Communication (and Coordination?) in a Modern, Complex Organization[•]

Incomplete Draft. Comments Welcome

Adam M. Kleinbaum
Harvard Business School

Toby E. Stuart
Harvard Business School

&

Michael Tushman
Harvard Business School

October 25, 2007

[•] The authors would like to thank Bill Simpson, Xiang Ao, Misiek Piskorski, Julie Wulf and members of the HBS Strategy Research Conference and MIT Technology, Innovation and Entrepreneurship Seminar for helpful discussions about this project. We most gratefully acknowledge the assistance of multiple individuals from the company that generously and trustingly provided the data for this project. Email comments and suggestions to akleinbaum@hbs.edu or tstuart@hbs.edu.

Abstract

This is a descriptive study of the pattern of communications in a modern organization. We analyze a dataset with more than 100 million electronic mail messages, calendar meetings and teleconferences for a sample of more than 30 thousand employees of a single, multidivisional firm during a three-month period in calendar 2006. The most basic question we examine in the paper concerns the role of observable (to us) boundaries between individuals in structuring communications inside the firm. We observe three general types of boundaries: organizational boundaries (strategic business unit and function memberships), spatial boundaries (office locations and inter-office distances), and social categories (gender, tenure within the firm). In dyad-level models of the probability that pairs of individuals communicate, we find very large effects of spatial proximity and formal organization structure on the rate of communication. Homophily effects based on gender, organizational tenure, and salary levels are much weaker. In individual-level regressions of the extent to which actors engage in category-spanning communication patterns, we find that women, upper middle managers, and medium tenured employees are most likely to participate in cross-group communications.

"*The social system is an organization, like the individual, that is bound together by a system of communication.*" Nobert Wiener, 1948, p. 24

I. Introduction

Coordination (and the communication it implies) is central to the very existence of organizations. Theories of the firm are variously rooted in the coordination benefits of hierarchical control (Thompson, 1967); relative to the market, the superior capacity of managerial hierarchies to efficiently coordinate transactions involving specific assets (Williamson, 1975); and the synergistic potential of coordinating multiple activities within a single corporate structure (Chandler, 1977). In all of these theories of the firm, the key managerial task is to affect coordination. And consistent with these theories, survey and ethnographic studies of managerial behavior have revealed that leaders spend upward of 80 percent of their time interacting with other people (Mintzberg, 1973, Kanter, 1977). The implication of our theories is that organizational members communicate to coordinate activities. Moreover, the quintessentially social nature of managerial work is evident in the vast proportion of management time devoted to interacting with others.

Despite the fundamental role of coordination—and the communication that enables it—to the purpose of organizations, we have little understanding of the interaction patterns among the myriad parts of modern, complex, multi-unit firms. To begin to fill this void, this paper presents a detailed, descriptive analysis of the network of communications among members of a large, structurally, functionally, geographically, and strategically diverse firm. The data comprise more than one hundred million electronic mail messages and tens of millions of electronic calendar entries for a sample of more than 30 thousand employees over a three-month period in 2006. For all individuals in the sample, we also possess basic organizational, demographic, and social information, including gender, salary band, tenure, business unit, function, and office location.

We emphasize that the aim of this analysis is descriptive. Although we report many tables and multivariate analyses, we refrain from formulating specific hypotheses. Rather, our theoretical development is limited to sketching the implications of different theories of organization for likely communication structures within complex organizations. Other than priors formed through our own anecdotal observations of large organizations, we have no a priori reason to give precedence to one theory over another. Moreover, at the level at which we can measure the communications network within the firm, there is not a conclusive, one-to-one mapping between the evidence we marshal and the different theories of organization that populate the literature. Thus, despite our belief that the evidence is consistent with one point of view—the dominant role of traditional conceptions of formal structure in shaping interaction—we will make no strong claim of proof.

Because of their importance in both general theories of social structure and specific theories of behavior in formal organizations, we structure the presentation of the data around the role of observable (to us) boundaries between individuals in structuring communications inside the firm. Given the data available to us, we are able to observe three general types of boundaries: organizational boundaries (e.g., business group, strategic business unit), spatial boundaries (e.g., office locations), and social categories (e.g., gender, tenure within the firm). We conduct many of the analyses in the paper at the level of the pair of individuals: we examine how organizational, spatial, and social boundaries affect the probability that, and the frequency with which, a given pair of individuals inside the firm communicate.

II. Theory

The literature offers starkly different possibilities for how communications might weave together a social fabric inside the firm. Weber provided us with the enduring image of the rationalized, formalized, hierarchical bureaucracy. This image carries through much of the classic work in organization theory and has implications for the structure of communications we might expect to observe within complex organizations. For instance, Chandler's (1962) famously characterized many of the large organizations since the turn of the last century as adopting M-forms, in which operational decisions occur within business units and strategic decisions are managed at the headquarters level. Adopting a Chandlerian view, which is also echoed in contemporary texts on value creation in the diversified firm, we should expect to observe a high density of communications nested within business units and, insofar as the organization is structurally integrated, the headquarters unit will play a central role in bridging communications among the discrete business units. While subsequent work, such as Lawrence and Lorsch's (1967) contingency-theoretic discussion of integration mechanisms and Galbraith's (1973) focus on lateral coordinating mechanisms do somewhat muddy the waters, it is nonetheless the case that classic organization theories imply that interaction patterns inside the firm will, for the most part, be dictated by the organization's formal structure. Thus, we can expect to observe hierarchical communication patterns that unfold within formal organizational units.

Albeit not extensive, there is evidence to suggest that formal structure does strongly influence communication patterns within firms. A few of the network survey-based studies of intra-firm networks have suggested that an organization's formal structure serves is the backbone of the actual relational structure of the firm. For instance, in an analysis of four different types of relations, Han (1996) found that the network of interactions was tightly bound to the formal, prescribed organizational hierarchy. Although not the primary purpose of the paper, a similarly influential role of formal structure in shaping interaction patterns is evident in Burt's (2004) recent analysis of social capital effects in the supply chain function of an electronics firm.

There are, however, views that compete with the potentially dominant role of formal structure in determining patterns of communications and influence within organizations. It is undoubtedly the case that one of organization theory's most taken-for-granted assumptions is that informal structures emerge within organizations. These informal structures are thought to exert significant influence on interaction patterns and, indeed, the "informal" organizational chart is often seen to be more important than the formal one (Selznick, 1943; Mayo 1949; Crozier, 1964; Lincoln, 1982; Krackhardt and Hanson 1993). Indeed, some of our field's most prominent scholars have viewed organizational structure and action as being only loosely coupled (Weick, 1976). In March and Olsen's (1976) influential garbage can model, for example, people, problems, and solutions admix almost by chance, and organizational action can appear almost random relative to formally prescribed decision hierarchies.

If communication patterns map to informal structure, the practical question for us is, what would the observable manifestation of this be in the intra-organizational interaction network? Although many different social dimensions might serve as the foundations of informal structure, because of the extensive evidence from both work and social contexts that actors exhibit homophilous interaction patterns (Blau and Schwartz 1984; Lazarsfeld and Merton 1954), we might expect to find that sociodemographic categories significantly delimit interaction patterns within organizations. For instance, there is considerable evidence of gender homophily in network composition: many individuals are immersed in networks comprising primarily same-sex ties (e.g., Ibarra 1992; Marsden 1988; McPherson and Smith-Lovin 1986; Ridgeway and Smith-Lovin 1999). Similarly, there is evidence that social ties concentrate within age-based and tenure-based strata, while relations that span such categories are much fewer and further between. Thus, if the communication pathways within the organizations are informally structured, we might anticipate that similarities along sociodemographic dimensions have a primary influence on who communicates with whom.

