being new and unusual to being widespread, individual agents will shift from being suspicious of those with the new ideas, to being relatively indifferent to them, to taking them for granted and being suspicious of others who do not hold those beliefs. These ego-based diffusion phenomena occur throughout the simulation continuously and at a different pace for each agent.

The difficulties of horizontal mergers may also arise at firms where the interaction of knowledge workers is not central to value production; for example, in manufacturing plants. However, in these cases the solutions are simple to describe (if not to perform). Since knowledge is explicit and already documented at such firms, it is much easier to simply perform mass layoffs and then socialize new workers into the new, shared organizational identity. Such firms were common in the past, but are growing rarer. I believe that one reason the merger failure rate has stayed relatively constant over the years is that any improvements in due diligence and monitoring have been offset by the increasing number of highly virtual firms that are attempting large horizontal mergers.

I believe that these problems are less common in vertical mergers — for example, a tire company buying a rubber plant — because there is less of an intention to integrate operations across the merged organizations. Such organizations can generate value by means external to each integrated organization — e.g., the rubber plant now sends most of its rubber to its tire company parent, improving the efficiencies between elements of the value chain. Integrating back-office functions is not central to the value proposition of the vertical merger and can therefore be executed when and if it feels low-risk to the organization and not before.

My research also, I believe, **contributes to organizational culture research by presenting an empirical and multilevel theory and tool for measuring changes in culture**. This research allows theorists and organizational change practitioners to quantify changes in organizational culture as long as they are willing to accept the underlying assumption that the symbols people use in their work reflect the culture of the organizations in which they are members. As such, the theory embodied in my tools inherently subscribes to symbolic interactionism (Stryker, 1980, 2008). While the modern literature on organizational culture has focused heavily on the use of the survey implement, I believe the natural language processing approach described in my dissertation has several advantages. For one, organizations may collect the necessary data without disruption to their workers and without worrying about demand characteristics — the effects on data when the respondents know what the "right" answers are and feel pressure to supply them (Orne, 1962). Moreover, my approach is inherently empirical, and change practitioners and theorists can aggregate to their preferred unit of analysis — whether that be project teams, departments, regional groups, or entire organizations — as long as that preferred unit can be aggregated from the unit of analysis inherent in the data.

Finally, my research extends the state of the possible in COT (Morgan & Carley, 2015) modeling. **To my knowledge, the Unified Network Model is the first active multilevel simulation of organizational change in the world.** Just as I have borrowed mechanisms from influential models of the past, including GARCORG (Carley, 1986), Construct (Carley, 1991), and the Mutual Learning Model (March, 1991), so too will other simulationists build models that borrow mechanisms and capabilities from the Unified Network Model. The Unified Network Model embodies the theory refined from the findings of Chapter 2 and described above: that large organizations must both perform and socialize their members into their identities continuously, and the shared organizational identity of a newly-merged organization must spread through the organization even while the its members continue to interact and work.

Interactions between Chapters

In the Unified Network Model in Chapter 3, I chose to limit individuals to evaluating specific alters only against their immediate neighborhood reciprocal partners. That was because of the surprising finding of Chapter 2 — that those who change the most are also those with the most positive sentiment in their language. This finding suggested that work is often a performance for the benefit of coworkers, and it therefore provided further support for Goffman's perspective (Goffman, 1961). An individual's choice of words communicates far more than the literal contents of the message; it can be used to affirm or disavow membership in an organization or to agree or to disagree with a specific position.

Another finding which recurred in both Chapter 1 and Chapter 2 was that legacy membership mattered. The pre-merger organizations, which no longer traded on the Dow Jones or held quarterly earnings calls, were still affecting individual employee behavior more than a year after their demise. This also influenced the decision in Chapter 3 to implement a fully active multilevel organizational simulation that incorporated legacy affiliation. The simulation, in turn, reflects the truth that other research has reported: that organizations with vastly different cultures and values are less likely to have successful mergers than those whose cultures are more similar (Stahl & Voigt, 2008).

The contributions and associated outputs of Chapters 1 and 2 also directly relate to the ability to model the MergedCo case study in the simulation in Chapter 3. Specifically, the reciprocal networks developed in Chapter 1 were used to create the "works with" network of the case study simulation, and the Morgan Corpora Comparison (MCC) method developed in Chapter 2 was used to identify language tokens for each individual. I then used the top 3 tokens for each individual to create the knowledge network for the case study simulation.

Methodological Contributions

A major theme of Chapters 1 and 2 of this dissertation is answering the question, "Can I take data directly from the organization at work and use it to understand that organization?" More specifically, "Can I take large sets of email records and use them to understand their originating organization?"

The answer is a strong yes.

In Chapter 1, I focused on the senders, recipients, and timestamps of emails. Email, like other modern communications, is typically responded to either promptly or not at all (Wuchty & Uzzi, 2011). The modern organization uses email for a wide variety of purposes, including making announcements to large groups, informing small groups without expecting interaction, creating a contingency against inaction by a peer, and supporting the completion of work within the organization. This last purpose was the most important to me, because I wanted to understand whether an individual's network of coworking partners had important effects for the individual in this context.

I therefore developed an algorithm that defined a network link to include not only the recipients of messages, but also the response time to those messages. Windowed reciprocal ties only exist between two people, A & B, when Individual A can routinely expect email replies from Individual B within 24 hours of Individual A's original message. This removed a huge number of nonreciprocal links from the network and clarified who really works with whom.

Because windowed reciprocity defines a dyadic property between two entities — a link — it can be aggregated to a higher level of analysis. As such, in Chapter 1, I showed that intra-organizational ties were changing between Time-1 and Time-2 the way one would expect, and I used windowed reciprocity at the functional group level to understand the dynamics between functional groups at MergedCo. Later in Chapter 1, I showed that the network composed of windowed reciprocity ties performs as well as or better than a network of raw email ties at predicting survey outcome measures at both the individual and functional group levels.

I implemented algorithms for calculating windowed reciprocity as networks between individuals in Java, which are available from a private GitHub repository and upon request.

In Chapter 2, I wanted to work with the subject and body of emails, which required developing natural language processing capabilities. Before doing that, I needed to build a process for sanitizing the tokens to pass individual privacy and corporate secrecy checks before making the data available to myself and others. This sanitization procedure is documented in Appendix A and is available as Java code from a private GitHub repository.

