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

Adam M. Kleinbaum
Harvard Business School
Boston, MA 02163
akleinbaum@hbs.edu
617.496.6968

Toby E. Stuart
Harvard Business School
Boston, MA 02163
tstuart@hbs.edu
617.496.4626

Michael Tushman
Harvard Business School
Boston, MA 02163
mtushman@hbs.edu
617.495.5442

[•] The authors would like to thank Pierre Azoulay, Mikolaj Piskorski, Bill Simpson, and Julie Wulf for helpful suggestions. We also benefited from the comments of seminar participants at Stanford, MIT, Columbia, Yahoo Research, and Harvard. Please direct correspondence to Adam Kleinbaum (akleinbaum@hbs.edu) or Toby Stuart (tstuart@hbs.edu).

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

Abstract

This is a descriptive study of the structure of communications in a modern organization. We analyze a dataset with millions of electronic mail messages, calendar meetings and teleconferences for many thousands of employees of a single, multidivisional firm during a three-month period in calendar 2006. The basic question we explore asks, what is the role of observable (to us) boundaries between individuals in structuring communications inside the firm? We measure 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 formal organization structure and spatial collocation on the rate of communication. Homophily effects based on sociodemographic categories are much weaker. In individual-level regressions of engagement in category-spanning communication patterns, we find that women, mid- to high-level executives, and members of the executive management, sales and marketing functions are most likely to participate in cross-group communications. In effect, these individuals bridge the lacunae between distant groups in the company's social structure.

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

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

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 enterprise (Chandler, 1977). In these and other theories of the firm, the key managerial task is to effect 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 extant 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 actual interaction patterns in modern, complex, multi-unit firms. To open the proverbial "black box" and begin to reveal the internal wiring of the firm, this paper presents a detailed, descriptive analysis of the network of communications among members of a large, structurally, functionally, geographically, and strategically diverse firm (hereafter, "BigCo"). The full dataset comprises more than 100 million electronic mail messages and over 60 million electronic calendar entries for a sample of more than 30,000 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, job function, and office location.

Because of the importance of boundaries in general theories of social structure and specific theories of organizational behavior, we focus on the role of observable (to us) boundaries between individuals in structuring communications inside the firm. Our data enable us to discern three types of boundaries: organizational unit (e.g., strategic business unit, function membership), 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 frequency of dyadic interactions.

After generating estimates of the effects of boundaries on the frequency of communication, we then flip this analysis on its head: we calculate person-specific measures of the degree to which each individual in the sample engages in communications that *deviate* from the modal pattern of intra- and inter-group interaction in the data. Aggregating across all of an individual's dyadic exchanges, people who score highly on this variable are category spanners – their interactions connect rarely traversed organizational and social groups. Because communication is dense within most categories and sparse between them, category spanners are far more likely to create the "weak" (Granovetter, 1973) and "bridging" (Burt, 1992) ties in the organization. Without these people, BigCo would devolve to a structure of clan-like silos separated by relational voids.

Although we devote considerable analytical attention to the boundaries that mold interactions, the paper's primary aim is descriptive and we refrain from formulating any specific hypotheses. Our theoretical development centers on the potential implications of different theories of organization for likely communication structures within complex organizations, but other than opinions formed from our own anecdotal observations of large organizations, we have no compelling *a priori* reason to give precedence to one theory over another. For this reason, we do not offer predictions. Moreover, at the level at which we can measure the communications network within BigCo, 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 preponderance of the evidence is most consistent with one point of view – classical organization theory's emphasis on formal structure in shaping interaction – we will make no strong claim of proof.

We present too many descriptions of the BigCo communication network to summarize in their entirety here, but a few findings are noteworthy and at least somewhat unexpected. First, relative to men, women participate in a greater volume of electronic and face-to-face interactions and do so with a larger and more diverse set of communication partners. This finding cannot be explained by gender sorting into different work roles, such as secretarial positions (although there is a gendered division of labor within the firm). Second, organizational boundaries – business unit, job function, and office location – have an enormous influence on who interacts with whom inside the firm. As a 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 dyadic distance, a pair of individuals that shares the same business unit, job function, and office location communicates at an estimated rate that is approximately 1,000 times higher. Third, among all employees, executive-level communication appears to be least (but still very heavily) delimited by the pathways of formal organizational structure. By contrast, in interactions amongst themselves, executives show a somewhat stronger tendency toward sociodemographically (gender and tenure) homophilous communication partners than do other employees. Finally, the category spanners in the firm are women concentrated in the upper-middle management ranks and in a few functions, most notably sales, marketing, and general executive management.

Theories of Organization & Communication

The literature offers starkly different possibilities for how communications might weave together a social fabric inside the firm. Weber (1924) 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 (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 many 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 should play a central role in bridging communications among autonomous 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, does complicate the story, classic organization theories nevertheless imply that interaction patterns inside the firm will, for the most part, be dictated by the organization's formal structure. Thus, classic organization theory leads us to expect hierarchical communication patterns that unfold within formal organizational units.

Albeit not extensive, there is some evidence to suggest that formal structure does strongly influence communication patterns within firms. A few of the survey-based studies of intra-firm networks have suggested that an organization's formal structure forms 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 reporting structure. Although not the primary purpose of the paper, a similarly central role of formal structure in shaping interaction patterns is evident in Burt's (2004) analysis of social capital effects in the supply chain function of a large electronics company.

There are, however, competing views. One of organization theory's most taken-for-granted assumptions is that informal structures of power, influence, and information exchange emerge within organizations. These informal structures are thought to significantly influence interaction patterns and, indeed, the "informal" organizational chart is often held to be more consequential than the formal one (Mayo, 1949;Krackhardt, and Hanson, 1993). Moreover, some of the most prominent students of organizations have viewed structure and action as being only loosely coupled (Weick, 1976). In Cohen, March and Olsen's (1972) garbage can model, for example, people, problems, and solutions admix by chance, and organizational action can appear almost random relative to formally prescribed decision hierarchies.

If communication patterns map to informal structure, the relevant 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 (Lazarsfeld, and Merton, 1954;Blau, and Schwartz, 1984), we might expect to find that communication is much more prevalent within sociodemographic categories. For instance, within formal organizations and in social networks more generally, there is considerable evidence of gender homophily: many individuals are immersed in networks comprising primarily same-sex ties (e.g., McPherson, and Smith-Lovin, 1986;Marsden, 1988;Ibarra, 1992;Ridgeway, and Smith-Lovin, 1999). Similarly, there is evidence that social ties cluster within age-based and tenure-based strata (e.g., Zenger, and Lawrence, 1989). Thus, if the communication pathways within organizations are informally structured, we might anticipate that similarities along sociodemographic dimensions have a primary influence on who communicates with whom.

If the formal and informal structure perspectives suggest, respectively, that organizational or sociodemographic boundaries will exert a first-order influence on interaction patterns within an organization, a third point of view – represented in the literature under rubrics such as the "boundaryless" firm, the "networked" organization (Powell, 1990;Nohria, 1996), and the knowledge-based firm (Kogut, and Zander, 1992) – reflect a contemporary image of organizations characterized by free-flowing, lateral, and collaborative interaction patterns. In this perspective, neither rigid hierarchies nor social categories necessarily play a dominant role in orchestrating intra-organizational interaction. One might trace the origins of these perspectives to Burns and Stalker's (1961) distinction between mechanistic and organic organizations. The notion of more organic organization structures gained momentum 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 the mechanism of governance.

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 cost and improving the quality of cross-geography, cross-social group interaction, have finally enabled the enactment of lateral, non-hierarchical interaction structures in complex organizations. For instance, Weisband (2008) argues that the technological advances of the past decade have vindicated the forward-looking vision of a networked organizational form that light-handedly coordinates the activities of heterogeneous, geographically diverse team members. In an environment of sophisticated, enterprise-wide information technology systems, myriad "web 2.0" collaboration tools, and massive bandwidth, the technology certainly exists to facilitate lateral organizational forms.

Thus, our survey of the extant literature suggests three very different bases for interaction within organizations. In classical work in organization theory and some contemporary theorizing it has inspired, formal structure reigns supreme; in more behaviorally oriented work with roots in mid-century sociology and social psychology, informal structure occupies a central position; and in a more recent stream of the literature, the image is one of a federation of organizational members woven together in lateral and fluid communication structures. We submit that the question of interest is one of degree: few proponents of the views that formal or informal structures would go so far as to argue that either one operates wholly at the expense of the other, and those who foresee the emergence of the boundaryless organization recognize that this metaphor represents the far end of a continuum rather than a present-day reality. As we see it, therefore, the relevant empirical question – and the one we hope to illuminate – is, to what extent do communication patterns map to formal organization structures, versus emerge organically in a manner that is unfettered by the proscribed authority structures of the organization, or by the geographic and organizational locations of members? Following reviews of a few recent studies with objectives similar to ours, we turn to a description and analysis of the BigCo electronic communication network.

Recent Literature

A handful of recent studies, including Kossinets and Watts (2006;2008), Marmaros and Sacerdote (2004), Guimera et al. (2006) and Tyler et al. (2003), have exploited the availability of large electronic communications databases to examine the emergence or the shape of intra-organizational social structures. Two of these articles, Guimera et al. (2006) and Tyler et al. (2003), are conducted at the network level: they develop algorithms to identify clusters of densely interacting individuals, in the former case within a university community and in the latter, within Hewlett Packard's research labs. We too will describe the communication network at BigCo at the group and individual levels, but like Kossinets and Watts (2008) and Marmaros and Sacerdote (2004), the bulk of our analysis is conducted at the dyad level. These two papers specifically analyze the emergence or intensity of communication in datasets of at-risk dyads on university campuses.