Moreover, since work as early as the 1920s, there is evidence of significant spatial limitations on interaction. Beginning with Bossard (1932), studies have investigated the importance of propinquity in determining the likelihood of friendship and marriage. These studies consistently find that the probability of a relationship increases sharply when two individuals live or work near to one another (e.g., Blau and Schwartz 1984; Festinger 1950; Kono, Palmer, Friedland and Zafonte 1998; Sorenson and Stuart 2001; Zipf 1949). In fact, although there has been rampant speculation that the proliferation of very inexpensive means of communication will reduce the influence of spatial propinquity on interaction propensities, the limited evidence on the issue contradicts the view that low cost communication reduces the impact of space on interaction probabilities (Marmaros and Sacerdote 2006).

The first and second perspectives suggest, respectively, that organizational boundaries or social-demographic-spatial boundaries will exert primary influences on interaction patterns. A third image presents a contrast—one in which rigid hierarchies, status orders, and social categories play a more modest role in structuring intra-organizational interaction. One might trace the origins of this view to Burns and Stalker's (1961) distinction between mechanistic and organic organizations. The notion of more organic organization structures clearly gained steam with Ouchi's (1980; 1981) provocative characterization of "clan structures", in which a commonly held and broadly internalized set of goals were thought to replace hierarchy as a governance principle. A related set of images—the "boundaryless" firm, the "networked" organization (Nohria 1996; Powell 1990), the knowledge-based firm (Kogut and Zander 1992), and so forth, reflect a more contemporary view of the more lateral, less siloed and more collaborative organization. Indeed, Ancona, Bresman, and Kaeufer (2002) argue that in contemporary organizations, leadership is disassociating from formal hierarchies and is migrating to lower levels of the organization.

Contemporary extensions of this line of theorizing in fact emphasize that new communication technologies, by lowering the costs and improving the quality of cross-geography, cross-social group interaction, have finally enabled the implementation of lateral, non-hierarchical structures in complex organizations. For instance, in a recent edited volume, Wiesband (2007) argues that the technological advances of the past decade have vindicated the forward-looking vision of the networked organization that light handedly coordinates the activities of heterogeneous team members who are distributed across a "flat" world. In an environment of sophisticated and widely adopted, enterprise-wide information technology systems, myriad "web 2.0" collaboration tools, and massive bandwidth and accessibility of multiple means of electronic communication, the technology certainly exists to facilitate lateral organizational forms.

Against this theoretical backdrop, a few caveats of this work merit emphasis before we state our tentative conclusions and then present our analysis. First, we know of no comparable study to this and thus we have no benchmarks against which to interpret the magnitude of our coefficient estimates. This means that interpretations of whether or not the data indicate that communications are strongly structured by the categories we examine is necessarily a function of the reader's priors about what the magnitude of these effects would be in a more siloed versus a more lateral organization. Alas, no baseline exists for comparative purposes. Second, although we analyze a vast dataset, we must not allow the enormous volume of the data to cause us to lose sight of the fact that we look at a single organization. At the moment, we have no basis for any claim of generalizability beyond the single organization we study. Third, we know memberships in formal organizational units (group, strategic business unit, and function) only. We do not know how authority relations, including direct lines of reporting or incentive systems put into place by senior managers, shape interaction patterns. Therefore, we operate off the assumption that cross-group communications represent informal ties, but we recognize that many of the lateral connections in the company are managed into existence by the organization's authority structure and incentive systems. By the same token, although we are inclined to interpret dense within organizational-unit ties as indicating that formal structure directs interaction patterns, it remains possible that many or even most of the communication links within formal organizational units in fact occur outside of those dictated by organization chart.

With these caveats in mind, we do observe what we judge to be evidence indicating strong effects of organizational boundaries in siloing communication, which seem to us to be more consistent with the traditional image of the organization, versus a more fluid pattern of interaction implied by contemporary visions of the firm. The categories which prove to strongly influence the likelihood of communications between two actors are business groups, business units, functions, and office locations. As an interesting summary statistic, we find that relative to two people that share none of these categories in common and who are geographically separated by the sample's mean distance, individuals that have all of these categories in common communicate at a predicted rate that is 326 times higher.

After generating estimates of the effects of observable boundaries on the probability of communication, we then flip this analysis on its head: we calculate person-specific predicted probabilities that capture the degree to which each person in the sample engages in communications structures that typify the modal pattern in the data. Aggregating across all of their dyadic exchanges, people who score highly on this variable by definition participate in typical communications within boundaries. Conversely, individuals who have low values on this variable are tantamount to category spanners. Because communication is dense within observed categories and relatively sparse between them, category spanners are far more likely to be responsible for the "weak" (Granovetter 1973) and "bridging" (Burt 1992) ties in the organization. Insofar as it occurs, these individuals are most responsible for the lateral, cross-group communications within the firm.

Before describing the data in greater detail, there are some close precedents in the literature for the analyses we undertake here. Four recent papers—Kossinets and Watts (2006), Marmaros and Sacerdote (2006), Guimera et al. (2006), and Tyler et al. (2003)—each analyze the emergence of network structures in electronic communication networks. The first three papers analyze email data gathered from university settings and the fourth examines an R&D unit at Hewlett Packard. Kossinets and Watts (2006) develop a predictive model of tie formation among members of a university community. Theirs is a model of endogenous network formation—they consider how the structure of the existing network shapes opportunities for interaction. Marmaros and Sacerdote (2006) also analyze the probability that two individuals on a university campus interact. Despite the small size of the campus, they find that geography (e.g., person i and j reside in the same dorm) and race strongly stratify electronic communication patterns. Guimera *et al*. (2006) and Tyler et al. (2003) adopt a more macro approach: they develop algorithms to identify clusters of densely interacting individuals, in the former case within a university community and in the latter, at HP. Finally, email is not the only data source available for analysis of electronic communications; Onnela et al. (2007) take a similar approach to examining the topological structure of an electronic network using data from a mobile telephone provider.

Also important for our endeavor, there is growing precedent for using email to proxy for social networks. For instance, Marmaros and Sacerdote (2006) conduct a survey that provides evidence that email networks in their university data closely parallel friendship networks. Quintane and Kleinbaum (2007) also use a survey to show that work, advice and friendship networks all closely parallel e-mail networks in the context of a not-for-profit organization, but with predictable differences. While we make no such claim in our data—we assume that the majority of the communications that take place within business firms such as the one we study are work related—we nonetheless take comfort in the fact that other scholars have also found that email communications within organizational communities appear to run alongside real-world social ties. As we will convincingly show, they do in these data as well.

Data and Methods

The research site chosen is one of the world's largest information technology companies. The company has over 30 product divisions, organized into four primary product groups: hardware, software, technology services and business services. In recent years, the company has pursued a corporate strategy of integration among its many diverse products and, correspondingly, interdependence among its many divisions; as a result, informal communication across formally-defined intraorganizational boundaries is considered a priority for the company. Although the firm is global in scope, our data collection was limited to the United States[1].

To examine the structure of intraorganizational communication, we draw on a unique data set of electronic communications. The data includes the complete record, as drawn from the firm's servers, of e-mail communications and scheduled meetings (both in-person and conference calls) among 30,328 people over an observation period of roughly three months. The default e-mail retention settings put in place by the firm's IT department cause messages to be deleted from the server after three months (though people can and usually do back up important messages locally).[2] All internal calendar and email information that was on the server at the time of data collection was included in our sample.