Once I had the sanitized tokens, I needed an approach for measuring the use and prevalence of various tokens as a proxy for organizational culture. However, emails vary widely in length, and the number of documents was also very different between Time-1 and Time-2. I therefore needed a method for measuring change in token usage that was robust to wide variances in document size and corpus size. I developed a method, again implemented in code, that I call the Morgan Corpora Comparison (MCC) method. Chapter 2 also includes a section comparing MCC to TF-IDF (Sparck Jones, 1972) and Jaccard similarity (Jaccard, 1912), showing that MCC complements but is not equivalent to either of those approaches. A utility script that implements the MCC method is available on GitHub and has already been used by other research teams.

One key advantage of the MCC is that it not only produces a list of terms that are distinct between two corpora, but it also provides a single value that quantifies the overall difference between any two corpora. It therefore produces a dyadic measure between any two entities, at any level of aggregation, based on the provided corpora. This allowed me to compare:

- Entities at Time-1 to themselves at Time-2: How different was their email language between Time-1 and Time-2?
- Individuals at Time-1 to the larger organization at Time-1: How different was the individual from the organization as a whole at Time-1?
- Individuals at Time-2 to the larger organization at Time-2: How different was the individual from the organization as a whole at Time-2?
- Two entities within the same time period: How different is their language from each other's at the same point in time?

Each of these comparison types proved valuable in the work. By comparing individuals at Time-1 to themselves at Time-2, I created a distribution of individual language change between Time-1 and Time-2. By comparing individuals to the larger organization in each time period, I was able to identify tokens that individuals used much more frequently than is typical at the organization. I used this output to inform my knowledge networks in Chapter 3, but in Chapter 2, I also combined that information with token sentiment scores from VADER (Hutto & Gilbert, 2014) to examine individual sentiment based on tokens used.

This combination of outputs led me to the surprising finding, discussed earlier, that the individuals who changed the most used the most positive sentiment words, and that token sentiment was negatively correlated with individual belief reported in surveys that the organization "needed to change." Individuals who had changed to meet their new circumstances felt they had changed enough. Perhaps unsurprisingly, high sentiment at Time-1 was negatively associated with being absent in Time-2.

In short, Chapter 1 and Chapter 2 showed me that structural and textual analyses of the data of the organization at work allowed me to create terms that correlated with self-reported survey data. This suggests that active monitoring of these data in future organizations could perhaps provide insight to executive teams interested in facilitating a merger.

Specifically, monitoring who is responsive to whom at the functional group level could help functional group leadership better understand where necessary functional group collaboration to deliver value for customers is succeeding or failing. It may also indicate the need for more formally recognized cross-functional teams to better support value delivery. This is particularly valuable when many departments are being re-organized and re-shaped, as occurred at MergedCo between Time-1 and Time-2.

Monitoring language sentiment based on distinctive terms of the individuals from their emails, as described in Chapter 2, may also help identify individuals likely to leave the organization, particularly when that overall sentiment changes from prior monitored periods. I would recommend deploying these indicators only to indicate 'at-risk' individuals and only to more senior leadership, who can then sound out middle management or conduct 1-over-1 meetings to assess actual risk. When over-deployed, this sentiment analysis tool can be harmful both to clear communication in the organization and to trust in the organization as individuals begin to be careful as to what they say and how they say it.

In addition to individual-level sentiment monitoring, the MCC method described in Chapter 2 could be used to assess the active spread of the organization culture. This would help managers assess whether existing interventions have been sufficient or if there is a need for more.

Chapter 1 and Chapter 2 also laid the groundwork for Chapter 3's multilevel simulation of organizations, the Unified Network Model, which included mechanisms and dynamics that would make it more feasible to model horizontal merger outcomes.

The Unified Network Model, fed data from the prospective merging organizations, could be used to compare possible outcomes when considering whether to perform the merger. These results would need to be contextualized against first more stylized outcomes and then other analyzed mergers. At a minimum, use of the tool could better indicate whether or more attention needs to be paid to supporting the cultural unification of the two organizations.

The Unified Network Model was able to replicate prior findings of the models on which it was based, as well as replicate six stylized facts important to organizational outcomes and mergers. In the case study, the model was able to predict cleanly who would not be present in Time-2 data. It also modeled the change in stock price at the onset of the merger (albeit using the best available email data, which was after the merger's onset). The Unified Network Model is coded in Java and is also on a private GitHub repository.

Throughout this effort, I have built algorithms and tools to solve problems related to studying the organization at work. These include:

- A sanitization procedure for processing text data to remove highly sensitive and confidential information from the text with minimal harm to its structure and meaning. This is described in Appendix A.
- A process and tool for using email metadata to generate a windowed reciprocity network. This method is described in Chapter 1.
- A process and tool for using sanitized email data to generate a list of most distinctive tokens as well as an overall summary of the differences between any two corpora. This method, called Morgan Corpora Comparison, is described in Chapter 2.
- A process for using the Morgan Corpora Comparison network to generate lists of low-value tokens—those that do not usefully discriminate between any two corpora—and automatically mark them for deletion. This is described in Appendix B.
- An active multilevel simulation model of merging organizations, called the Unified Network Model. This model is the focus of Chapter 3.

All of these tools are available upon request and are coded in Java. Many of them are available as scripts that can be called at the command line.

Limitations and Future Work

Of course, the statistical findings from all three chapters originate from the data of a single organization. To extract this data, even from a willing organization that was able to provide technical support, was nontrivial and required the cooperation of multiple talented individuals.

Significant work remains to validate these findings in other organizations, but I believe that the advent of cloud-based infrastructure may simplify the process of gathering this data in the future.

Key outcomes of interest, such as turnover, are reflected in this work only by the absence of individuals at specific points. In both Chapter 2 and Chapter 3, a key outcome of interest is “present in the data at Time-2” — as opposed to the more direct “left the organization before Time-2.” With the data I have, I could not provide that distinction. It is possible that my findings are an artifact of the sample I was able to extract from MergedCo.

If the goal of this work was to replace surveys, I am afraid it has failed. Even if the technical challenges are conquered, this work cannot supplant the use of survey methodologies in the foreseeable future. Currently, the structural and textual measures in this work correlate with survey measures, but cannot replace them.

My reliance on surveys leaves me vulnerable to their weaknesses, which are hard to mitigate in the context of a high-stakes organizational change like a horizontal merger. In this effort, I found that while survey participation was generally quite high, many individuals simply chose not to answer questions that were of great interest to the research team. For example, LuxuryCo, the acquiring firm, had very few employees respond to the survey instrument measuring commitment to the merger in Time-2. At this point, it was clear they would be highly integrated with MergedCo at all levels. More troublingly, key individuals who, based on the structural analysis, experienced great disruption during the merger avoided the survey entirely.