Both Kossinets and Watts (2008) and Marmaros and Sacerdote (2004) find that sociodemographic proximity undergirds tie formation, and the latter show that even on a contiguous, single-campus university campus (Dartmouth), geographic proximity (measured at the floor and residence hall levels) is enormously consequential for friendship formation. We too find effects of geographic space and sociodemographic proximity on pairwise interaction probabilities at BigCo. However, the differences between our analysis and these predecessors loom as large as the similarities. Most important, our study is situated in a large, multi-location, for-profit company, and the (probably vast) majority of the millions of communications in the data are work-related. By contrast, Kossinets and Watts and Marmaros and Sacerdote examine e-mail networks on geographically compact university campuses, in which a much larger share of interactions are likely to be friendship ties rather than task relations. Therefore, we see our study as joining a small group of other papers in establishing a baseline picture of the structure of interaction in different types of current-day organizational and social communities.

Before describing the data in further detail, it merits note that a few studies have also established that e-mail network data reliably can be used to proxy for physical-world social networks in organizational settings. Quintane and Kleinbaum (Quintane, and Kleinbaum, 2008) show that e-mail data corresponds to survey reports of work networks at least as closely as do other sources of observational data (e.g., ethnographic observation of inter-personal interactions). Marmaros and Sacerdote (2004) also conduct a survey that demonstrates a close correspondence between e-mail and self-reported friendship networks among college students. Similarly, we will show that the BigCo e-mail network very much runs parallel to the face-to-face work and social relationships in the company. In fact, the extraordinarily high similarity between the e-mail and face-to-face network at BigCo is a striking finding in its own right.

Data, Variables and Methods

BigCo is a large information technology and electronics company with 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 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].

The data we analyze include the complete record, as drawn from the firm's servers, of e-mail communications and scheduled meetings (both face-to-face and conference calls) among 30,328 people during 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 older messages locally). All internal calendar and e-mail information that was on the server at the time of data collection was included in our sample.

BigCo provided the data in the form of 30,328 text files, each representing the communication activity of a single person, which we cleaned and parsed. 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 senders and recipients were then replaced with hashed identifiers. We then consolidated these files, split them according to communication type, and expanded each multiple-recipient message or recurring meeting to include one entry for each unique dyad and each occurrence. The final files contain 114 million e-mails and 68 million meetings.

We focus our analyses on e-mails that are sent to four or few recipients. In the core models, we exclude sender-to-BCC pairs, mass mailings[2] and direct interactions with

[1] Law protecting employee privacy precludes data collection for a project like this in most of Europe and parts of Asia, so the company was only able to provide data for U.S.-based employees.

administrative assistants. Imposing these screens shrinks the data set by almost an order of magnitude to 13 million e-mails and 3 million meetings.[3] (In Appendix 2, we show that findings are very consistent with those we report in the main text when we change these assumptions.) The mean employee in our sample exchanged 1,178 non-mass, non-BCC e-mails with 93 other members of the sample during the observation period (a median count of 3 and a mean count of 12.9 within each communicating dyad), as well as 866 e-mails with employees not in our sample. These distributions have long right tails: the maximum number of correspondents was 2,097 and the maximum number of e-mails was over 20,000.

BigCo also provided demographic and HR information about each employee, which we are able to link to the communication data through encrypted employee identifiers. The HR data include each employee's business unit, major job function, job sub-function, firm tenure, salary band, state, location code[4] and gender.

Our sample contains 24% of the firm's U.S. employee population and was collected through a snowball sampling procedure. Our initial point of entry into the organization was the

---

[2] We define "mass" as messages or meetings with more than 4 recipients/participants. Consistent with results reported in Quintane and Kleinbaum (2008), 83% of e-mails in our data set have 4 or fewer recipients.
[3] Of the original 114 million dyadic e-mails, 31 million involved a person either outside the United States or otherwise not included in the sample, and about whom we have no demographic data; 3.5 million involved an administrative assistant, 64 million were mass mails (i.e., they included more than four recipients; mass mailings represent just 17% of total e-mails sent but they are a much larger faction—just over 50%—of pairwise exchanges based on the expansion of the message to include all sender-to-recipient-list ties); and 1.2 million were BCCs (in these instances, we retain the message but do not treat the sender-to-BCC recipient as a realized tie). The total number of dyadic meetings is, depending on assumptions, very large. Unlike e-mail, a meeting involves interactions not only between sender and each recipient but also among all recipients, so the size of the data set grows exponentially with the number of participants in a meeting, not geometrically. However, if we consider the 68 million dyadic meetings involving 10 or fewer people to be our starting point, 8 million involved a person outside the U.S. or otherwise not included in the sample, and about whom we have no demographic data; 1.4 million involved an administrative assistant, 56 million were large meetings (more than 4 recipients); and 80,000 were "FYI notifications" about a meeting (i.e., BCCs).
[4] For privacy, the firm provided an encrypted location code that allows us to determine whether two employees are in the same location, but does not allow us to pinpoint the specific location of any employee. However, BigCo provided an inter-office distance matrix for all 289 office location codes in the dataset to allow us to measure the exact geographic separation between all pairs of employees in the dataset.

corporate sales force: 180 people involved in cross-divisional projects were identified and invited to participate in the study. Although all projects were organized by the sales force, because all were cross-divisional, only half of the 180 invitees hailed from corporate sales. The other half were distributed across the full range of BigCo's other business units. Of the total group, 91 individuals agreed to participate, 25 of whom we were unable to include because they worked outside the U.S. In turn, the remaining 66 individuals communicated with an additional 30,262 U.S. employees during the preceding three months. The company then provided *complete* e-mail and calendar data for *all* 30,328 U.S.-based employees (the 66 core members plus their 30,262 direct contacts).

Although this is not a simple random sample, it is compellingly large. Moreover, the randomness of the sample is improved by mass mailings, which serve to cast a wide net in drawing other employees into the sample. To cite an example, one of the 66 individuals in the original sample received an e-mail that was sent to him and 1,214 co-recipients. This e-mail alone accounts for over 4% of our full sample because the sampling procedure sweeps the complete e-mail records of all 1,214 co-recipients into our dataset. More generally, the 66 individuals in the original sample possess an average of 3,415 direct contacts in the full sample, although just 137 of these contacts, on average, appear to represent actual direct contacts (i.e., the focal individual and the partner exchanged one or more non-mass, non-BCC communication during the observation period).

The possibility remains, though, that use of the full sample could produce findings that are biased in unknowable ways relative to the true patterns of interaction in BigCo. To guard against the risk that our findings are driven by the non-random sampling procedure, we use our knowledge of the firm's population of U.S.-based employees to create a near-random sub-sample of employees. In effect, we develop a sub-sampling algorithm that maximizes the correspondence between the sub-samples we draw and a set of population parameters along the dimensions that are likely to be most meaningful in our analysis. (Because this algorithm is somewhat complex, we provide details of the sub-sampling procedure in Appendix 1.) Unless otherwise noted, the analyses we will present will be based on the more conservative random sub-sample, but as we show in Appendix 2, the findings do not substantively change in the full sample.

DYAD-LEVEL COMMUNICATION RATES

After cleaning and parsing the data, we collapsed them into a single cross-section and created counts at the dyad level of the total number of $i \leftrightarrow j$ messages, where i and j index all individuals in the sample. In other words, we constructed a cross sectional dataset with counts of the number of communications within 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. Even with the time axis compressed to treat the data as a cross-section, the communication matrices are large and sparse – less than 0.3% of the approximately 112 million possible unordered cells in the sub-sample e-mail matrix are non-zero. Likewise, only 0.12% of cells in the unordered meeting matrix are non-zero. Given the computing power available to us, it is not expeditious to work with the full matrix.

Random sampling from the set of the 112 million communicating dyads is one potential solution to this problem. However, this approach 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 (King, and Zeng, 2001), drawn at a 1:1, actual-tie : zero-cell, ratio, which are then weighted according to their probability of being drawn into the analysis sample[5]. We do not stratify on the sampling of zeros; we simply draw the zero cells at random.

Our dependent variable is a count of the number of e-mails (or, in separate analyses, meetings) exchanged within each dyad. To accommodate the case cohort data structure, we use a weighted quasi-maximum likelihood (QML) Poisson model. 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; no assumptions about the distribution of the data are required[6] (Gouriéroux, et al., 1984;Wooldridge, 1997;Silva, and Tenreyro, 2006). Moreover, robust standard errors also are consistent even if the mean is mis-specified. Thus, we estimate the likelihood that dyad-level covariates affect the frequency of interaction using models of the form:

$$\mathrm{E}\left[Y_{ij} \middle| X_{ij}\right] = \exp\left((X_{ij} + Z_{ij})\beta\right) \qquad (1)$$

where Y_{ij} is the count of e-mails exchanged (in both directions) between individuals i and j, X_{ij} is a vector of pair-level covariates, Z_{ij} a vector of control variables, and β a vector of regression coefficients.

[5] Our results are robust to alternate case cohort datasets with 5:1 and 10:1 ratios of zeros to non-zeros.
[6] In particular, unlike the maximum likelihood Poisson estimation, Poisson QML estimation does *not* assume that the data are distributed with the mean equal to the variance of the event count. Unless the data are known to be distributed negative binomial, Poisson QML estimation is preferable because it is consistent even if the data are negative binomial and its assumption is limited to the conditional mean of the data.