The communication data were received from the firm in the form of 30,328 text files, each representing the communication activity of a single person, which were cleaned and parsed in Perl. These files were consolidated, split according to communication type, the duplicates removed[3], and each multiple-recipient message or recurring meeting expanded to include one entry for each unique dyad or occurrence. Finally, BCCs, mass mailings[4] and dyadic interactions including an administrative assistant[5] were removed. The resulting communications data sets included 47 million e-mails and 18 million meetings. To protect the privacy of individual employees, messages and meetings were stripped of all content, leaving only information about the sender and recipient(s), time/date sent, size of the message and any attachments, and whether the message referenced any prior message. The identities of sender and recipients were disguised using an SHA-512 hash algorithm.

---

[1] Law protecting employee privacy makes the data collection for this project virtually impossible in much of Europe, so the company was only able to provide data for U.S.-based employees.

[2] Individual users can adjust their own settings to remove messages sooner (usually to stay beneath the space quota). Conversely, they can also apply for an expanded quota to preserve messages on the server for longer than the default period.

[3] Duplicates resulted from communications between two people who are both in the sample population; in such cases, one copy of the communication would arise from the outbox of the first person and a second copy would arise from the inbox of the second person. Such duplicates were identified and removed using a unique identification number for each message.

[4] Defined as messages or meetings with more than 4 recipients, consistent with results reported by Quintane & Kleinbaum (2007).

[5] Note that in the e-mail data, dyadic interactions occur only between the sender and each recipient, but in the calendar data, dyadic interactions also occur among recipients; the data parsing strategy reflects these different realities.

The second part of the data set, also provided by the company, includes demographic and HR data about each employee and is tied to the communication data through the same encrypted employee identifiers. Data include each employee's: business unit, job function, tenure with the firm, salary band, state and location code[6], and gender.

Collectively, this sample represents 24% of the firm's U.S. employee population, collected through a snowball sampling procedure. Initial access to the organization was provided by members of the corporate sales force: 180 people were identified who were explicitly involved in cross-divisional activity and were invited to participate in the study; 91 of them agreed. Collectively, they communicated with an additional 30,237 employees during the preceding three months; all 30,328 (the 91 core members plus the 30,237 2nd degree members) were included in the sample. Although this is not a simple random sample, it is a compellingly large sample and is roughly representative of the firm's U.S. operations (see Table 1).

Once the data were cleaned and parsed, we collapsed them into a single cross section and created counts at the dyad level of the total number of $i \leftrightarrow j$ messages. In other words, we count the number of communications among unordered pairs of individuals. We then undertook two primary sets of multivariate analyses. In the first set, we model the frequency of dyadic communication based on common group memberships and other pairwise attributes of each dyad of actors. Even when we compress the time axis to treat the data as a single cross section, however, the communication matrices are both enormous and extremely sparse – less than 0.5% of the approximately 450 million possible unordered cells in the e-mail matrix are non-zero. Likewise, fewer than 0.2% of cells in the unordered meeting matrix are non-zero. Given the computing power available to us, it is not feasible to work with the full matrix.

Sampling randomly from the set of the 450 million communicating dyads offers one potential solution to this problem. However, this approach is not ideal because it ignores the fact that the realized ties provide most of the information for the estimation of the factors that affect tie likelihood (Cosslett 1981; Imbens 1992; Lancaster and Imbens 1996). We therefore construct a "case cohort" dataset: our regression models include all non-zero cells and a random sample of zero cells drawn at a 1:1 ratio, which are then weighted according to their probability of being drawn into the analysis sample. We do not stratify on the sampling of zeros; we simply randomly draw the zero cells and each regression column is based on a different random draw. The dependent variable is a count of the number of e-mails (or, in separate analyses, meetings), so we used a weighted Poisson model[7]. Because the Poisson is in the linear exponential family, the coefficient estimates are consistent as long as the mean of the data is correctly specified (Gouriéroux, Monfort and Trognon 1984). Moreover, robust standard errors are also consistent even if the mean is mis-specified.

---

[6] For privacy, the firm provided an encrypted location code that allows us to determine which employees are in the same location, but does not allow us to determine the specific location of any employee, other than to know what state they are in; however, they also provided a distance matrix to allow us to determine the geographic separation between any pair of employees without knowing their specific location.
[7] These findings are robust to alternate specification as a negative binomial model.

Specifically, we estimate models of the form:

$$E\left[Y_{ij} \mid X_{ij}\right] = \exp\left(X_{ij}\beta + \sum_{j} Y_i\beta_1 + \sum_{i} Y_j\beta_1\right) \quad (1)$$

where Y_{ij} is the count of emails exchanged in both direction between individuals *i* and *j*, X_{ij} designates a matrix of pair-level covariates (e.g., dummies for whether actors *i* and j are members of the same business unit), and the sums of Y_i and Y_j indicate the total volume of emails sent or received by the actors *i* and *j*. These counts are included to absorb actor-level heterogeneity in communication volumes, which will influence the likelihood of communication within any pair of actors.

Results from equation 1 give us the likelihood that dyad-level covariates affect the probability of interaction. We also, though, wish to understand the individual-level correlates of category-spanning communication patterns. Following a large amount of research in the social networks literature on boundary spanning, we assume that individuals who communicate across boundaries are most likely to be key brokers within the firm because these individuals are likely to bridge otherwise disconnected groups. This means we need a measure of the degree to which each individual participates in category-spanning ties. A simple approach to approximate category spanning is to take the set of each individual *i*'s email exchanges with alters and count the average number of categories spanned by these ties. The shortcoming of this approach, however, is that it fails to adjust for the fact that some types of categories, e.g., office location and business unit, appear in our data to have a much stronger influence on communication patterns than do other categories, such as gender.

To derive a measure of category spanning, we estimate a dyad-level model as in equation (1) of the rate of communications in the data. Specifically, we estimate the model with six categorical variables that indicate whether the actors in the dyad are in the: same business unit, same function, same office, same salary band, same gender, and same tenure (within 5 years); and one continuous variable, indicating the geographical distance between them. After obtaining parameter estimates from this equation, we then delete all cases in the random sample (i.e., we retain only the 1.2 million undirected dyads in which communication actually takes place, or equivalently, $Y_{ij} \geq 1$). For each of these realized dyads, we then compute:

$$\hat{Y}_{ij} = \exp(\alpha + \beta_1 \times BU + \beta 2 \times fx + \beta_3 \times office + \beta_4 \times dist + \beta 5 \times gender + \beta_6 \times band + \beta_7 \times tenure$$

where $\hat{Y}_{ij}$ is the predicted count in the *ij*th dyad. When $\hat{Y}_{ij}$ is high, the *ij*th dyad represents a pair of actors that would be expected to communicate relatively frequently, given the patterns in the overall data. Conversely, when the predicted count is low, the *ij*th pair shares an attribute profile that leads it to be unlikely to communicate. Thus, the former pair of actors participate in the clustered forms of communication that typify the organization, whereas the latter pair represents a communication link between infrequently interacting groups.

Next, we aggregate $\hat{Y}_{ij}$ to the individual level by computing person-specific averages of the predicted count, or:

$$\hat{Y}_i = \left(\sum_{j=1}^{n} \hat{Y}_{ij} \right) / n$$

where, for each focal actor *i*, *n* indexes the total number of other individuals *j* with whom *i* communicates. When a focal individual *i* communicates exclusively within categories and never across, $\hat{Y}_i$ assumes its maximum possible value. If an individual never communicates across categories, i.e., $\hat{Y}_i$ =exp(α), $\hat{Y}_i$ achieves a minimum.

We then run a second stage regression[8] at the individual level, with the dependent variable defined to be:

$$-1 \times \log(\hat{Y}_i)$$

We log transform $\hat{Y}_i$ to remove skew and then flip the distribution so that positive coefficients in this regression indicate that a covariate is associated with an *increased likelihood that an individual is a boundary spanner*.