While I note these concerns and admit the need for replication and validation in other work contexts, I still see this work as offering many potential extensions for future research.

In the domain of applied organizational change, it would be interesting to compare the language of executive leadership with the resulting language of the new shared organizational identity. Such work could take advantage of existing constructs in this space, such as Schoennauer (1967)’s “absorb, blend, and combine” or the more detailed categories employed by Giessner et al. (2006) — assimilation, integration-proportionality, integration-equality, and transformation. The proposed scholarship would be able to assess whether the intentions of the executive teams leading a merger were reflected in their output. For example, if an executive team stated that they intended to create an entirely new organizational culture — the “transformation” strategy from the list above — but the result borrowed a great deal from one of the two merging firms, that may signal a disjunction between intention and execution that could aid in understanding these organizations’ merging dynamics.

For those interested in stylized organizational forms and organizational agility, the Unified Network Model or a successor could be used to compare how organizations with various organizational forms are able to adjust to a “black swan” event (Taleb, 2007). In a “black swan” event, reality changes abruptly. The Unified Network Model could be used to measure the length of time it takes for the organization to return to equilibrium performance and how that varies across organizational forms, strengths of socialization, and code learning rates.

Many extensions are available for simulationists interested in extending the Unified Network Model. A model of tie development and destruction would give the Unified Network Model more ability to model organizational outcomes over longer time spans or in instances when the organization is changing substantially over time, such as in a startup. Developing and implementing a theory of individual resistance to organizational socialization would also improve the model and allow for modeling individuals who are explicitly rejecting the new organizational identity. Finally, improvements to the resistance to change mechanism to better account for the salience of group identities would allow the impact of original organizational identities to fade gradually over the course of the simulation, better modeling individuals who are no longer invested in that identity.

Implications for the C-Suite and Merger Consultants

This dissertation has several important implications for organizational leadership and managerial consultants, particularly those who lead, advise, or support a horizontal merger at a large organization.

The deduplication of functional capabilities is often a key element of the value proposition of a horizontal merger. MergedCo, for example, predicted tens of millions of dollars in cost savings from deduplication. As this work shows, in networked or virtual organizations, where experts in a variety of domains work together to produce value, a decision to simply retain high performers and remove seemingly lower-performing staff is a likely mistake. Oftentimes, those identified as high performers are near the endpoint of a system of interactions which other, uncredited parties make possible. By removing the other parties that support the work of the high performers, you are destroying the system of value creation. My recommendation: Meet with high performers and walk through who supports them in their work. You are likely to find key functions that must be retained in order to sustain pre-merger performance. Focus on the system of value creation rather than the specific individuals who get credit for that value.

The organization is much larger and more difficult to traverse than the ability to email anyone at the organization would suggest. What this means for leaders of organizations is that your understanding of the organizations you lead is much more local and fragmentary than you might think from looking at those with whom you've interacted with over the year. You cannot measure the state of the organization merely by keeping your "ear to the ground."

To deploy a new identity during a horizontal merger, develop a more comprehensive strategy than simply relying on trickle-down diffusion from leadership. The longer the period where your organization sits between new and old identities, the higher the chance of significant waste and underperformance. Instead of an unplanned, organic diffusion strategy, I would strongly recommend delivering the new identity all at once through onsite intensives for contiguous chunks of the organization, where the new identity can then be installed within a short period of time. Also, you will acknowledge reality and develop trust in your candor if you are willing to admit that reframing one's work to a new organizational identity can take substantial effort, during which organizational output may be lower than the company's targets. Properly delivered, this message makes the change both more palatable and less threatening to team members who are affected.

Merger consultants, you know to avoid taking too much comfort in excellent numbers when doing surveys at an organization, but I would ask you to interrogate yourselves as to who is avoiding answering specific questions — or answering the survey at all. In the case study motivating this dissertation, survey respondents from the acquiring firm who had been participating normally became unwilling to answer questions about merger commitment once it was clear that, contrary to their expectations, both the acquired and acquiring firms would be subservient to a new shared identity and executive team.

Finally, the simulation results confirm a fact that may seem implausible to many hard-nosed business types: Corporate culture matters. More specifically, corporate cultures that are very different make a successful horizontal merger less likely. At a minimum, they lengthen the period of underperformance as individuals who work together in the new organization struggle between their old and new identities.

Summary

In this dissertation, I introduced the idea that a horizontal merger requires the diffusion (Rogers, 2010) of a new organizational identity. In Chapter 1, I refined email data to create a works-with network of those who interact with each other routinely, which shows likely routes by which the new organizational identity will spread. In Chapter 2, I used the text of the emails to understand how organizational culture changed over time and articulated the central premise of this dissertation: that a horizontal merger requires the diffusion of a new organizational identity. In Chapter 3, I embodied this theory in a computational model of organizational change I called the Unified Network Model and used data from Chapters 1 and 2 to create a detailed and intricate simulation of MergedCo. In doing this work, I created multiple tools that are embodied in code and available for reuse. I believe I advanced scholarship in the areas of organizational simulation, organizational change, and organizational culture.

References

Ahuja, M. K., & Carley, K. M. (1999). Network Structure in Virtual Organizations. *Organization Science, 10*(6), 741-757. doi:10.1287/orsc.10.6.741

Carley, K. M. (1986). Efficiency in a garbage can: Implications for crisis management. In J. G. March & R. Weissinger-Baylon (Eds.), *Ambiguity and Command: Organizational Perspectives on Military Decision-Making* (pp. 195-231): Pitman Publishing Inc.

Carley, K. M. (1991). A theory of group stability. *American Sociological Review*, 331-354.

Cartwright, S. (2002). Why mergers fail and how to prevent it. *Business: The Ultimate Resource*.

Drucker, P. F. (1988). The coming of the new organization. *Harvard Business Review*, 3-12.

Giessner, S. R., Viki, G. T., Otten, S., Terry, D. J., & Täuber, S. (2006). The Challenge of Merging: Merger Patterns, Premerger Status, and Merger Support. *Personality and Social Psychology Bulletin, 32*(3), 339-352. doi:10.1177/0146167205282151

Goffman, E. (1961). Asylums: Essays on the Social Situation of Mental Patients and Other Inmates. New York, NY: Doubleday Anchor.

Hutto, C., & Gilbert, E. (2014). *Vader: A parsimonious rule-based model for sentiment analysis of social media text.* Paper presented at the Proceedings of the International AAAI Conference on Web and Social Media.