Dyad-level Variables

The independent variables in the dyad-level regressions are all properties of the *ij*th pair of employees. Of primary interest in our analysis is a set of dummy variables that indicate whether or not two individuals, employees *i* and *j*, share the same affiliation across six different organizational and social groups. First, we include *SameBU*, defined to be one when *i* and *j* are in the same strategic business unit. BigCo has 31 business units, 29 of which are organized into four business groups; the remaining two are corporate headquarters and the corporate sales force, both of which are treated as business units by the company and in our data. Given the centrality of corporate headquarters in theories of the multi-business firm, we include two additional dummy variables to reflect membership in HQ. *BothCHQ* is defined to be one when the two individuals in a dyad are both members of corporate headquarters. This variable can be considered as an interaction between *SameBU* and "business unit = corporate headquarters". We also include *OneCHQ*, which has value one when exactly one member of the dyad is in the headquarters unit; *OneCHQ* can be considered as an interaction between "not *SameBU*" and "*i* or *j* is in headquarters". Similarly, we include dummy variables for *SameGender* and *BothFemale*, where the latter covariate is equivalent to an interaction between *SameGender* and "gender = female".

We include a *SameFunction* dummy variable to indicate whether employees *i* and *j* are in the same job function. BigCo classifies each employee in one of 13 different job functions: administration (consisting primarily of secretaries and other support personnel), communications, finance, general executive management, human resources, legal, manufacturing, marketing, research & development, supply chain, sales, services and a catch-all “other” category. These 13 job functions are further sub-divided into 60 subfunctions, which we account for in the regressions with a *SameSubfunction* dummy variable.

Employees in our sample work in 289 offices scattered across all 50 U.S. states. We include a *SameOffice* dummy to indicate pairs of actors who are physically located in the same building. We also include *logDistance*, the logarithm of the estimated door-to-door (driving) distance between offices, plus one mile.

The company has a 15-band salary hierarchy ranging from 0 (for employees in training) to 14. Members of bands 7-10 are considered to be middle managers and those 11 and above are considered to be executives. We include a *SameBand* dummy variable to indicate that both members of a dyad are in the same salary band. Finally, we include an indicator *TenureWithin5*, a dummy variable set to one when the absolute value of the difference between *i*’s and *j*’s tenure with the company is less than or equal to five years.

We include two control variables to absorb dyad-level heterogeneity. First, *logInSample* is the natural logarithm of one plus the number of e-mails the two actors exchanged with all other (non-*i*-*j*) partners in our sample. One can think of this covariate as serving the function of a dyad-level fixed effect. By including it, we condition on the total count of individual *i*’s plus *j*’s e-mails. After conditioning on the total e-mail count, the variance remaining to identify the other regression parameters will relate to the distribution of communications across potential partners, rather than being driven by the overall communications volume of the two actors in a dyad. Likewise, we include *logOutOfSample*, the natural logarithm of one plus the number of e-mails the two actors sent or received from other employees of BigCo that are not in our sample. This covariate adjusts for the fact that the individuals within the sample have differential propensities to communicate beyond it.[7]

Finally, for all categories in the regressions, we control for the combined sizes of the groups to which the members of the dyad belong. These group sizes define the risk set of possible local and cross-group communication partners. For instance, we include *logAvgBUSize*, the log of the average number of people in the sample in the business units to which i and j belong. In general, when i and j are members of large business units (or other large groups), the probability that they specifically will interact will decline because of the large number of available substitute communication partners (assuming that individuals' interaction frequencies do not scale proportionately with group size). We include similar, log-scaled group size controls for function, subfunction, office, salary band, and the number of people within five years of the firm tenure of individual i or j.

BOUNDARY SPANNING (PERSON-LEVEL) ANALYSES

After examining factors that influence dyadic interaction rates, we then analyze, at the individual level, the correlates of a category-spanning communications profile. To conduct this analysis, we construct two person-level dependent variables, each rooted in different assumptions about the nature of intra-organizational boundaries. The simplest measure of category-spanning would be to take the set of each individual i's e-mail exchanges with all communication partners and then to compute over this set the number of categories spanned in the average message. For instance, an e-mail from a male employee in business unit 22 to a

---

[7] Based on the way the sample was constructed, we have all within-firm e-mail communications for the original 66 members of the sample and all of their communication partners. However, we possess demographic data only for these individuals and their 30,262 direct, U.S.-based partners. Among these 30,328, we know of all messages sent and received outside of the U.S. and outside of the sample, but we cannot characterize the identities of alters beyond the 30,328.

female employee in business unit 29 would span two categories, gender and business unit, if all other group memberships held by these two individuals (function, sub-function, office location, tenure, and salary band) were shared.

The shortcoming of this approach, however, is that it fails to adjust for the fact that some categories in the data, for example, office location and business unit, will prove to have a *much* larger influence on communications in BigCo than do others, such as gender and tenure. Simply counting or averaging the number of spanned categories fails to account for the relative impedance of the six boundaries we measure on the observed communication frequencies within the company. Because the aim of our analysis of category spanning is to gain insight into who bridges the infrequently connected groups in the firm, we will need to account for the baseline level of interaction between groups in generating the measure of category spanning.

To identify the individuals in BigCo who are responsible for coordinating between groups that rarely interact, we create a variable, Improb_i , which gauges the *improbability* of each individual's overall communication profile, where high improbabilities occur when an employee frequently spans a specific group boundary that is rarely crossed in the company. By definition, individuals with high, composite improbability scores disproportionately form linkages between groups that communicate infrequently. To construct Improb_i, we begin by creating matrices, one for each category (BU, function, subfunction, gender, office, salary band), with elements defined as the observed probabilities that members of group x communicate with those in group y. For example, because there are 31 business units at BigCo, we construct a 31x31 matrix $\text{Pr}BU$, in which the xyth cell of the matrix is the probability that members of business unit x exchange e-mail with those in business unit y. After producing similar matrices for all six categories, we can then calculate, for all actually communicating dyads, the "improbability" that employees *i* and *j* will communicate as:

$$\text{Improb}_{ij} = 1 - \left[\Pr BU_{ij} \times \Pr fx_{ij} \times \Pr subfx_{ij} \times \Pr office_{ij} \times \Pr gender_{ij} \times \Pr band_{ij}\right] \qquad (2)$$

where each of the Pr__ variables reflect the actual incidences of communication at BigCo between two members of the specific pairs of business units, functions, … represented in each *ij* dyad. Improb_{ij} is computed for each communicating dyad and it assumes its greatest value when the *ij*th dyad**,** given *i* and *j*'s group affiliations and the actual interaction frequencies in BigCo, have group membership profiles that make them least likely to communicate. Two specific examples from the data—one low, one high—may help to illustrate the measure. In one dyad with a very high 55% probability (and hence low improbability) of communication, the two members of the pairing are both in the research software group, both work at corporate headquarters, they are in adjacent salary bands, and they work in the same office building in New York. More than half of the dyads that share these characteristics are live communication links, and this is one of the highest baseline probabilities of interaction among all pairings of group memberships in the data. Thus, Improb_{ij} is a low 0.45 (= 1 - 0.55) for this pairing. By contrast, a second dyad that actually communicates in the data has just a 0.0001% baseline probability of interacting—one of the lowest in the dataset. One person in this dyad is in the General Executive Management function, the other in sales; one is an executive (salary band 14), the other a middle manager (salary band 10); they work in offices on opposite coasts; and they are in different business units. The only group membership they share is gender. This is an extremely improbable pairing, with a high Improb_{ij} value of 0.9999 (= 1 – 0.0001). These contrasting examples illustrate the intuition of the measure: the first dyad spans only two categories, gender and salary band, and communication across these two boundaries is relatively common at BigCo. The second dyad represents a link that jumps four levels in the salary distribution, crosses the geographic expanse of the country, and spans functional and divisional boundaries. It connects two individuals who are highly unlikely to interact, and thus represents a bridging tie.

To move from the e-mail dyad to the person level, for each employee *i*, we then take the e-mail-volume-weighted average across all alters *j* to get Improb_i, the average improbability of *i*'s overall communication pattern:

$$\text{Improb}_i = \frac{\sum_{j=1}^{n_i} \left(\text{Improb}_{ij} \times \text{Freq}_{ij} \right)}{n_i \times \text{Freq}_{ij}} \tag{3}$$

where, for each focal actor *i*, n_i indexes the total number of other individuals *j* with whom *i* communicates and Freq_{ij} measures the number of e-mails exchanged between *i* and *j*. The higher the value of Improb_i, the more actor *i* communicates across categories that are only infrequently spanned at BigCo.

As a second measure of boundary spanning and, for sake of comparison to the existing literature, we also calculate Burt's (1992) structural constraint measure. An actor's structural constraint, C_i , is the sum, across all alters *j*, of c_{ij}:

$$c_{ij} = (p_{ij} + \sum_{q \neq i,j}^{N} p_{iq} p_{qj})^2 \tag{4}$$

where p_{ij} represents the "proportional strength network," or the proportion of actor *i*'s time invested in actor *j*. We calculate the constraint measure using the algorithm implemented in Pajek (Batagelj, and Mrvar, 2006).

Structural constraint conceptually differs from the improbability measure of equation (3) because it is purely network-based: it increases when an individual possesses just a few direct contacts and when those contacts are densely connected among themselves. The measure is agnostic to the category memberships of an individual's e-mail partners. By contrast, Improb_i assesses the degree to which ego's contacts are separated by social, geographic or organizational boundaries. There is good reason to expect these measures to be correlated—if individuals concentrate their interactions within organizational and sociodemographic groups, then those who communicate across groups likely will exhibit low levels of constraint arising from two-step connections among their direct ties.