Results

We begin with a few descriptive looks at the data. Figure 1 shows the risk-set-adjusted conditional probability that a message sent by an individual in a given salary band will be bound for an individual in each other salary band. Given the distribution of individuals across salary bands, cells are color-coded as red for communications that occur less frequently than expected; blue for communications that occur more frequently than expected; and blue and bold for communications that occur at least 50% more frequently than expected. As the color scheme clearly indicates, communications are heavily concentrated along and around the main diagonal. In aggregate, people communicate more frequently with others in similar salary bands. It is interesting to note that this company defines executives to be in salary ranks eleven to fourteen. We observe consistently higher than expected frequencies in the block diagonal defining executive management, and we see a similar but less pronounced pattern of interaction among lower salaried employees.

Figure 2 provides a first glimpse at the role of organizational structure in directing communications. This figure positions divisions in the company according to their proximities in the inter-divisional communication network. Each node in the figure

[8] For robustness, we also estimated first-stage models that include all two-way interactions and, separately, all two- and three-way interactions, then carried these coefficients through the second stage model; the vectors of regression coefficients of these models were correlated .97 and .96, respectively, with the simpler first-stage model.

represents one strategic business unit and nodes are color-coded according to their memberships in larger business group. Nodes are connected by arcs representing the risk-set adjusted conditional probability of communication volumes between business units; intra-business-unit communications are dropped from this analysis. The nodes and arcs in Fig. 2 are projected into two dimensions by the Kamada-Kawai spring-embedding algorithm, as implemented in Pajek (Batagelj and Mrvar 2006), where close proximity of nodes represents high probability (i.e., relatively frequent) communication between a pair of divisions. The main result of Figure 2 is that nodes are clustered closely together according to super-divisional groupings; this suggests that at the business unit level also, inter-unit communication is largely shaped by the formal organizational groupings that join strategic business units into a broader group structure.

Two nodes in this graph are the sole members of their business groups and warrant special attention: the corporate sales force and corporate headquarters (both treated as "business units"). The sales function of the company is largely centralized, with a massive corporate sales force designed to sell the products of all the company's product divisions. Internally, the sales force is organized according to corporate clients' industry sector. Many of the product and service divisions nevertheless have their own sales force as well. The corporate sales force appears in Figure 2 in blue and is rather peripheral to the inter-unit communication network[9]. Corporate headquarters appears in yellow in Figure 2; it is close to the dense core of the network, but is not highly central. One explanation for the peripheral location of both of these important corporate functions is that by communicating frequently with relatively many other nodes, they appear as less closely tied to any single node, in contrast with the product and service units, who may interact intensively with a small number of alters and very little with the rest. Indeed, results indicate that both the corporate sales force and corporate headquarters have lower standard deviations across their tie strengths than other business units.

One of the interesting findings from our analysis is that the overlap between the email and calendar data is nearly perfect: we find no dimensions along which these two modes of communication appear to be substitutes. While we cannot decouple face-to-face meetings from scheduled teleconferences—the calendar data record both types of communication with no discernable distinction made between them—the correspondence between email and these two other modes of communication is striking. In unreported regression analyses, we find a 0.98 correlation between the vectors of coefficients in a set of regression specifications in which the dependent variable is defined to be counts of dyadic communications based on, respectively, email and calendar ties. The one dimension along which we might most expect patterns of communication to vary across media is geographic space: a reasonable hypothesis would be that email is used more frequently among geographically dispersed colleagues separated by multiple time zones, and other forms of communication dominate closer to home. However, even along this dimension, the two forms of communication appear to be remarkably similar. Figure 4,

[9] When asked about this, a company representative suggested that sales people might routinely include customers in their communications, causing those communications to be omitted from the data set. However, there is little support for this suggestion: members of the corporate sales force have, on average, substantially more communications in the data set than most other business units.

for instances, plots the proportion of both email and calendar communications that connect two actors separated by the geographic distances indicated on the horizontal axis (log scale). Obviously, the two curves are very similar; anecdotally, numerous informants at the company have described conference calls at odd hours as a key challenge of global teams.

Results from the dyad-level Poisson regressions appear in Table 2. The coefficients in the table indicate unit differences in the log of the expected event count, which can be converted into incidence ratios by exponentiating coefficient values. The single largest organizational effect on the probability that two individuals communicate is being in the same business unit. When i and j are in the same BU, they interact at exp(1.90)=6.69 the rate of two otherwise similarly related individuals in different business units. The effect of being in the same function is large, but much smaller than same BU: two individuals in the same function communicate at 3.36 times the rate of those who are in different functions, *ceteris paribus*. Adding an extra dummy to capture individuals who are in both the same business unit and the same function reduces these estimated rates only modestly.

In theories of corporate-level strategy, the headquarters unit of the firm plays a particularly salient role in value creation (insofar as the firm does, in fact, create value in excess of its constituent parts). From the strategy and planning role of the headquarters unit in Chandler's (1962) characterization of the M-form, to more contemporary theories of the need for an active role of headquarters in generating business-unit competitive advantage (Collis and Montgomery 1998; Goold, Campbell and Alexander 1994), the coordinating activities of the headquarters unit is at the center of theories of value creation, and more fundamentally, of the very desirability of the existence of multi-business firms. To explore the role of headquarters in this firm, we include dummy variables to indicate whether one or both of the individuals in a dyad are in the headquarters unit (respectively, "onechq" and "bothchq"). The coefficient on bothchq—which can be interpreted as an interaction effect between same business unit and business unit equals headquarters—is insignificant. This indicates that members of corporate HQ have roughly the same propensity than members of other BUs to communicate internally. Moreover, the positive coefficient on "onechq" indicates that, relative to other cross-BU pairs, those in which one member of the pair is from the headquarters unit are 1.14 times more likely to interact. This effect is substantially stronger in the OneExec case: dyads consisting of one member of headquarters and one executive (who may or may not be the same person) interact 1.4 times more frequently. Thus, there is some evidence that members of the firm's headquarters staff are, compared to other business units, more outwardly focused in their interaction patterns.

Turning next to geography, at the present time we have available three different measures of spatial proximity: we know whether i and j are in the same office, whether they work in the same state[10], and we know the distance separating each pair of actors. Sameoffice is in fact the largest estimated effect on the rate of communication. Two

[10] We experimented with numerous specifications and found the results to be very similar, but we achieved the best model fit using sameoffice and distance; we present those results here

individuals communicate at [exp(1.98*sameoffice)=] 7.24 times the rate as otherwise identical, cross-office and cross-state pairs. Relative to the same comparison—two individuals in different offices and states—pairs of individuals in the same state but not the same office interact at 1.9 times the rate (results not shown). The distance measure also indicates that proximate people communicate more frequently: a one standard deviation (845-mile) increase in distance reduces communication by a factor of 0.82. Thus, by either measure there is a pronounced effect of spatial propinquity on the probability of interaction.

We next turn to measures of social categories: gender, tenure, and salary band. The gender composition of the sample is approximately 66 percent male and 34 percent female. Surprisingly, we find that the conditional effect of same gender is modest—male-male and female-female dyads are only marginally (1.12 times) more frequent communicators than are cross-gender pairs. Moreover, this modest increase in the rate appears to be substantially larger for female-female pairs than for male-male pairs; the effect of "bothfemale" in the full regression boosts the samegender effect from 1.12 for male-male dyads to 1.39 for female-female dyads. These effects appear to differ, however, within the executive ranks. In dyads where both members are executives, the samegender effect is roughly triple, increasing the frequency of communication by 1.36 over male-female dyads. However, the insignificant bothfemale variable indicates that any homophily among same-sex executive dyads is shared equally between women and men.