Jaccard, P. (1912). THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.1. *New Phytologist, 11*(2), 37-50. doi:10.1111/j.1469-8137.1912.tb05611.x

Kansal, S., & Chandani, A. (2014). Effective Management of Change During Merger and Acquisition. *Procedia Economics and Finance, 11*, 208-217. doi:10.1016/s2212-5671(14)00189-0

Liang, D. W., Moreland, R., & Argote, L. (1995). Group Versus Individual Training and Group Performance: The Mediating Role of Transactive Memory. *Personality and Social Psychology Bulletin, 21*(4), 384-393. doi:10.1177/0146167295214009

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*(1), 71-87.

Morgan, G. P., & Carley, K. M. (2015). Computational Organizational Theory. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (Vol. 4, pp. 485-492). Oxford: Elsevier.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American psychologist, 17*(11), 776.

Rogers, E. M. (2010). *Diffusion of innovations*: Simon and Schuster.

Schoennauer, A. W. (1967). Behavior patterns of executives in business acquisitions. *Personnel Administration*.

Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation, 28*(1), 11-21.

Stahl, G. K., & Voigt, A. (2008). Do Cultural Differences Matter in Mergers and Acquisitions? A Tentative Model and Examination. *Organization Science, 19*(1), 160-176. doi:10.1287/orsc.1070.0270

Stryker, S. (1980). *Symbolic interactionism: A social structural version*: Benjamin/Cummings Publishing Company Menlo Park, CA.

Stryker, S. (2008). From Mead to a Structural Symbolic Interactionism and Beyond. *Annual Review of Sociology, 34*(1), 15-31. doi:10.1146/annurev.soc.34.040507.134649

Taleb, N. N. (2007). The black swan: The impact of the highly improbable (Vol. 2): Random house.

Wuchty, S., & Uzzi, B. (2011). Human communication dynamics in digital footsteps: A study of the agreement between self-reported ties and email networks. *6*(11), 8. Retrieved from https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0026972&type=printable

Appendix A: Tokenization and Sanitization Procedure for Email Data

This algorithm was developed as a separate process written in Java. It incorporates technologies from third-parties, including SQLite (SQLite, 2015), SQLite4Java (ALM Works, 2015), Tika (Apache Software Foundation, 2015) and the Stanford Core NLP (Stanford Natural Language Processing Group, 2015).

Currently, the Tokenization procedure assumes a SQLite (SQLite, 2015) database that holds the input data, but this is an implementation decision based on extant technologies, not a strong demand of the tool itself. Any database or even any specially delimited file can be used to store text content and be sanitized with this tool, given appropriate input data extensions or modifications.

The current tokenizer only works on English texts, but it would be relatively straightforward to extend tokenization across language barriers given appropriate classifiers. The majority of the texts, as classified by Tika (Apache Software Foundation, 2015) are in English, while the remaining most relevant languages did not have classifiers.

Expected Input:

1. A database table with texts, where the fields with texts to sanitize are given. Tokens should be consistent across texts.
2. An extant tokenization thesauraus – if tokens that need to be anonymized have already been identified, then you can include them in a separate tab-delimited file, where each line of the file is assumed to be a separate entry. The first element is the word to tokenize, the second element is the token ID, and the third element is the token's class (Person, Place, Organization).
3. An extant ignore list – a separate file where each token that should be ignored is placed on its own line

Expected Output:

The output will be a specially delimited text file ("__;__" is the current delimiter) that can be used for further processing of the text. Because the id of the text message is also recorded, it's straightforward to re-attach content to senders and receivers as needed.

Algorithm Summary

- For each record *r*:
  - For each field *f*:
    - Use Tika to identify language of r_f
    - If language = English
      - Replace all numeric characters with '#' symbol
      - Use the Stanford NER (Named Entity Recognizer) to identify people, places, or organizations
      - Add new people, places, or organizations to entity dictionary. Note that tokens may be nGrams, and as such, may have white-space in the token id. If token is an nGram, add to nGram list.
- Sort nGram list by longest to shortest
- For each record r:

- For each field f:
    - Attempt to apply every nGram
    - Split f into words W by white-space
    - For each word in W
        - Check word against entity dictionary (e.g. will, ;will.)
        - Capitalize word, check against entity dictionary (e.g., Will, ;will.)
        - Remove forward and trailing punctuation, check against entity dictionary (e.g., will, will)
        - Remove forward and trailing punctuation, capitalize, check against entity dictionary (e.g. Will, Will)
        - If word or variants match entity dictionary:
            - Identify replacement points
                - Use a forward-search to identify places where this word exists
                - Filter out forward-search results that are not bounded by white-space on both sides
                - Filter out replacement points that are inside an already-replaced token
            - Replace tokens at each indicated point
                - Track change in movement of all forward replacement points
    - Write out f to f_{name}_temp.txt

Appendix B: The Use of Morgan Corpora Comparison for Stop-Word Identification

Abstract: Delete Lists are lists of words that have been determined to have little useful meaning for textual analysis. One subset of words that are frequently deleted are stop-words. Stop-Words are textual tokens, such as "and", "a", or "the", that provide structural or grammatical impact to a sentence but do not themselves have significant inherent meaning. Identifying stop-words is a routine process in most text-cleaning applications, but frequently is done via user-maintained word lists. I suggest that the corpora comparison technique I devised for word-score polarization can be used to identify low-value words while preserving the bulk of the text tokens. I will use both known and random draw corpora comparisons for this process. By "known" corpora, I mean corpora drawn from explicit data-sources, the emails of one company and the emails of another, for example. "Random-Draw" corpora are created by drawing document sets at random, and therefore this technique could be applied to any sufficiently large text corpus of interest. I use the ability to identify stop words as a proxy for performance in generating useful delete lists.

Introduction

In textual analysis, word removal is a common cleaning process. Words are usually removed because they have little value to the analyst' goals. Stop words are one such set of low-value words. These stop words, such as "a", "and", "the", and "or", have significant structural and grammatical importance, but do not themselves have or convey intrinsic meaning.

Word removal through delete lists significantly reduces the size and complexity of resulting analysis products, such as semantic networks and linguistic tree structures, making it easier to get useful results from these later analyses.

Delete word lists vary from group to group, and these lists are maintained with significant analyst effort. The exact composition of this stop word list may depend on the goals of the group, the text medium, and the analysis goals. There may not be one delete list that every analysis group could agree to.

Because it requires significant human effort to maintain delete words lists, I offer an automated algorithmic technique for inferring which words are of low-value and can thus be discarded safely. I do this via corpora comparison – taking two sets of documents and identifying terms that do not usefully distinguish the two corpora. This technique should be applicable across various mediums, as long as the two corpora are comparable. I explore this approach with both "known" corpora, where there is a reason to distinguish between the documents, and "random draw" corpora, where document sets are drawn at random.