Empirically, however, in the BigCo e-mail network, constraint (equ. 4) and improbability (equ. 3) appear to gauge somewhat different characteristics of actors' communication networks. Specifically, Burt's constraint measure in this company's e-mail network is driven almost entirely by the P_i term in equ. 4, which is jointly determined by ego's number of direct contacts and the distribution of his or her network energy across these ties. In fact, there is a .93 correlation between overall constraint, C_i, and P_i in the BigCo e-mail network. In other words, Burt's (1992) constraint measure depends almost entirely on an employee's direct ties; it is affected only on the margin by the actual pattern of (two-step) ties among a focal actor's direct contacts. Because of this, we believe that looking at direct ties through a different lens – i.e., the degree to which they span important social and organizational boundaries – may be a preferred approach to identifying the individuals who play a vital coordinating role in the firm. Nevertheless, we do assess the correlates of C_i and compare them to those of the category-spanning covariate.

PERSON-LEVEL ESTIMATIONS

Constraint is a continuous dependent variable and we model it in an OLS framework. $Improb_i$, however, is bounded on the [0 ,1] interval. Because OLS estimation is biased and inconsistent when the outcome variable is a proportion, we estimate fractional logit models[8] (Papke, and Wooldridge, 1996) of $Improb_i$.

In analyzing the correlates of category-spanning network positions, our interest is in learning whether an individual's group memberships (e.g., gender, salary band level) associate with the extent to which they serve as bridges between otherwise infrequently connected groups in the company. There is, however, a complication with regressing $Improb_i$ on category memberships: its realization may be influenced by the demographics of the population across categories. To see this, consider the case of gender. Women comprise 30 percent of the analysis sample. Because women are a smaller group than men, it is quite likely (although not axiomatic) that women will communicate across the gender divide more frequently than will men.[9] If we were to include a gender dummy variable on the right hand side of the regressions while also factoring in the actor's gender in influencing the improbability of his/her communications profile, we may discover that women are more likely to span categories simply because they are members of a numerical minority group. To avoid this problem, we construct four different dependent variables of category-spanning, one for each of the four categories (gender, salary band, business unit, and function) we include in the regressions. Specifically, we calculate:

[8] The distribution of $Improb_i$ is concentrated near 1. Even though post-estimation diagnostics suggest that the models fit well, for robustness we also estimate fractional logit models on linear and non-linear transformations of $Improb_i$ and OLS models using $\ln(1\text{-}Improb_i)$ as the dependent variable. Results are substantively similar across this range of models.
[9] To illustrate this, simply assume that women and men both send an average of ten e-mails and that recipients are random with respect to gender. In this scenario, seven of ten e-mails sent by women cross the gender boundary, versus three of ten sent by men.

$$\text{Improb - exBU}_{ij} = 1 - \left\lfloor \Pr Func_{ij} \times \Pr Subfunc_{ij} \times \Pr Office_{ij} \times \Pr Gender_{ij} \times \Pr Band_{ij} \right\rfloor \quad (7)$$

$$\text{Improb - exFunc}_{ij} = 1 - \left\lfloor \Pr BU_{ij} \times \Pr Subfunc_{ij} \times \Pr Office_{ij} \times \Pr Gender_{ij} \times \Pr Band_{ij} \right\rfloor \quad (8)$$

$$\text{Improb - exGender}_{ij} = 1 - \left\lfloor \Pr BU_{ij} \times \Pr Func_{ij} \times \Pr Subfunc_{ij} \times \Pr Office_{ij} \times \Pr Band_{ij} \right\rfloor \quad (9)$$

$$\text{Improb - exBand}_{ij} = 1 - \left\lfloor \Pr BU_{ij} \times \Pr Func_{ij} \times \Pr Subfunc_{ij} \times \Pr Office_{ij} \times \Pr Gender_{ij} \right\rfloor \quad (10)$$

We then estimate the coefficient of *female* in regressions whose dependent variable, e.g., Improb-exGender$_i$, is not directly affected by the gender distribution of the sample. Thus, we estimate the effect of *gender* on the likelihood that an individual's communications profile will span atypically connected categories *other than gender*; the effect of business unit membership on spanning all categories other than business unit, and so forth.[10]

Results

We begin with descriptive cuts of the data. In total, 44% of e-mails at BigCo are addressed to a single recipient, which is significantly lower than the 82% Kossinets and Watts (2006) report for the university they analyze. This difference may reflect the more team-based nature of work in companies like BigCo relative to a university setting, as well as a higher proportion of work-related versus social communications in the two respective contexts. The distribution of the number of recipients per e-mail (not shown) is highly skewed, with a very long right tail: the median number of recipients is 2 and the maximum number of recipients is 1,214. 83% of e-mails are non-mass communications, which we define to be those addressed to four or fewer recipients.

[10] Subfunction is a fifth category that we incorporate in calculating the value of Improb(i), but because we do not include subfunction dummy variables in the category-spanning regressions, it is not necessary to compute a variant of the dependent variable that excludes subfunction.

Table 1 provides summary statistics for communication patterns in BigCo, broken out by three categories: men and women; employees in corporate headquarters versus the line organization; and executives relative to middle managers and rank-and-file employees. The gender differences in communications activity are pronounced and run counter to what we would have expected based on documented gender differences of network composition in other types of data, such as the GSS (Marsden, 1987), survey data inside of firms (Podolny, and Baron, 1997), and academic collaboration networks (Ding, et al., 2006). As shown in Panel A in Table 1, women interact with more communication partners than do men (96 versus 80) and they e-mail at a higher rate with each partner. Moreover, on every reported measure of partner diversity, women exhibit broader-reaching communication profiles than do men: their e-mail partners are more geographically distant and they are slightly further away in the salary distribution. Likewise, relative to men, a greater proportion of women's communication activity spans the boundary of their organizational functions, strategic business unit, and office. Finally, women's extra-group communication partners are, on the whole, more diverse, as measured by Herfindahls of the diversity of their interaction patterns across partners' group memberships.[11]

Two possibilities may alter the interpretation of the gender differences in Table 1. First, it is well documented that there is a gendered division of labor in the workforce and that men and women are sorted into different work roles within organizations (Baron, and Bielby, 1980). In consequence, women may e-mail more actively and do so more broadly because of the demands of their work roles within BigCo. In fact, this is the case. Women are more likely than men to

[11] The Herfindahl index is a summary measure of the dispersion of a variable (in this case, communication partners' group affiliations) across a set of categories; larger values indicate that alters are relatively concentrated across a small number of groups whereas smaller values indicate that alters are more evenly spread across a large number of groups. We normalize our Herfindahls to range from 0 to 1, rather than the standard 1/N to 1, for greater comparability across variables.

occupy administrative assistant positions and other support roles; two-thirds of the administrative support personnel in the sample are female, which is more than two times the number we would expect if the position were gender neutral. Moreover, individuals in administrative support roles in fact are heavier-than-average users of e-mail. However, the data reported in Table 1 already exclude all administrative support personnel; when we include them, the male-female differences become even more pronounced. These data are shown in Table 1 in square brackets. To compute these quantities, we retain in the data individuals in administrative support roles and calculate averages for number of unique communication partners and number of e-mails. In both instances, the difference in male-female counts increases. Subsequently, we will show that the gender difference in the breadth of communication persists in multivariate models that more fully account for organizational location and work role.

A second possibility is that there are gender differences in usage of communication media – women may e-mail more frequently than men, but do so at the expense of other forms of interaction. To explore this possibility, Panel B in Table 1 reports gender differences in communication profiles based on calendar data that record face-to-face meetings and scheduled telephone calls. Although the indicators of the organizational diversity of communication partners are less systematically different, women nonetheless have a greater number of unique contacts (45 versus 41) and a larger number of scheduled meetings. Thus, we believe that the gender differences in Table 1 are real – they reflect actual differences in the networks of women and men at BigCo.

We also break out employees in corporate headquarters (CHQ) to begin to examine whether CHQ plays an active role in coordinating across organizational groups. 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 (Goold, et al., 1994;Collis, and Montgomery, 1998), the coordinating activities of the headquarters unit are at the center of theories of value creation, and more fundamentally, of normative (from a shareholder value maximization perspective) accounts of the desirability of multi-business firms.

At BigCo, the data indicate slightly broader communications profiles for members of headquarters. Relative to the rest of the sample, employees in CHQ have a greater number of unique e-mail partners; their e-mails traverse larger geographic distances; and they have a lower proportion of intra-business-unit (defined as within CHQ) and intra-office communications. However, CHQ interactions are slightly more heavily concentrated within functions than are those of other employees.

The relationship between e-mail activity and hierarchical level is striking; the average executive (members of the top four salary bands) in our sample sent and received more than twice as many e-mails as the average middle manager who, in turn, sent and received more than twice as many as the average rank-and-file employee. Likewise, executives communicated with many more partners than middle managers or rank-and-file employees and their e-mails were substantially more dispersed across business unit, function, and office boundaries.

Tables 2a-2c illustrate the average volumes of e-mail interaction across, respectively, salary bands, business groups, and functions. Each cell in these matrices is the total number of e-mails sent and received between members of the categories represented on the row and column, denominated by the total number of people in the respective row and column categories. Thus, the matrices are symmetric and each cell can be interpreted as the per-person average number of e-mails sent and received between members of the categories on the row and column. The shading in the tables highlights the highest 20% of cell values.