There is also evidence of both intra-salary-band and within-cohort homophilous tendencies, but these effects as well are much smaller than either the spatial or organizational influences on communication patterns. Individuals in the same band in the company's fourteen-level salary grades communicate at a [exp(.34)=] 1.40 times higher rate than more distant cross-salary-rank dyads. The effect of *i* and *j* both being within within five years of one another in their tenure with the firm has a similar estimated effect of [exp(.27)=]1.30 on the rate of communication.

The overall impression to emerge from the dyad-level analysis is that organizational structure and geographic space sharply delimit patterns of exchange. Social categories—gender, tenure, and salary grade—also shape propensities to interact, but the magnitudes of their effects are much lower than are the effects of organization and space. Taking into account the fact that the models are multiplicative, individuals in the same office, same function, and same business unit are far more likely to interact than a randomly chosen pair of people. Sociodemographic categories, while positively influencing proclivities to communicate, do so to much less of an extent.

To provide an overall sense of the data, we have created distributions of the proportions of categories spanned by actually communicating pairs of individuals, relative to a random baseline given by the empirical distribution of individuals within and across observed categories. This analysis appears in Figure 3, in which we focus on the following five categories within the data: business unit, function, office, salary band, and gender. Blue bars represent the realized ties, and red bars the distribution of all possible pairs of communicating dyads (*i.e.*, all approximately 450 million cells of the ½*n*(n-1) network of potential dyads). Of course, the distribution of possible ties is shifted far to the right of the distribution of realized links. Slightly more than 50% of the actual communicating dyads in the company span none, one, or two boundaries. By contrast, if we examine the distribution of possibly communicating pairs of actors within the company (which implicitly adjusts for the risk set of actors within and across categories) less than 10% of pairs drawn at random would span two or fewer boundaries. At the other end of the distribution, just 5% of actually communicating dyads span all five boundaries, whereas 25% of the randomly generated pairs are neither same business unit, same function, same office, same gender, nor same salary band.

If the dominant patterns of communication in the organization are clustered within organizational unit and geographic space, the interesting question becomes: who are the individuals that form bridges across these clustered groupings by participating in cross-category exchanges? To answer this question, we turn our attention to the regressions of $\hat{Y}_i$, the individual-level measure of propensity to communicate across categories, on dummy variables for the categories themselves. Results of this regression appear in Table 3. Before describing them, however, we note one important caveat: our analyses thus far consider only the mean of $\hat{Y}_i$, but it is not obvious that the mean is the optimal summary statistic. Almost across the board, the results of this analysis are surprising to us.

First, despite the fact that evidence from other contexts has repeatedly found that men have more structurally diverse networks, the reverse appears to be the case in these data. *Ceteris paribus*, the women in the company are engaged in more category spanning ties. Note that this result is net of controls for the 14 salary grades and job functions. As well, administrative assistants have been deleted from the data. Thus, this result is unlikely to be a spurious result of women being clustered into roles in which they provide administrative support for men.

The omitted salary band category is fourteen, which is the highest in the company. Relative to band 14s, we find that individuals in the high-middle sections of the salary distribution, most notably those in bands 10 to 12, engage in the most boundary spanning communications. And we find that those in the middle segments of the tenure distribution are also most likely to be high boundary spanners. Thus, we can tentatively conclude that the middle management ranks within the firm are central in the cross-cluster communication stream.

Beginning with business units, we find that members of corporate headquarters are in fact *less* likely than members of most other business units to have statistically improbable, cross-category communication patterns. In the table, corporate headquarters is the omitted business unit category and almost all of the other BU dummies are positive and statistically significant, indicating that membership in that unit is associated with a more category-spanning pattern of communication. Compared to each of these other units, people in corporate headquarters are less likely to have category-spanning communication patterns. The sole exceptions are bu_13, representing the corporate sales force and to a lesser degree, bu_2 representing the business consulting division; the coefficients on the bu_13 and bu_2 dummy variables are negative and significant, representing less category-spanning patterns of communication among members of the sales force and consulting staff, compared with members of the corporate headquarters. Because we know from Table 2 that members of corporate HQ are more likely to engage in business unit-spanning ties, this implies that their cross-BU communications are predominantly homophilous with respect to function, geographic space, and social demographic categories.

Conversely, we find that individuals whose job function is described as general executive management, the default category for the function dummy variables, have communication patterns that are less conventional. Every other job function dummy had a negative coefficient in our analysis (several were insignificant), indicating that members of every other function have less cross-boundary communication patterns than members of the general executive management function.

Discussion and Conclusion

We began with the observation that although theories of communication and coordination are central to field of organization theory, we know little about the structure of communication in the modern, complex organization; with this paper, our first and primary objective is to contribute a first look at the way communications are structured in one large multi-unit firm using a unique data set of e-mail communications and electronic calendar data. Reiterating the caution stated at the outset that we have no metrics against which to compare our results, we believe that the communication structures observed within the firm we study are most compatible with traditional theories of relatively rigid formal organization structures and sharp spatial limits on interaction patterns. We find that communication is largely structured by the formal, bureaucratic organization structure (business unit memberships and functional roles). We also find that spatial boundaries—office locations and distances between individuals' locations—also have a large effect on the shape of the network inside the firm. Social categories matter as well, but to a much lesser degree.

We focused some of our analyses on the coordinating role of the corporate headquarters and find mixed results. Although cross-business unit pairs of individuals are more likely to communicate if one member is in CHQ, people in CHQ are actually appreciably less likely than members of many other BUs to occupy boundary-spanning positions in the broader social structure of the firm. It seems that members of CHQ communicate across business unit boundaries relatively freely, but they do so less frequently than average along functional, geographic and social boundaries. Of course, the mere fact that members of corporate headquarters are not involved in boundary-spanning social networks is not necessarily diagnostic of the role of CHQ staff in facilitating inter-unit coordination. It is possible, for instance, that CHQ may play the role of brokering introductions (Obstfeld 2005; Simmel 1902). Or, the existence of boundary spanning ties at lower levels of the organization may be the direct result of lateral coordination mechanisms put into place by CHQ but not observed by us. Nevertheless, the image of CHQ staff as cutting across clusters of social relations within the firm does not appear to be compatible with the actual communication patterns in this organization.

The limitations of the data and analysis notwithstanding, the descriptions provided here should provide a fertile basis for future research. We see a number of possible directions in which to extend this work. First, all the analysis presented here is at the aggregate level; further research is warranted into the egocentric networks of those actors with very high, category-spanning values of $\hat{Y_i}$. Aside from the gross categorical variables explored here, what individual variables lead some actors to occupy boundary-spanning positions more than others? In future work, we hope to explore this question using the career histories of each employee to determine more specifically the antecedents of, literally, improbable ties.

Second, given the relatively short observation period of our electronic data, we have opted to treat the data as a cross-section, but clearly boundary-spanning communications must come and go over time. Longitudinal analysis can explore the dynamic nature of intraorganizational communication and, perhaps, help demonstrate causality in a domain ripe with endogeneity. We hope to begin to explore the dynamics of boundary-spanning using a second panel of electronic communication data gathered nine months after the first. And third, this paper has focused on the antecedents of boundary-spanning behavior in an organization. While there is an extensive literature on the consequences for an individual who occupies a brokerage position, there is little research into the consequences for organizations of individual-level boundary spanning (cf. Kleinbaum and Tushman 2007).