The base case, and the first I explore, is the "known" case. The performance in the "known" case represents "good" performance, and thus that the "random draw" performance could, at best, achieve "known draw" performance.

In the remainder of this chapter, I identify the algorithm used to identify these low-value words, explore the performance from the "known" case, and then identify the additions of the algorithm used to address the random case and report performance.

The Motivating Case: Comparing Two Corpora with Known Differences

This algorithm relies on the odds-ratio of the two corpora in comparison. A token is highly useful and/or interesting if the token is highly distinctive. Essentially, the quantity of interest would answer the question: "if I saw this token, would I immediately know which corpus this word is from?" I use the normalized odds ratio to compare the frequency of the word in each corpus.

I will use the following notation. For each corpora, C, there is a set of tokens, T. Each token, t, appears a certain quantity of times and is noted as the quantity t_C. The sum of these quantities would be noted as T_C. In this case, there are two known corpora, A and G.

The equation, assuming t_A and t_G are both non-zero, ranges from -0.5 to 0.5. If the token appears only in A, the score is .5, if the token appears only in G, then the score is -0.5.

Equation 12. The Transformed Normalized Odds Ratio

$$odds(t) = \left(1 - \left(\frac{1}{\left(\frac{|t_A|}{|T_A|} \Big/ \frac{|t_G|}{|T_G|}\right)}\right)\right) - .5$$

I can also "pre-treat" the corpus by removing words that occur less than a certain number of times; this affects the ultimate distribution of the words. Removing words that occur less than 3 times is often recommended, but I present results across multiple cleaning thresholds. Histograms of token scores across multiple thresholds are presented in Figure 27. I expect to see a normal distribution with a spike at both tails. The effect of cleaning is evident and definitely valuable when moving from no filter to the "3 or more" words filter.

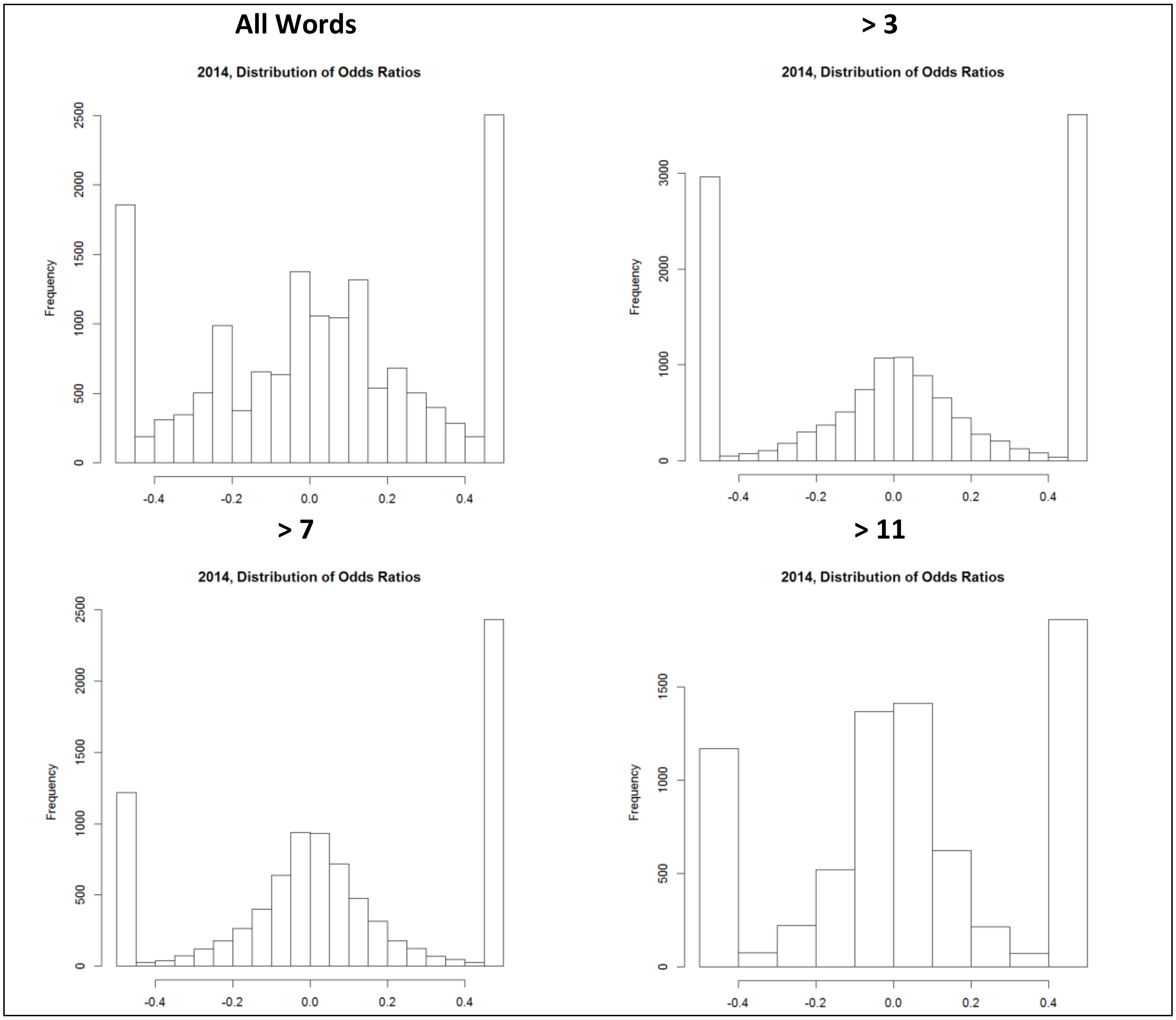


Figure 27. Distributions of token scores across multiple cleaning thresholds

The results in Figure 27 indicate the tested corpus' robustness to word removal. Without special knowledge of the corpora involved, I would expect to see a normal distribution with spikes at both ends. These spikes represent words that only one group or the other uses. Words in each spike usually represent the identities of products, trademarks, or objects related to the company's work. Many words will be used interchangeably across corpora, and this is represented by the bulge at 0.

"Known" Case Performance

In the "Known" case, I have two sets of emails at similar periods of time. The authors of these emails are from two merging companies. I compare the corpus of emails from Group A against the corpus of emails written by Group G from the same time-period.