All three panels in Table 2 reflect one of the very pronounced patterns evident in Table 1: a great deal of interaction takes place within groups, which is manifest in the comparison between main diagonal and off-diagonal cells. Consider Table 2a. In aggregate, BigCo employees communicate much more frequently with others in the same or similar salary bands – the cells on the main diagonal show that within-band interactions are much larger than the off-diagonal cells. Conversely, the cells on the northeast perimeter of the matrix, which pair high to low salary bands, demonstrate that communication between these groups is virtually non-existent. For example, the average number of e-mails sent and received during the sample period among two members of band 14 is 332; by contrast, and average is 0.01 messages for pairings of band 14 members with those of band 5. Overall, Table 2a reveals a block of interaction among executives (bands 11-14) and among middle managers (bands 7-10), with some indication that the members of bands 10 and 11 bridge the two groups. In other words, the dominant pattern in the data is to communicate within level and between adjacent levels, implying a hierarchical communication pattern with virtually no open pathways of interaction between high and low pay grades. Directives from high ranking employees appear to work their way down the chain of command by steps of one or two levels. Furthermore, the final two columns in Table 2a indicate that the highest pay band is also the one that exhibits the largest proportion of within-band communications; 29% of band 14 e-mails are with colleagues in the same pay grade.

Table 2b also presents average communication frequencies, but for intra- and inter-business group interactions. The company operates four primary business groups (software, hardware, business services, and technology services), which together contain 29 business units.

We have added to the matrix the two business units that sit outside the four business groups: the corporate sales force and corporate headquarters (both treated as "business units" by the company, but neither belonging to a business group). The sales function of the company is highly centralized, with a very large corporate sales force that sells the products of all the company's business units. The organizational design of BigCo presumes that both CHQ and corporate sales play important coordinating roles within the firm.

Once again, the values on the main diagonal dwarf those to its side. In fact, the largest off-diagonal cell, which is between technology services and corporate headquarters, is less than one fifth the size of the smallest on-diagonal cell (that for the corporate sales group). However, consistent with their intended roles, members of headquarters and the corporate sales force do communicate more broadly across the business groups than do members of the business groups themselves. To see this, bold cells in the matrix highlight all values above the median in the table (while the shaded cells continue to represent the top quintile). The highest concentration of bold cells is on the rows/columns for CHQ and sales.

Panel C in the table reports per-person average communication volumes between the functions in the company. Here too, the largest cell values are on the main diagonal, and no off-diagonal cell approaches the magnitude of within-group communications. And among functions, once again we see that employees in sales play the greatest role in coordinating across other functions.[12]

[12] Every employee at BigCo is assigned to a business unit, function, and sub-function. The sales function includes both those members of the corporate sales force (treated as its own business unit) who are actively engaged in selling and salespeople in the product divisions. Of note for the multivariate analyses, many members of the corporate sales business unit are not in the sales function, so it is possible to independently identify effects for membership in the corporate sales business unit and the sales function.

It is evident from the panels in Tables 1 and 2 that communications within BigCo are very much structured by organizational and sociodemographic categories. But here, the question arises: compared to what? One potential comparison can be found in the calculation of Newman's (2003) assortativity coefficient, r, which is a univariate statistic designed to gauge the degree to which relationships in a network cluster within groups. This quantity is defined:

$$\mathrm{r} = \frac{\sum_{\mathrm{i}} e_{\mathrm{ii}} - \sum_{i} a_i b_i}{1 - \sum_{i} a_i b_i} \tag{12}$$

where e_{ij} is the proportion of all ties in the network that link members of group i to group j; a_i is a vector of row sums of the **e** matrix or, equivalently, a vector of the proportions of all ties sent by each group i; and b_i is a vector of column sums of the **e** matrix, or the proportion of all ties sent to each group i. Newman demonstrates the use of the coefficient in an examination of a network known to be highly assortative – that of racial mixing among sexual partners. In an analysis of a sexual partners network, he finds an assortativity coefficient of 0.621 and concludes that "this network is strongly assortative by race – individuals draw their partners from their own group far more often than one would expect on the basis of pure chance," (Newman, 2003, p. 26126).

Table 3 reports Newman's assortativity coefficient for each of the primary categories in our data. Remarkably, the assortativity coefficients for both business unit (0.6841) and function (0.6500) within BigCo exceed 0.621, the value Newman observed for racial mixing within sexual networks. By contrast, the assortativity coefficient on gender (0.1756) is much lower. As we will confirm in the multivariate regressions that follow, organizational structure consistently exerts a stronger influence on interaction propensities than do sociodemographic group memberships.

Graphically, we can provide another glimpse of the large effect of group membership on interaction intensities at BigCo. We create distributions of the number 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. In other words, we compare the distribution of category memberships of those who actually e-mailed one another at BigCo to the overall distribution of category pairs in the data assuming random matching. Figure 1 shows this for four categories: business unit, function, office, and salary band. Blue bars represent the realized ties, and red bars the distribution formed by all possible pairs of communicating dyads (*i.e.*, all approximately 112 million cells of the ½(n)(n-1) network of potential dyads) – in effect, the benchmark assuming random mixing. Of course, the distribution of possible ties is shifted far to the right of the distribution of realized links. Slightly more than 70% of the actual communicating dyads in the company span zero, one, or two of these four boundaries; by contrast, less than 20% of pairs drawn at random would span two or fewer boundaries. At the other end of the distribution, just 10% of actually communicating dyads span all four boundaries – that is, they are in different business units, different functions, different offices and different salary bands, whereas, 46% of the randomly generated pairs of potential communicators share none of these category memberships.

There is, in short, strong evidence that interactions within BigCo are heavily influenced by organizational structure, and the descriptive statistics presented thus far provide strong clues regarding the organizational locations and types of individuals most likely to span the sharply delineated communication clusters in the company.

Multivariate Results – Dyad Level

Table 4 presents dyad-level Poisson QML regressions of the rate of e-mail exchange in BigCo. In Model 1, the single largest organizational effect on the rate of communication is sharing the same business unit affiliation. When individuals *i* and *j* are in the same business unit, they interact at exp(2.427)=11.32 times the rate of otherwise similar dyads that span different business units.

Considering the primacy of the corporate headquarters unit in theories of value creation in the multi-business firm, we single out the effect of CHQ in the regressions. 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*). (Recall that CHQ is treated as a business unit at BigCo.) The coefficient on *BothCHQ* is insignificant, indicating that members corporate headquarters staff have roughly the same propensity as members of other business units to communicate internally. 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 communicate at exp(0.729)= 2.07 times the rate. This suggests that, compared to members of other business units, BigCo's headquarters staff are more outwardly focused in their interaction patterns.

The effects of being in the same function and subfunction are large and together approach the magnitude of the same BU effect: two individuals in the same function communicate at exp(1.034)=2.81 times the rate of those who are in different functions, *ceteris paribus*. Because the models are multiplicative, two individuals who also are in the same subfunction communicate at a total of 9.37 (=2.81 × 3.33) times the baseline rate.

Turning next to geography, we include a dummy variable indicating whether employees *i* and *j* are in the same office and the log of the distance between them. *SameOffice* has a very large effect on the rate: two individuals communicate at [exp(1.154**SameOffice* – 0.161*ln(distance + 1 mile))] times the rate as otherwise identical, cross-office pairs. Note that the coefficients imply that the effect of *SameOffice* increases in comparison to pairs that are in different offices and are geographically separated; the coefficient on *SameOffice* alone gives the hazard ratio of a pair in the same office relative to a pair of individuals located in different offices but separated by zero miles. Compared to two employees separated by just 100 miles, two people in the same office communicate 6.67 times more frequently. Relative to a dyad separated by the mean geographic distance in the sample, two people in the same office communicate at 9.46 times the rate.

Model 1 also shows that pairs of individuals in the same salary band exchange e-mail at a higher rate, although the effect of salary band is considerably smaller in magnitude than that of business unit, office, and function. Members of the same band in the fourteen-level salary hierarchy communicate at a 1.21 times [=exp(.269)] higher rate than cross-salary-rank dyads[13].

We next turn to the two sociodemographic categories in the data, gender and tenure with the firm, which we add in Model 2 of Table 4.[14] The gender composition of the population (and of our subsample) is 70.1 percent male and 29.9 percent female. We do find a positive effect of gender homophily: male-male dyads exchange e-mails at 1.23 times [=exp(0.206)] the rate of cross-gender pairs. The same-gender increase in the rate is substantially larger for female-female

[13] In other results (not reported), we find a strong and significant effect of being in adjacent salary bands, but this effect is smaller than the same band effect; unlike other categories that are unordered, there is a proximity effect with respect to salary band.

[14] We assume, but cannot confirm, that proximity in the tenure distribution will correlate highly with proximity in the age distribution. BigCo is an organization that is known for treating its workforce well and avoiding layoffs, which likely results in a higher tenure-age correlation than in many other firms.

pairs; the effect of *BothFemale* in the full regression boosts the *SameGender* effect from 1.23 for male-male dyads to 1.48 [=exp(0.206 + 0.186)] for female-female dyads. Likewise, the effect of two employees being within five years of one another in their tenure with the firm also has a 1.65 times estimated effect [=exp(.501)] on the rate of communication.

Model 3 in Table 4 is the full model, including organizational, geographic, and sociodemographic covariates. Here, we find that, relative to the parameter estimates in Models 1 and 2, there is virtually no change in any of the estimated effects of formal structure—same business unit, same function, same office, and same salary band coefficients remain constant between models 1 and 3. However, there is significant attenuation in the effects of the shared sociodemographic groups, tenure and gender, when we estimate these effects net of the organizational structure covariates. The percent increase in communication (i.e., the rate multiplier minus one) on "tenure within 5 years" falls by more than half between models 2 and 3; the increase on male-male pairings drops by three fourths and becomes insignificant; and the female-female multiplier decreases by 10 percent. This suggests that there is considerable sorting by gender and tenure into specific divisional, functional, or spatial locations, which are not accounted for when we estimate the effects of same sociodemographic categories without accounting for formal organizational location.