Variable	Correlation
Business Unit	0.737
Function	0.904
Salary Band	0.816
State	0.970

Table 1: Representativeness of the sample; the correlation reported is between the distribution vector of individuals in the sample compared to the population. The sample is slightly over-representative of the corporate sales force (which was the origin of the snowball sample) and the executive salary bands.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	**0.05**	0.00	0.00	0.00	0.00	**0.10**	**0.20**	0.01	0.13	0.07	**0.43**	0.01	0.00	0.00
2	0.00	**0.03**	**0.05**	**0.05**	**0.01**	0.01	**0.32**	0.31	0.22	0.00	0.00	0.00	0.00	0.00
3	0.00	**0.01**	**0.14**	**0.06**	**0.01**	0.05	0.08	0.24	0.11	0.14	**0.16**	0.01	0.00	0.00
4	0.00	0.00	0.01	**0.15**	**0.01**	0.03	0.07	0.14	0.15	0.20	**0.12**	**0.08**	**0.02**	**0.03**
5	0.00	**0.00**	**0.01**	**0.07**	**0.20**	**0.06**	0.14	0.27	0.09	0.07	0.02	0.01	**0.02**	**0.05**
6	**0.00**	0.00	0.00	0.01	0.00	**0.15**	0.14	0.20	0.25	0.18	0.05	0.01	0.00	0.00
7	**0.00**	0.00	0.00	0.01	0.00	0.05	**0.18**	0.23	0.27	0.20	0.04	0.01	0.00	0.00
8	0.00	0.00	0.00	0.01	0.00	0.02	0.08	0.27	0.32	0.24	0.05	0.01	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.16	0.35	0.31	0.09	0.03	0.01	0.00
10	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.10	0.25	**0.38**	**0.15**	**0.06**	**0.01**	**0.00**
11	**0.00**	0.00	0.00	0.01	0.00	0.01	0.02	0.05	0.18	0.36	**0.19**	**0.12**	**0.04**	**0.01**
12	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.03	0.11	0.30	**0.25**	**0.16**	**0.09**	**0.04**
13	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.02	0.07	0.20	**0.23**	**0.24**	**0.12**	**0.11**
14	0.00	0.00	0.00	0.02	**0.01**	0.01	0.00	0.00	0.03	0.11	**0.13**	**0.21**	**0.24**	**0.23**

	CHQ	GBS	GTS	S&D	SWG1	SWG2	STG
CHQ	**0.70**	0.06	0.05	0.08	0.01	0.01	0.02
GBS	0.07	**0.79**	0.02	0.05	0.00	0.00	0.00
GTS	0.10	0.04	**0.69**	0.12	0.00	0.00	0.00
S&D	0.07	0.05	0.06	**0.71**	0.01	0.01	0.01
SWG1	0.04	0.03	0.01	0.09	**0.66**	0.02	0.01
SWG2	0.05	0.02	0.00	0.11	0.03	**0.70**	0.01
STG	0.12	0.01	0.01	0.11	0.01	0.01	**0.58**

	AD	FI	GM	HR	MK	RD	SC	SL	SV
Administration	**0.27**	0.05	**0.01**	**0.02**	0.06	0.06	0.02	0.16	0.30
Finance	0.09	**0.55**	**0.02**	0.02	0.02	0.01	0.01	0.14	0.14
General Exec Mgmt	0.05	**0.09**	**0.22**	**0.07**	0.07	0.09	0.00	0.21	0.09
Human Resources	0.08	0.04	**0.04**	**0.51**	0.02	0.03	0.01	0.10	0.15
Marketing	0.06	0.01	**0.01**	0.00	**0.67**	0.05	0.00	0.11	0.05
R&D	0.04	0.01	**0.01**	0.01	0.04	**0.68**	0.00	0.08	0.10
Supply Chain	0.10	0.04	0.00	0.01	0.01	0.02	**0.48**	0.13	0.19
Sales	0.04	0.02	**0.01**	0.01	0.03	0.03	0.01	**0.71**	0.13
Services	0.07	0.02	0.00	0.01	0.01	0.03	0.01	0.12	**0.69**

Figure 1: Probability of e-mailing to the column group, conditional on the sender being in the row group by salary band, business unit, and function, respectively. Red cells are below expected counts; blue cells are larger than expected counts; bold numbers are at least 50% above expectation. Selected business units and functions shown.

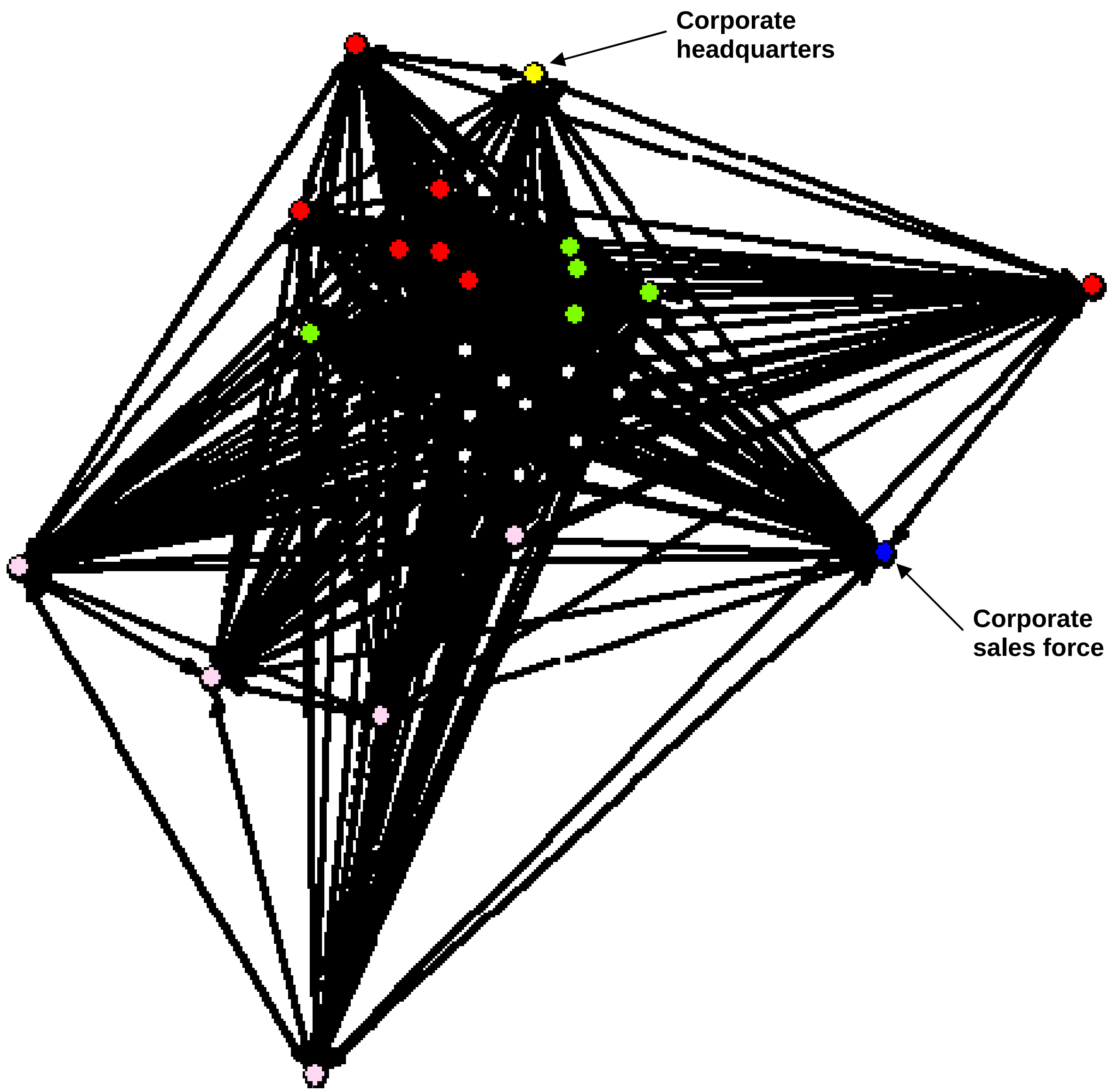

Figure 2: The communication network among business units, color coded by each unit's membership into a larger group.