If I remove terms based on their odds score, I see the following performance from this known corpora. The X-Axis is the percentage of all terms removed, while the Y-Axis is percentage of known stop words removed. Random performance is the diagonal – if I remove 30% of the terms at random, I should get around 30% of the stop words removed as well. The best performance is at the top left of the graph.

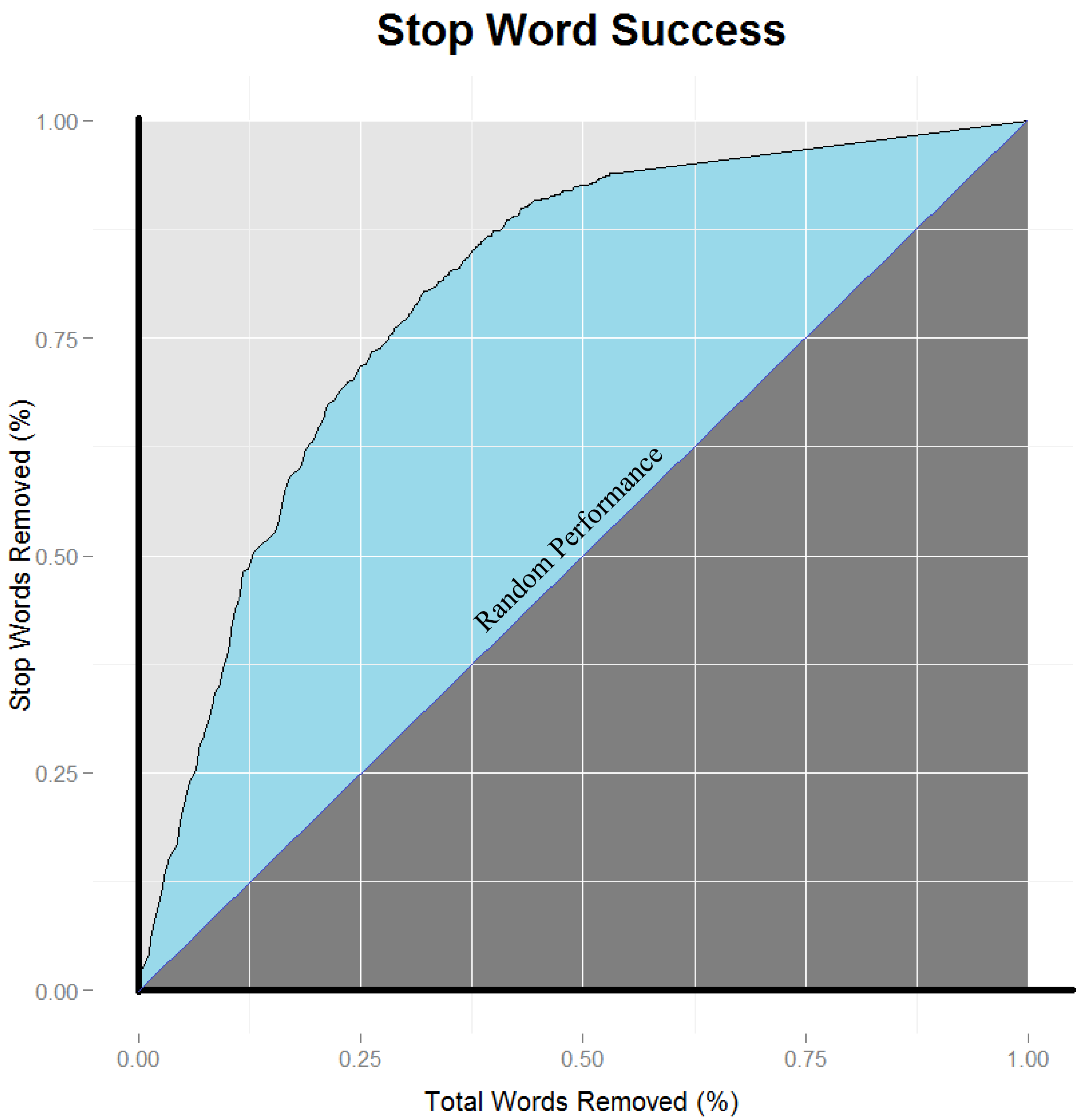


Figure 28. The Performance of the Known Case on Time-1 data The area in blue indicates the relative performance gains compared to a random classifier.

This represents a realistic ceiling to expected performance from the random classifier.

I can evaluate a ROC Curve in two straightforward ways – we can calculate the area under the curve, which indicates the overall fitness of the classifier, and we can calculate the closest point to the ideal optimum: 0,1, which indicates the best performance achieved by this classifier across the entire curve. Another way to think of them is the area serves as a proxy for typical performance, while the closest point gives a measure of best performance. These quantities are highly correlated, but there may be trade-offs between them. For application purposes, there may be a sub-region of the curve that can be considered (it may not be feasible to retain 50% of the words, even if that's the best found performance). Both measures can be calculated for a sub-region.

The General Case: Identifying Low-Value Words by Harnessing the Law of Large Numbers

I don't always have two corpora to work with, and I would like to identify low-value words automatically even when I only have a single set of documents (a corpus) to work with. I will do this by taking the set of documents, creating random subsets as comparison corpora and identifying low-value words based on those random draws.

I use multiple (even many) random corpora based on the original data because I expect words without the ability to distinguish corpora to do so relatively consistently over time. There may be the occasional outlier where "the" is not identified as a low-value word, but on draw after draw, it will be. Evidence accretes over each draw. This gradual accretion of evidence is what allows the algorithm to harness the Law of Large Numbers.

Algorithmic Extensions

Essentially, the algorithm is as follows:

- Identify an appropriate corpus, ***C***.
- Gather evidence via random draws
    - Select documents at random without replacement (each comparison is guaranteed to be non-overlapping sets) to form two corpora, R1 and R2. The number of documents in each random draw is another parameter, ***D***. Larger corpora are less likely to be noisy (in terms of odds ratios), and more likely to include many terms.
    - Calculate the transformed odds ratio (Equation 1) for R1 and R2, for each unique token
    - Sort the token set based on absolute value of the odds ratio and take the ***A***% lowest tokens. The value of ***A*** indicates how much evidence I attempt to gather per random draw. For each token in the ***A***% lowest, note it was marked lowest by iterating its "found lowest" count by 1
- Merge found tokens with the master set of all tokens ever found
- Generate the ROC Curve by starting from the highest realized "found lowest" count and iterating down until all tokens are removed.