The final three regressions in Table 4 report interaction effects between the same business unit, same sub-function, and same office dummy variables, along with (logged) sizes of each of these groups. Across all three groups, the hazard of within-group communications declines with the size of the group. For instance, the estimates imply that on average, two employees from the same business unit that is at the third quartile of the business unit size distribution have a rate of communication that is 0.76 times [= exp(7.51 × -0.361) / exp(8.29 × -

0.361)] that of an otherwise similar within-BU dyad from a smaller division at the first quartile of the size distribution. The *SameOffice* and *SameSubfunction* effects are also decreasing in group size (see Table 4). Thus, smaller business units, sub-functions, and office locations in BigCo have substantially greater rates of internal interaction than do larger groups.

Turning to Table 5, we report estimations of the full specification on two mutually exclusive subsets of the data: middle manager dyads (salary bands 7-10) and executive (bands 11-14) interactions.[15] In other words, the data for Model 1 comprise only dyads in which both members of the pair are middle managers; those for Model 2, all dyads in which both members are executives.

The interesting finding to emerge from the table is the relative influence of the organizational and sociodemographic covariates across the salary rank distribution. We find that executives are less constrained in their interactions by their respective organizational locations – while the magnitude of the effects of same business unit and same function are large in both models in table 5, they are smaller in executive-to-executive interactions.[16] Conversely, we find that, while the coefficients are not precisely estimated, executive-level interactions appear to be more strongly influenced by sociodemographic categories: the "same gender" and especially, "same tenure" dummy variables have larger coefficients among executives. These findings are

[15] In Table 5, we compute significance levels based on the addition of interaction effects in a pooled dataset that includes both middle-manager/middle-manager and executive-executive dyads. Our significance tests are run by including a *BothExec* covariate in the model and then interacting it with all other covariates in the model. A significant interaction between *BothExec* and *SameBU*, for example, indicates that the *SameBU* coefficient is significantly different between middle-middle and executive-executive dyads.

[16] Because all executives within a function also are classified in the same subfunction, there is no independent variation of *SameSubfunction* in the executive-to-executive model in Table 5. Thus, to compare the effect of function membership across the two models, we must drop the *SameSubfunction* variable. When we do so, we find that *SameFunction* has a significantly larger effect on interactions among middle managers than among executives (unreported results). Therefore, although the *SameFunction* effect is significantly larger among executives than among middle managers, it would be misleading to conclude that executives are more siloed by function; rather, we must compare the coefficient on *SameFunction* in the executive column against the sum of the coefficients on *SameFunction* and *SameSubfunction* in middle manager regressions.

consistent with classic sociological observations (Kanter, 1977) about the role of sociodemographic similarity in creating the interpersonal trust that managers need to handle uncertainty.

Before turning to the person-level category-spanning regressions, to test the robustness of these results, we examined six alternate ways of specifying our dependent variable and estimated similarly specified models on each. First, we admitted a broader set of ties to the analysis by relaxing the mass mailing threshold to include messages with as many as 10 recipients. Next, we restricted the analysis to "strong ties" by including only single-recipient messages and, in a separate analysis, by including only dyads that exchanged at least the sample mean of 12 messages. We also analyzed a count of the number of e-mails that were limited to correspondences including file attachments. Whereas all of our other analyses focused on direct dyadic communications, in one analysis we examined co-receipt of mass mails as a proxy for structural similarity in the communication network. Finally, we analyzed data from our entire sample to test the robustness of our sub-sampling procedure. Full explanation, results (which are substantively similar to the core results) and discussion can be found in Appendix 2. Finally, we compared the core e-mail results against results of calendar data and found few meaningful differences; these results are in Appendix 3.

The overall conclusion to emerge from the dyad-level analysis is that organizational structure and geographic space sharply delimit patterns of exchange. Social categories also influence propensities to interact, but the magnitudes of their effects are modest relative to those of organizational structure and the (organizationally assigned) spatial organization of BigCo.

Results on Person-Level "Category-Spanners"

In light of the overwhelming evidence that the dominant lines of communication at BigCo are within organizational units and geographic space, the question then follows: who are the individuals that form bridges across these clustered groupings by participating in cross-category exchanges? To answer this question, we turn to the regressions of the individual-level measures of propensity to communicate across infrequently spanned categories on dummy variables for an employee's membership categories themselves. Results of these regressions appear Table 6, which reports the fractional logit regressions of the "improbability" of individuals' communication profiles. In reporting these results, recall that is necessary to estimate the effect of each category membership on a specification of the category-spanning dependent variable (Improb_i) that is calculated without reference to that category. For convenience, we have assembled all of the correct coefficients into a single column (5) in the table and the reader need only examine that column, but each of the actual regressions are reported in Models (1)-(4) in the table.

First, despite evidence from other contexts showing that men have more structurally diverse networks, the reverse is indeed the case in these data. After accounting for job function, salary band, and business unit (as well as removing administrative support personnel from the data), women in the company are engaged in significantly more category-spanning ties than men. The gender coefficient suggests that women have communication patterns that are 51% more "improbable" than men. What does this mean? It implies that men in BigCo communicate within and across the group boundaries that are well trodden; their e-mail exchanges either stay within group boundaries or span category memberships that are commonly crossed. Women are substantially more likely to be involved in communication dyads that span offices, business units, job functions, and salary bands.[17]

With respect to salary bands, we find that individuals in the high – but not the very highest – salary bands engage in the most boundary-spanning communications. Middle managers have communication patterns that are 1.19 times as improbable as those of the rank and file; Junior and mid-level executive in bands 11 through 13 have successively higher likelihoods of bridging organizational and social boundaries: their communication is 1.63, 1.87, and 3.82 times more improbable than that of rank-and-file employees, respectively. The results on band 14, the firm's most senior managers, are somewhat surprising: their communication patterns are less improbable than the omitted category, rank-and-file employees. Although this implies that senior executives are not directly inter-group brokers in the firm, one possible explanation for this result is that the most senior executives effectively define the basic patterns of interaction in BigCo, and the rest of the organization then follows suit. In other words, if the senior-most executives in the firm emphasize coordination between particular group boundaries, those in the lower ranks then enact this coordination by communicating along the same pathways. By the construction of our improbability measure, this would render band 14 communications low on our dependent variable because they would track the commonly traveled inter-group links in the firm, albeit in a leading manner that is not captured in our cross-sectional analysis.

With respect to business units, we find that members of corporate headquarters are more likely than members of other business units to have statistically improbable, cross-category

[17] It is possible to estimate the effect of gender (and every other category membership) on the propensity to span each separate boundary. $Improb_i$ is a composite of all categories, but we can separately estimate the effect of "gender is female" on the probability of spanning: business unit, function, subfunction, office location, and salary band. When we do this, we find, for example, that women are 11% [=exp(0.1060)] more likely to span functional boundaries than are men. These results are too lengthy to report but are available from the authors upon request.

communication patterns: the coefficient on *CHQ* is positive and significant and the magnitude of the effect is about 28%. Similarly, we find that individuals whose job function is described as general executive management have communication patterns that are remarkably unconventional – they are over eight times as likely as other job functions to span boundaries. And consistent with the results in the descriptive statistics, members of the marketing, sales, and services job functions also are more likely to have boundary-spanning communication patterns, though the magnitude is far less pronounced that that of the general executive management function.

Finally, it is worth noting the association between our measures of boundary-spanning and Burt's (1992) measure of structural constraint, the inverse of brokerage. These results are reported in Model 8 and are generally quite similar to those on the improbability of individuals' communication profiles: as with our category-spanning measures, women, employees in the marketing and sales functions, middle managers, and executives in bands 11 through 13 are all significantly more likely to have low constraint networks[18]. Unlike the category-spanning results, however, those in the highest salary band and those in the general executive management are not statistically different in their levels of brokerage.

Discussion and Conclusion

In his letter to shareholders in the 1989 annual report, Jack Welch described plans for what would become one of the cornerstone initiatives in his long-tenured leadership of General Electric. Specifically, he wrote of his intention to mold GE into a boundaryless organization, stating, "The boundaryless company we envision will remove the barriers among engineering,

[18] Positive effects on brokerage are indicated by statistically negative coefficients because the structural constraint measure is the inverse of brokerage.

manufacturing, marketing, sales and customer service; it will recognize no distinction between domestic and foreign operations – we'll be as comfortable doing business in Budapest and Seoul as we are in Louisville and Schenectady. A boundaryless organization will ignore or erase group labels such as 'management,' 'salaried' or 'hourly,' which get in the way of people working together."

Leaders of the organization we analyze here similarly have stressed the importance of lateral, cross-division, cross-function, and cross-rank coordination within the company, but the reality is that interaction patterns at BigCo (as we imagine they would at GE as well) appear to follow those of a standard bureaucratic model in which formal organization structures and office locations sharply delimit interaction patterns. Repeating the statistic we reported earlier, relative to two people who share none of these categories in common and who are geographically separated by the sample's mean geographic distance, pairs of individuals that are in the same business unit, subfunction, and office location communicate at an estimated rate that is 1,000 times higher. Social categories also matter at BigCo, but to a much lesser degree. Moreover, the formal authority structure of the firm clearly forms the vertical column of interaction: employees communicate within salary levels and with those in adjacent salary bands, but only rarely do they e-mail beyond this range: fully 76% of all e-mails in our sample are sent to a member of the same or an adjacent salary band as the sender; just 6.4% of e-mails span at least two salary bands and a mere 1.6% of e-mails span at least three salary grades. In all these results, we find few meaningful differences between e-mail data and calendar data, which gives us confidence that the social interactions we observe are, indeed, descriptive of the core business of the firm.