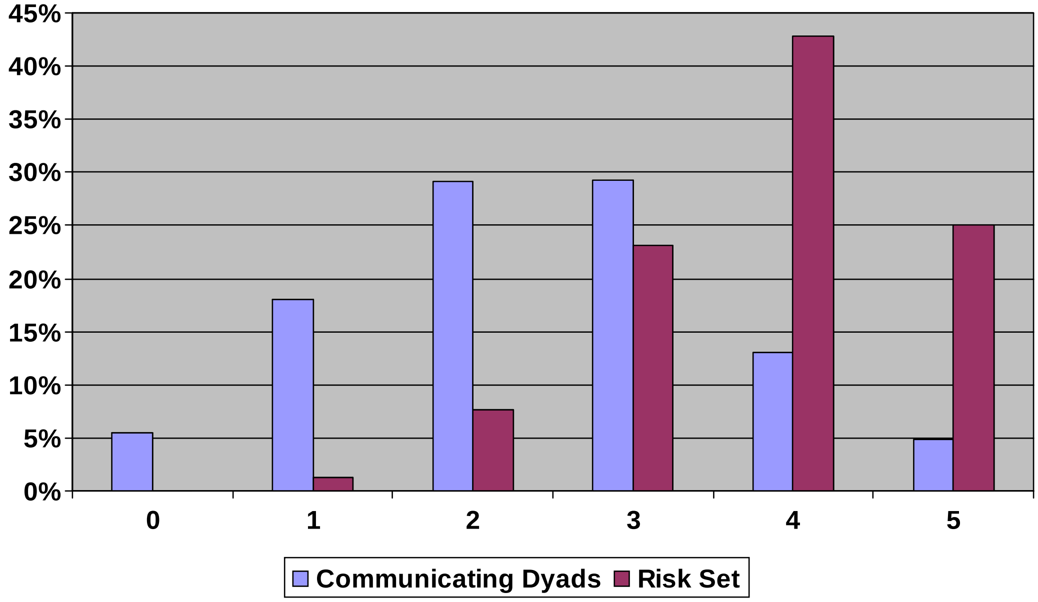

Figure 3: Distribution of communications by the number of boundaries spanned (among business unit, function, salary band, location, and gender) compared to the risk set of possible communication dyads

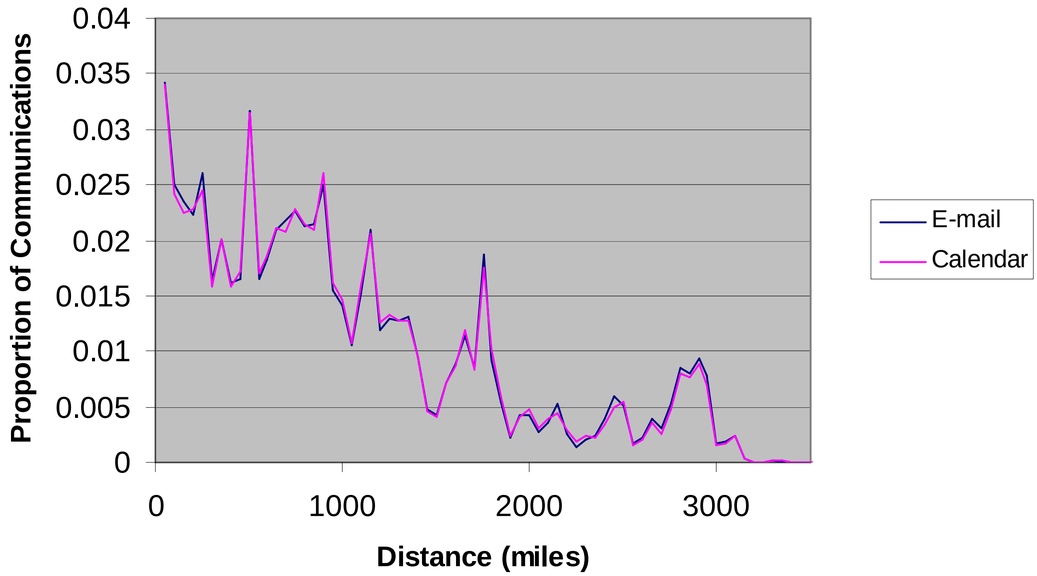

Figure 4: Effect of distance on frequency of dyadic communication by e-mail and meetings.

	BothExec	OneExec	NoExec	FullDataSet
samebu	1.415691	1.925668	1.914878	1.896008
bothchq	*0.1569907*	*-0.1047929*	*0.0407452*	*0.0281654*
onechq	*0.0585527*	0.3400052	0.1040957	0.1332222
samegender	0.3104138	0.0995876	0.1029532	0.117246
bothfemale	*0.1525218*	0.1612418	0.2364077	0.2134689
samefunction	1.208735	1.201079	1.221384	1.213741
sameoffice	1.030925	1.990626	1.990492	1.980202
distance	-0.0001593	-0.000156	-0.0002438	-0.0002348
sameband	-0.3493169		0.2999946	0.3397616
tenurewith~5	*-0.0165058*	0.3647628	0.2553843	0.267812
lnrowsumi	1.162008	1.111796	0.9471679	0.9694572
lnrowsumj	1.268665	1.350679	0.9765275	1.029462
_cons	-20.38919	-22.69379	-18.16086	-18.7356
N	30,982	209,981	2,101,218	2,342,181

Table 2: Poisson regressions of frequency of dyadic e-mail communication on similarity variables for the dyad, run on the full data set and on subsets of the data in which two, one or zero members of the dyad are executives (salary band > 10), respectively. All coefficients are significant at the p<.01 level, except where indicated by italics.

	Coef.	Std. Err.	P>|t|
female	0.094336	0.00988	0
bu_1	1.251549	0.088149	0
bu_2	-0.04014	0.016697	0.016
bu_3	1.231297	0.060905	0
bu_4	0.788773	0.040439	0
bu_5	1.252332	0.092304	0
bu_6	0.968056	0.043468	0
bu_7	0.255184	0.017097	0
bu_8	0.2374	0.063863	0
bu_9	1.28033	0.054886	0
bu_10	1.15432	0.05477	0
bu_11	0.696312	0.093622	0
bu_12	0.845038	0.051228	0
bu_13	-0.18056	0.019892	0
bu_14	0.704902	0.029754	0
bu_15	0.439324	0.032793	0
bu_16	0.323322	0.036569	0
bu_17	0.677626	0.066599	0
bu_18	0.396695	0.049339	0
bu_19	0.981864	0.07497	0
bu_20	0.191187	0.033208	0
bu_21	0.606961	0.113556	0
bu_22	1.041061	0.094848	0
bu_23	1.301686	0.149514	0
bu_24	1.903104	0.179546	0
bu_25	0.683971	0.085755	0
bu_26	0.901484	0.088759	0
bu_27	0.265122	0.150157	0.077
bu_28	0.320633	0.075642	0
bu_29	1.206506	0.111527	0
bu_30	0.780542	0.068209	0
bu_31	-0.04651	0.098927	0.638
function_0	-0.04503	0.095057	0.636
function_1	-0.73452	0.108669	0
function_2	-0.49197	0.097711	0
function_4	-0.64599	0.102993	0
function_5	-0.75868	0.116335	0
function_6	-0.21961	0.1231	0.074
function_7	-0.26011	0.094355	0.006
function_8	-0.2582	0.09945	0.009
function_9	-1.30547	0.095624	0
function_10	-0.40646	0.101234	0
function_11	-0.44324	0.093621	0
function_12	-0.55353	0.094162	0
band_1	-0.40078	0.134342	0.003
band_2	-0.53647	0.315524	0.089
band_3	-0.38896	0.173228	0.025
band_4	-0.25561	0.1437	0.075
band_5	-0.45816	0.170872	0.007
band_6	0.095379	0.128165	0.457
band_7	0.18404	0.125806	0.144
band_8	0.206019	0.125026	0.099
band_9	0.314562	0.124746	0.012
band_10	0.422103	0.124553	0.001
band_11	0.445362	0.125061	0
band_12	0.385837	0.12517	0.002
band_13	0.33416	0.113458	0.003
tenure_q1	-0.03157	0.015551	0.042
tenure_q2	0.063384	0.014787	0
tenure_q3	0.062153	0.013774	0
tenure_q4	-0.01022	0.013359	0.444
rowsum	0.000145	1.01E-05	0
rowsumsq	-1.27E-08	1.83E-09	0
_cons	1.760417	0.079738	0