Evaluating Performance across Parameters

I want to evaluate this technique across multiple settings of ***C***, ***D***, and ***A*** to provide a more complete evaluation of the approach. Please see Table 28 for the virtual experiment. Theoretically, very low values and very high values of ***A*** should produce sub-optimal performance closer to random – I added more values to the testing of ***A*** in order to characterize the inverse-U shaped curve indicated by theory. I do multiple runs of each parameter setting to evaluate noise.

Table 28. Virtual Experiment for testing the Random Draw Corpora Case

Factor	# of Values	Values
Corpus (*C*)	3	Early 2013, Late 2013, 2014
Document Draw Size (*D*)	4	1000, 5000, 10000, 20000
Accretion Rate (*A*)	8	1, 5, 10, 20, 40, 60, 80, 99
Constants	Setting	
Filter Value	3	
Number of Draws	1000	
Outcomes		
Best Performance	The distance of the performance point closest to 0,1	
Average Performance	Area under the curve	
	Total Combinations	96
	Repetitions	10
	Total Runs	960

The results indicate that, in general, accretion rate has the expected behavior, with area under curve describing a U-Shaped Curve, while optimal performance describes an inverse-U-Shaped curve – see Figure 29. Very small corpora (1000 documents) may benefit from very high accretion values, because there is relatively little information to be gained from each random draw. Larger corpora are usually better, although it appears that small corpora sometimes find a higher optimal performance than medium size corpora. The best performance is found when the accretion rate is about 40%.

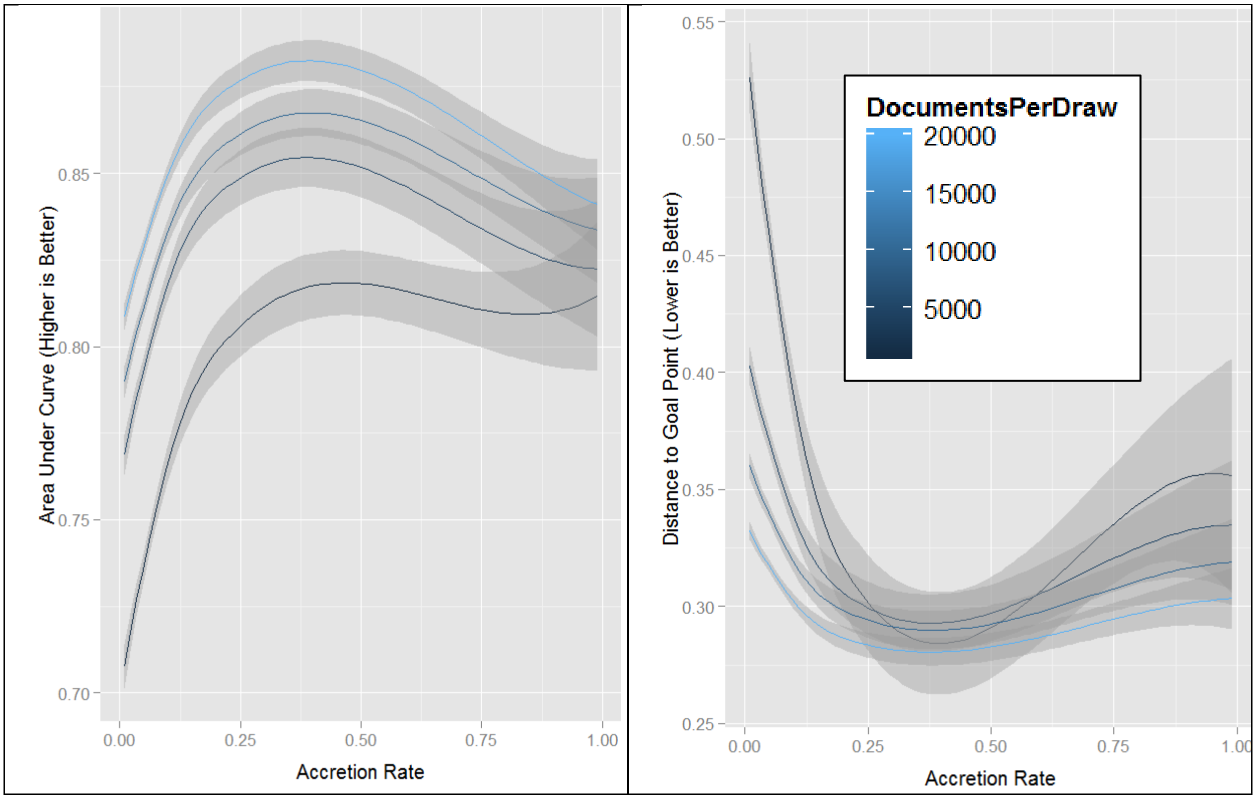


Figure 29. Performance based on Accretion Rate (A) – each line indicates a different value of D.

Figure 30 shows the results for different corpora, ***C***. I do not see a statistically significant difference between performance on these three corpora.

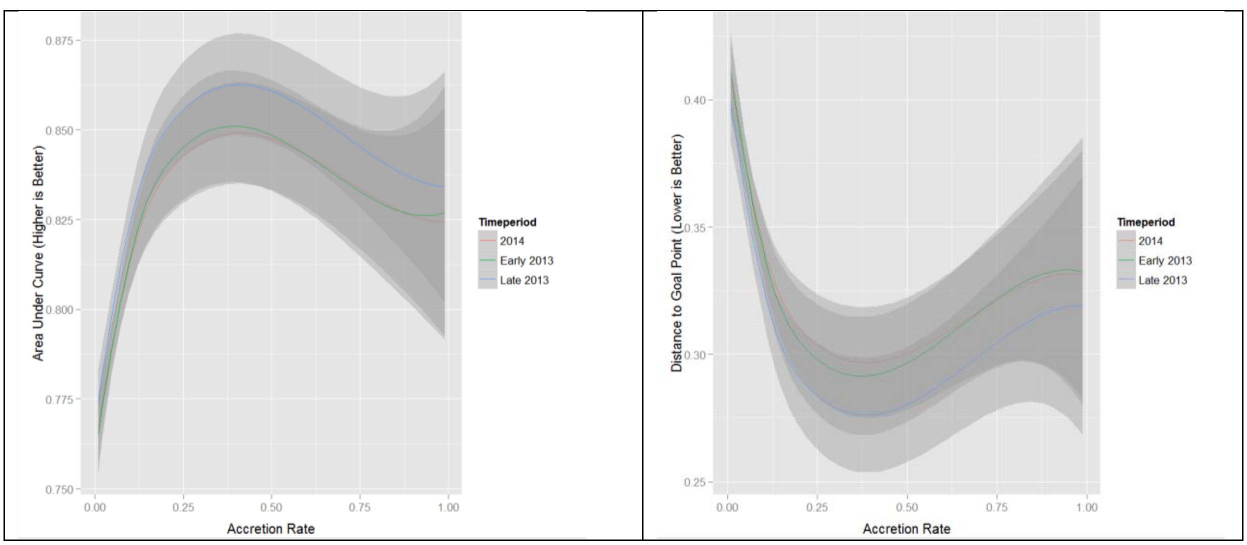


Figure 30. Corpora do not show statistically significant differences.