When we invert our perspective to focus on those who span the densely interacting groups within the firm, we were surprised to discover that women at BigCo are more likely to bridge the communication silos in the company. The available evidence suggests that this finding is neither an artifact of gendered sorting by job function, nor is it indicative of a gender difference in preference for communication media. Similarly, those in the general executive management, sales, and marketing functions, as well as junior- to mid-level executives, were also more likely to engage in category-spanning communication patterns that run across the less-frequently traversed boundaries in the firm.

The results we report are based on overall effects, averaged over a large sample of employees. An alternative possibility is that coordination on a small scale, involving relatively few key actors – such as, perhaps, the few members of the top management team – has major consequences for the organization. Indeed, one of our most surprising findings is the modest role that the firm's most senior executives seem to play in coordinating the activities of the enterprise. But if coordination can occur through the actions of a few key people managing the interfaces of otherwise modular organizational units, then it is possible that there is extensive and highly consequential coordination occurring in spite of the relatively little communication across formal organizational boundaries. Although it seems unlikely in the complex world of information technology products and services that meaningful coordination among large organizational units could occur without widespread interaction among middle management and technical personnel, the analysis we presented thus far does not exclude this possibility. Therefore, we separately looked at the 200 individuals with the most improbable communication profiles in the sample (i.e., those that are most likely to bridge functions, business units, and office locations that interact infrequently). When we did this, we reproduced our core results: women are over-represented (35.2% of the 200, compared to 30.1% of the firm's population); mid-level executives (bands 11-13) are dramatically over-represented (28% compared to 2.7% of the population); and sales, marketing and general executive management are also over-represented (25%, 11% and 2.0%, respectively, compared to 13%, 1.7% and 0.1% of the population). So regardless of whether coordination is effected through widespread communication or communication among a select group of important individuals, we find that the same categories of people are crucial in spanning organizational boundaries.

Analyses of the coordinating role of the corporate headquarters yield relatively weak but consistent results. Cross-business-unit pairs of individuals are more likely to communicate if one member is in CHQ, and members of CHQ are more likely than members of many other business units to occupy category-spanning positions in the broader social structure of the firm. Ceteris paribus, members of CHQ appear to communicate across business unit and functional boundaries. Of course, the fact that members of corporate headquarters have somewhat (but not dramatically) more category-spanning communication profiles is not necessarily diagnostic of the overall role of CHQ staff in facilitating inter-unit coordination—these results may understate the true impact of headquarters. It is possible, for instance, that CHQ members may play the role of brokering introductions within BigCo. 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. The bottom line, however, is that the organization we study is one that would be largely familiar to Chandler.

Before describing avenues for extending this study, we note a few of its more significant limitations. First, we know memberships in formal organizational units (strategic business unit, function and sub-function) only. We do not know how authority relations, including direct lines of reporting, incentive systems or structural overlays put into place by senior managers, shape interaction patterns. Therefore, when we present the analysis we operated with 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 the flow of work.

We began with the observation that although theories of communication and coordination are central to the field of organization theory, we have theories and assumptions but little empirical evidence about the structure of communication in the modern, complex organization. In some ways, the fact that there is no directly comparable analysis of the communication structure of a large group of company employees is a limitation of this work. Notwithstanding the Marmaros and Sacerdote (2004) and Kossinets and Watts (2008) papers, we know of no truly comparable study to this and thus we have few benchmarks to serve as a baseline for comparing 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 one's prior about what the magnitude of these effects would be in a more siloed versus a more lateral organization. Although there are some rough benchmarks in the literature, neither a directly comparable nor a well-established baseline exists.

In a related vein, 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 but a single organization. At the moment, we have no basis for any claim of generalizability beyond the single organization we study. However, we see this limitation as an opportunity for future research. There are myriad types of organizations – to name a very few of the endless distinctions, large and small, young and old, high- and low-technology, product and service companies, for-profits and not-for-profits, domestic and international. Will our findings extend to any or many of these other types of organizations?

With the availability of e-mail datasets such as the one we examine here, we believe that this last limitation in fact also presents a tremendous opportunity for organization theorists to initiate a new research program on the nature and diversity of communication and coordination in modern organizations. Recalling two famously contrasting perceptions of the nature of organizational communities that has animated a great debate in organization theory during the past few decades – that of Hannan and Freeman (1977), which posed the classic question, why are there so many organizational forms, and DiMaggio and Powell's (1983) rejoinder, why are there are so few – we now wonder about the level of *internal* diversity of organizations. Electronic communications data should offer an unprecedented window into the social and work relations inside firm. Not only will this offer an opportunity for us to develop taxonomies of internal organizational structures, it will also enable analysis of many individual-, group-, and organization-level outcomes.

There is a second set of analyses that clearly merit attention. Given the relatively short, three-month observation period of this study, we have opted to treat the data as a cross-section. Longitudinal analyses promise insight into the dynamic nature of intraorganizational communication and, with the right research designs, are likely to yield causal analyses of the evolution of social networks inside organizations. Similarly, longitudinal analyses can shed light on the dynamic sequence of social interaction and evolving "conversations" in asynchronous communication media – as well as offer a fascinating window into social norms and social roles in electronic networks. We hope that this cross-sectional description of macro-organizational communication will begin to lay the foundation for future longitudinal analyses.

Appendix 1: Sub-sampling Procedure

Because our data were collected as a snowball sample, they were not necessarily representative of the overall population of U.S. employees at BigCo. Initially, we were concerned that our sample would be biased because its core is a group of people who were chosen specifically because of their involvement in cross-divisional activities. Despite the firm's stated goal of encouraging cross-divisional coordination, there is nevertheless reason to believe that such people are unusual. We believe these concerns are mitigated, however, by the large number of second-order contacts in the data. Due to large-n meetings and broadcast e-mails – which are unrelated to the core actors' cross-divisional activities – our snowball core of 66 people yielded a sample of over 30,000 based on just single-step connections.

We were also concerned about the degree to which our sample was representative of the firm's employee population. Specifically, because our snowball originated disproportionally with executives in the corporate salesforce, we found that we had over-sampled corporate headquarters and the corporate salesforce; dramatically over-sampled executives (our sample captured over 80% of U.S. executives) and correspondingly under-sampled rank-and-file employees; and sampled the functions at different rates. As a result, the population of dyads might over-represent certain types of interactions while under-representing other types.

One approach to dealing with this problem would be to employ sampling weights in our statistical analysis, but this solution was complicated by our dyadic unit of analysis and by the fact that we have already used sampling weights for the zeroes in our case cohort data set. Instead, we exploited the large size of our sample to randomly choose individuals to create a sub-sample that is representative of the population with respect to important demographic variables, namely: middle managers (bands 7-10) and each of the executive salary bands (11-14); the general executive management, marketing, sales and services functions; and corporate headquarters.

To assemble the representative sub-sample, we created a three-dimensional matrix of salary band (middle managers, 11, 12, 13, 14, everyone else), function (general executive management, marketing, sales, services, everyone else) and business unit (corporate headquarters, everyone else). For each of the 60 cells of this 6×5×2 matrix, we calculated the sampling probability that would be needed to achieve a sub-sample rate of 15.9% of the U.S. population of the firm (compared to 23.8% of the U.S. population of the firm in the original sample). We chose to make our sub-sample representative of only selected groups in order to maintain a large sample size; had we made our sub-sample representative across the board, we would have diminished our sample to just 2.9% of the U.S. population of the firm.

Once we had these probabilities, we used a random number generator to determine whether each person in the overall sample, given her salary band, function and business unit, would be included in the sub-sample. Across multiple models, we frequently created new sub-samples – each with slight, random differences in size and composition – to ensure that our results do not depend on random idiosyncrasies in sub-sample selection. We also test the robustness of our results against the full sample in Appendix 2.

Appendix 2: Robustness

We have conducted numerous tests to assess the robustness of the results. In general, we find that across a broad range of different specifications and assumptions regarding the e-mail data, the results we report in the paper prove to be highly robust. Table A1 in this appendix details some of these alternatives, with a particular focus on alternative definitions of the dependent variable.

Mass Mailings. In the paper we analyzed only messages with four or fewer recipients. We chose to exclude "mass mailings" because we assume that they are less important to the social structure of the firm than are more targeted communications. Although there is empirical precedent in the literature for four as a threshold (Quintane, and Kleinbaum, 2008), our first robustness check relaxes that assumption, admitting all e-mails with as many as 10 recipients[19]. The results of this analysis yield only a single substantive change relative to the full baseline model in Table 4 in the text (reproduced in Table A1 for ease of comparison). The dummy variable for both members of the dyad being in corporate headquarters is statistically significant in the 10 recipients regression, but not in the results reported in the paper.

Strong Ties. Conversely, we also reconstructed our dependent variable to capture only the (presumed) strongest ties in the data. We adopted two different measurement approaches to gauge strong ties. First, we limited the analysis to messages with just a single recipient, or one-to-one correspondences. We reasoned that, on average, these are more likely to contain more meaningful work or social content than messages addressed to multiple recipients. In the second

[19] Each robustness check in this section required a different data set with different numbers of realized dyads; as a result, each data set was re-balanced separately to maintain the 1:1 ratio of zeros to non-zeros, as described in the *Methods* section above.

approach, we included only dyads who exchanged at least 12 messages (with up to 4 recipients, as in our primary analysis) during the observation period (i.e., an average of at least one message per week). As before, the overall picture resulting from our analysis of these "strong ties" networks is largely the same as the results of our baseline models, but there are some subtle differences. As in the paper, the dominant effects in the "strong tie" network are business unit, function, and office location. The differences between the regressions is limited to the less precisely estimated sociodemographic effects; for instance, the same gender effect loses significance in the regression of 12+ e-mail exchanges, but this (and all other) point estimates remain very close to those in the baseline specification.