Table 3: OLS regressions of propensity to communicate across categories on dummies for the levels of the categories. The reference category for functions is function_3, General Executive Management. The reference category for business units is bu_0, corporate headquarters. The reference category for salary bands is band_14, the most senior employees. The reference category for tenure is tenure_q5, the longest-tenured quintile. N = 27,751. R^2 = .4140.

References

Batagelj, Vladimir, and Andrej Mrvar. 2006. "Pajek: Program for Large Network Analysis."

Blau, Peter Michael, and Joseph E. Schwartz. 1984. *Crosscutting Social Circles : Testing a Macrostructural Theory of Intergroup Relations*. Orlando: Academic Press.

Bossard, James H. S. 1932. "Residential Propinquity as a Factor in Marriage Selection." *The American Journal of Sociology* 38:219-224.

Burns, Tom, and G.M. Stalker. 1961. *The Management of Innovation*. London: Tavistock Publications.

Burt, Ronald S. 1992. *Structural Holes : The Social Structure of Competition*. Cambridge, MA: Harvard University Press.

Chandler, Alfred Dupont. 1962. *Strategy and Structure: Chapters in the History of the Industrial Enterprise*. Cambridge, MA: M.I.T. Press.

Collis, David J., and Cynthia A. Montgomery. 1998. "Creating Corporate Advantage." *Harvard Business Review* 76:70-83.

Cosslett, Stephen R. 1981. "Maximum Likelihood Estimator for Choice-Based Samples." *Econometrica* 49:1289-1316.

Festinger, Leon. 1950. *Social Pressures in Informal Groups : A Study of Human Factors in Housing*. New York: Harper.

Galbraith, Jay R. 1973. *Designing Complex Organizations*. Reading, MA: Addison-Wesley.

Goold, Michael, Andrew Campbell, and Marcus Alexander. 1994. *Corporate-Level Strategy: Creating Value in the Multibusiness Company*. New York: J. Wiley.

Gouriéroux, C., A. Monfort, and A. Trognon. 1984. "Pseudo Maximum Likelihood Methods: Theory." *Econometrica* 52:681-700.

Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78:1360-1380.

Guimera, R., L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas. 2006. "The Real Communication Network Behind the Formal Chart: Community Structure in Organizations." *Journal of Economic Behavior & Organization* 61:653-667.

Ibarra, Herminia. 1992. "Homophily and Differential Returns: Sex Differences in Network Structure and Access in an Advertising Firm." *Administrative Science Quarterly* 37:422-447.

Imbens, Guido W. 1992. "An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling." *Econometrica* 60:1187-1214.

Kleinbaum, Adam M. 2006. "Measuring Mail: New Analyses of E-Mail Data for the Study of Cross-Divisional Innovation." in *Academy of Management Best Paper Proceedings*, edited by K. Mark Weaver. Atlanta, GA; (Finalist, Best Paper Award, Technology and Innovation Management Division).

Kleinbaum, Adam M., and Michael L. Tushman. 2007. "Building Bridges: The Social Structure of Interdependent Innovation." *Strategic Entrepreneurship Journal* 1.

Kogut, Bruce, and Udo Zander. 1992. "Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology." *Organization Science* 3:383-397.

Kono, Clifford, Donald Palmer, Roger Friedland, and Matthew Zafonte. 1998. "Lost in Space: The Geography of Corporate Interlocking Directorates." *The American Journal of Sociology* 103:863-911.
Kossinets, Gueorgi, and Duncan J. Watts. 2006. "Empirical Analysis of an Evolving Social Network." *Science* 311:88-90.
Krackhardt, David, and Jeffrey R. Hanson. 1993. "Informal Networks: The Company Behind the Charts." *Harvard Business Review* 71:104-111.
Lancaster, Tony, and Guido Imbens. 1996. "Case-Control Studies with Contaminated Controls." *Journal of Econometrics* 71:145-160.
Lawrence, Paul R., and Jay William Lorsch. 1967. *Organization and Environment: Managing Differentiation and Integration*. Boston: Harvard Business School Press.
Lazarsfeld, P. F., and Robert K. Merton. 1954. "Friendship as Social Process: A Substantive and Methodological Analysis." Pp. 18-66 in *Freedom and Control in Modern Society*.
Marmaros, David, and Bruce Sacerdote. 2004. "How Do Friendships Form?"
Marsden, Peter V. 1988. "Homogeneity in Confiding Relations." *Social Networks* 10:57-76.
Mayo, Elton. 1949. "Hawthorne and the Western Electric Company." in *Organization Theory, Selected Readings*, edited by Derek Salmon Pugh: Penguin Books.
McPherson, J. Miller, and Lynn Smith-Lovin. 1986. "Sex Segregation in Voluntary Associations." *American Sociological Review* 51:61-79.
Nohria, Nitin. 1996. *From the M-Form to the N-Form : Taking Stock of Changes in the Large Industrial Corporation*. Boston, MA: Division of Research Harvard Business School.
Obstfeld, David. 2005. "Social Networks, the *Tertius Iungens* Orientation, and Involvement in Innovation." *Administrative Science Quarterly* 50:100-130.
Onnela, J. P., J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, and A. L. Barabasi. 2007. "Structure and Tie Strengths in Mobile Communication Networks." *Proceedings of the National Academy of Sciences* 104:7332-7336.
Ouchi, William G. 1980. "Markets, Bureaucracies, and Clans." *Administrative Science Quarterly* 25:129-141.
—. 1981. *Theory Z : How American Business Can Meet the Japanese Challenge*. Reading, Mass.: Addison-Wesley.
Powell, Walter W. 1990. "Neither Market nor Hierarchy: Network Forms of Organization." Pp. 295-336 in *Research in Organizational Behavior*, edited by Barry M. Staw and L.L. Cummings. Greenwich, CT: JAI Press.
Quintane, Eric, and Adam M. Kleinbaum. 2007. "Mind over Matter? E-Mail and Survey as Representations of Observed and Perceived Networks." Harvard Business School.
Ridgeway, Cecilia L., and Lynn Smith-Lovin. 1999. "The Gender System and Interaction." *Annual Review of Sociology* 25:191-216.
Simmel, Georg. 1902. *The Sociology of Georg Simmel*. Glencoe, IL: Free Press.
Sorenson, Olav, and Toby E. Stuart. 2001. "Syndication Networks and the Spatial Distribution of Venture Capital Investments." *The American Journal of Sociology* 106:1546-1588.

Zipf, George Kingsley. 1949. *Human Behavior and the Principle of Least Effort; an Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley Press.