I can compare this technique to filtering based on TF-IDF. TF-IDF uses the number of times a document appears in text against the number of documents the text is a part of, and can calculate a score. Basic TF-IDF does very poorly at identifying stop-words, with an area under the curve of 0.135 (significantly below random chance) and a best distance at 1, also significantly worse than random chance.

Appendix C: Equations to identify Consensus

These equations are drawn from *Modeling Formal and Informal Ties With an Organization: A Multiple Model Integration*, which specifies the Unified Hierarchical Model. These equations were not changed in the development of the Unified Network Model.

The set *H* constitutes members, *m*, of the organization, *O*, which are more accurate than the organization in perceiving the environment.

Equation 13. Identifying high performers (Equation 1) from Morgan and Carley (2012)

$$H = \{acc(m) > acc(O) : m \in O\}$$

The set C_b are high performers, members of *H*, who do not agree with organizational code on bit *b*.

Equation 14. High performers who disagree with the Organization (Equation 2) from Morgan and Carley (2012)

$$C_b = \{m_b \neq O_b : m \in H\}$$

The set S_b are high performers, members of *H*, who do agree with organizational code on bit *b*.

Equation 15. High performers who agree with the Organization (Equation 3) from Morgan and Carley (2012)

$$S_b = \{m_b = O_b : m \in H\}$$

The probability of the organization the code's value for bit *b* to that of C_b, those who disagree with the organization, is determined by the level of consensus in that set of high performers, and only if the cardinality of C_b is larger than S_b. The term, *a*, is the CODE_LEARNING_RATE defined in Appendix D.

Equation 16. Probability of Change, Part 1 (Equation 4) from Morgan and Carley (2012)

$$p_b = 1 - (1 - a)^{(|C_b| - |S_b|)}, \quad (|C_b| - |S_b|) \geq 1$$

If the cardinality of S_b is as large as or larger than C_b, bit *b* will not change.

Equation 17. Probability of Change, Part 2 (Equation 5) from Morgan and Carley (2012)

$$p_b = 0, \quad (|C_b| - |S_b|) < 1$$

Appendix D: Full List of Parameters for the Unified Network Model

Constant	Range	Meaning
AGENT_NUMER	$1 \leq v \leq \infty$	The number of agents that should be in the simulation. Any setting of this parameter is updated if a network link-list is read in to match the number of unique Node-IDs in that Link-List
FACT_NUMER	$1 \leq v \leq \infty$	The size of the bit-vector that describes reality, which is the same size as the organizational and individual codes
MAX_TURNS	$1 \leq v \leq \infty$	How many turns will the simulation run? This is in addition to the number of turns the Simulation should be in "warmup" mode, defined in WARMUP_TURNS
REALITY_SHIFT_RATE	$0 \leq v \leq 1$	Independent probability of each bit in the reality vector flipping each turn
TURNOVER_RATE	$0 \leq v \leq 1$	Probability of each agent leaving each turn
GRACE_PERIOD	$1 \leq v \leq \infty$	Number of turns before an agent is held responsible for whether the organization is doing well in what they're experts at – if set above MAX_TURNS, it effectively turns off this mechanism
USE_HIRING_COMMITTEE	True/False	If true, use a hiring committee to replace people. If false, replace people who leave at random
USE_POSITION_BASED HIRING_COMMITTEE	True/False	Requires USE_HIRING_COMMITTEE to be set to True to have an effect. If True, the hiring committee is selected dynamically from people who interact with the individual. If False, a static hiring committee is used.
COMMITTEE_SIZE	$1 \leq v \leq$ AGENT NUMBER	Requires USE_HIRING_COMMITTEE to be set to True to have an effect. How large is the hiring committee?
POOL_SIZE	$1 \leq v \leq$ AGENT NUMBER	Only applicable if hiring committees are using the Minimum Aggregate Distance method, how large is the pool of candidates?
USE_HOMOPHILY_BIAS INFORMATION_SPREAD	True/False	Should the homophily bias mechanism affect likelihood of accepting a message from someone?

USE_RESISTANCE_BIAS INFORMATION_SPREAD	True/False	Should the resistance to change mechanism (based on group affiliation) affect likelihood of accepting a message from someone?
UPDATE_CAPACITY	$1 \leq v \leq \infty$	How many messages can an individual accept in a turn? This is a strong mechanism to implement individual bounded rationality.
EXTRA_MESSAGE SEND_RATE	$0 \leq v \leq 1$	The likelihood of sending further messages, which decays based on this parameter. Sending the first message is 0.8^0, the second 0.8^1, the third 0.8^2, etc.
PERCEPTION_ACCURACY	$0 \leq v \leq 1$	How accurate is the perception of others at the individual level? This means that for each bit an alter has, the agent gets right about 90% of them.
STAFF_PERCEPTION ACCURACY	$0 \leq v \leq 1$	How accurate is the organization about understanding its staff? This means that for each bit an agent has, the organization gets right about 90% of them.
CODE_LEARNING_RATE	$0 \leq v \leq 1$	How aggresive is the organization towards learning from its members? 0.5 is only moderately aggressive. High values indicate more aggressive.
SOCIALIZATION_RATE	$0 \leq v \leq 1$	How effective is an organization's socialization? This setting effects the independent probability of any individual bit of an agent has for changing to what the socializing organization believes. Note every organization can have different values for this, but this is the default.
USE_SUBGROUP SOCIALIZATION	True/False	Should each individual organization socialize its members? This is a default setting for the run and can be set per organization.
KNOWLEDGE_COMPLEMENT SIZE	$1 \leq v \leq$ FACT NUMBER	Is knowledge interdependent? A value of 1 means it is not. Higher values indicate more interdependence.
WARMUP_TURNS	$0 \leq v \leq \infty$	How many turns are run with the organizations interacting separately from each other before the full "merged" simulation begins?

WARMUP_SOCIALIZATION RATE	$0 \leq v \leq 1$	How effective is an organization's socialization?