Work Ties. Next, we attempt to test the robustness of our results against a subset of the data that may be more likely to be explicitly work-related. Reliably identifying work-related communications in the absence of e-mail content is impossible, but as a rough approximation, we attempt to leverage the header information that is available to us by including in this analysis only e-mails containing file attachments. In a corporate environment, file attachments most often include spreadsheets, presentations, and other documents relating to performing, sharing and communicating of work. The results of this analysis again are substantively similar to the baseline models.

Co-occurrence in Mass Mail Headers: Whereas all of our other analyses focused on direct dyadic communications, in one analysis we examined co-receipt of mass mails. Mass mails, those addressed to more than four recipients, are often sent to large numbers of individuals, but they are rarely sent indiscriminately. The set of individuals who co-receive the same mass mails have something in common, so the count of mass mails co-received can be construed as a rough proxy for structural similarity in the communication network. Three coefficients have substantially different point estimates across the two models. The *SameBU* effect is about one third weaker and the *TenureWithin5* coefficient is about half as strong in the co-recipient model than the baseline model; additionally, the *BothCHQ* effect, insignificant in the baseline model, is positive and significant in the co-recipient model.

Subsampling. Finally, we test the robustness of our results against assumptions implicit in our sub-sampling procedure. In most of our analyses, we use a randomly-drawn sub-sample of individuals that better reflects the demographics of the population of the firm, but we find that the results are substantively similar using the full sample, which is over three times larger. The only substantive change relative to the baseline model is in the coefficient of *BothCHQ*. In our baseline model, the coefficient was insignificantly different from zero, suggesting that *BothCHQ* dyads communicate at the same rate as other *SameBU* dyads. These data suggest that *BothCHQ* dyads communicate at 0.733 [= exp(-0.310)] times the rate of other *SameBU* dyads, a significant effect that is relatively modest in magnitude.

Appendix 3: Calendar Data

The data we received from the company include logs of e-mail and meetings scheduled on electronic calendars. In the main text of this paper, we have focused exclusively on the e-mail data. One of the interesting findings from our analysis is that the overlap between the e-mail 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 e-mail and these two other modes of communication is striking[20].

To document these results, we ran the same set of baseline models on both the e-mail and calendar data sets (Table A2). The correlation between the vectors of coefficients in the set of regression specifications in which the dependent variable is defined to be count of dyadic communications based on, respectively, e-mail and calendar ties is in excess of 0.99 in the middle manager population; the executives population shows a lower correlation (0.56) primarily due to greater deviations in the insignificant coefficients. 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 e-mail 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 A1, for instance, plots the proportion of both e-mail and calendar communications that connect two actors separated by the geographic distances

[20] Note also that this similarity is not the spurious result of e-mails used to schedule appointments; such e-mails are coded in the data and were excluded from the analysis.

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, reinforcing our suggestion that e-mail is a complement to, rather than a substitute for, other modes of communication. The only evidence we see of substitution is in the *SameOffice* effect among executives; that coefficient is insignificant in the e-mail data, but positive and significant in the calendar data. This small caveat notwithstanding, we believe that this evidence bolsters our claim that, at least in this organization, e-mail provides an excellent representation of the true intraorganizational communication structure.

REFERENCES

Ancona, D., H. Bresman, and K. Kaeufer
2002 "The comparative advantage of X-teams." MIT Sloan Management Review, 43: 33.

Baron, J. N., and W. T. Bielby
1980 "Bringing the Firms Back in: Stratification, Segmentation, and the Organization of Work." American Sociological Review, 45: 737-765.

Batagelj, V., and A. Mrvar
2006 "Pajek: Program for Large Network Analysis." Version 1.12 ed.

Blau, P. M., and J. E. Schwartz
1984 Crosscutting social circles : testing a macrostructural theory of intergroup relations. Orlando: Academic Press.

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

Burt, R. S.
1992 Structural holes : The social structure of competition. Cambridge, MA: Harvard University Press.

2004 "Structural Holes and Good Ideas." American Journal of Sociology, 110: 349-399.

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

1977 The visible hand : the managerial revolution in American business. Cambridge, Mass.: Belknap Press.

Cohen, M. D., J. G. March, and J. P. Olsen
1972 "A Garbage Can Model of Organizational Choice." Administrative Science Quarterly, 17: 1-25.

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

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

DiMaggio, P. J., and W. W. Powell
1983 "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." American Sociological Review, 48: 147-160.

Ding, W. W., F. Murray, and T. E. Stuart
2006 "Gender Differences in Patenting in the Academic Life Sciences." Science, 313: 665-667.

Galbraith, J. R.
1973 Designing complex organizations. Reading, MA: Addison-Wesley.

Goold, M., A. Campbell, and M. 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, M. 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.

Han, S.-K.
1996 "Structuring relations in on-the-job networks." Social Networks, 18: 47-67.

Hannan, M. T., and J. Freeman
1977 "The Population Ecology of Organizations." American Journal of Sociology, 82: 929-964.

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

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

Kanter, R. M.
1977 Men and women of the corporation. New York: Basic Books.

King, G., and L. Zeng
2001 "Logistic Regression in Rare Events Data." Political Analysis, 9: 137-163.

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

Kossinets, G., and D. J. Watts

2006 "Empirical Analysis of an Evolving Social Network." Science, 311: 88-90.

2008 "Origins of Homophily in an Evolving Social Network." Unpublished Manuscript.

Krackhardt, D., and J. R. Hanson
1993 "Informal networks: The company behind the charts." Harvard Business Review, 71: 104-111.

Lancaster, T., and G. Imbens
1996 "Case-control studies with contaminated controls." Journal of Econometrics, 71: 145-160.

Lawrence, P. R., and J. W. Lorsch
1967 Organization and Environment: Managing Differentiation and Integration, 1986 ed. Boston: Harvard Business School Press.

Lazarsfeld, P. F., and R. K. Merton
1954 "Friendship as Social Process: A Substantive and Methodological Analysis." Freedom and Control in Modern Society: 18-66.

Marmaros, D., and B. Sacerdote
2004 "How Do Friendships Form?"

Marsden, P. V.
1987 "Core Discussion Networks of Americans." American Sociological Review, 52: 122-131.

1988 "Homogeneity in confiding relations." Social Networks, 10: 57-76.

Mayo, E.
1949 "Hawthorne and the Western Electric Company." In D. S. Pugh (ed.), Organization Theory, Selected Readings: Penguin Books.

McPherson, J. M., and L. Smith-Lovin
1986 "Sex Segregation in Voluntary Associations." American Sociological Review, 51: 61-79.

Mintzberg, H.
1973 The nature of managerial work. New York: Harper & Row.

Newman, M. E. J.
2003 "Mixing patterns in networks." Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), 67: 026126.

Nohria, N.
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.

Ouchi, W. 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.

Papke, L. E., and J. M. Wooldridge
1996 "Econometric Methods for Fractional Response Variables With an Application to 401 (K) Plan Participation Rates." Journal of Applied Econometrics, 11: 619-632.

Podolny, J. M., and J. N. Baron
1997 "Resources and Relationships: Social Networks and Mobility in the Workplace." American Sociological Review, 62: 673-693.

Powell, W. W.
1990 "Neither Market Nor Hierarchy: Network Forms of Organization." In B. M. Staw, and L. L. Cummings (eds.), Research in Organizational Behavior: 295-336. Greenwich, CT: JAI Press.

Quintane, E., and A. M. Kleinbaum
2008 "Mind Over Matter? The Correspondence Between E-mail and Survey Measures of Social Networks." Unpublished Manuscript.

Ridgeway, C. L., and L. Smith-Lovin
1999 "The Gender System and Interaction." Annual Review of Sociology, 25: 191-216.

Silva, J. M. C. S., and S. Tenreyro
2006 "The Log of Gravity." Review of Economics and Statistics, 88: 641-658.

Thompson, J. D.
1967 Organizations in action; social science bases of administrative theory, 2nd ed. New Brunswick, NJ: Transaction Publishers.

Tyler, J. R., D. M. Wilkinson, and B. A. Huberman
2003 "Email as Spectroscopy: Automated Discovery of Community Structure within Organizations." In M. H. Huysman, E. Wenger, and V. Wulf (eds.), Communities and Technologies: 81-96.

Watts, D. J., and G. Kossinets
2008 "Origins of Homophily in an Evolving Social Network." Unpublished manuscript.

Weber, M.
1924 "Legitimate Authority and Bureaucracy." In D. S. Pugh (ed.), Organization Theory: Selected Readings: 3-15. New York: Viking Penguin.

Weick, K. E.
1976 "Educational Organizations as Loosely Coupled Systems." Administrative Science Quarterly, 21: 1-19.

Weisband, S. P.
2008 Leadership at a distance : research in technologically-supported work. New York: Lawrence Erlbaum Associates.

Wiener, N.
1948 Cybernetics; or, Control and communication in the animal and the machine. New York: J. Wiley.

Williamson, O. E.
1975 Markets and hierarchies, analysis and antitrust implications : a study in the economics of internal organization. New York: Free Press.

Wooldridge, J. M.
1997 "Quasi-Likelihood Methods for Count Data." Handbook of Applied Econometrics, 2: 252-406.

Zenger, T. R., and B. S. Lawrence
1989 "Organizational Demography: The Differential Effects of Age and Tenure Distributions on Technical Communication." The Academy of Management Journal, 32: 353-376.

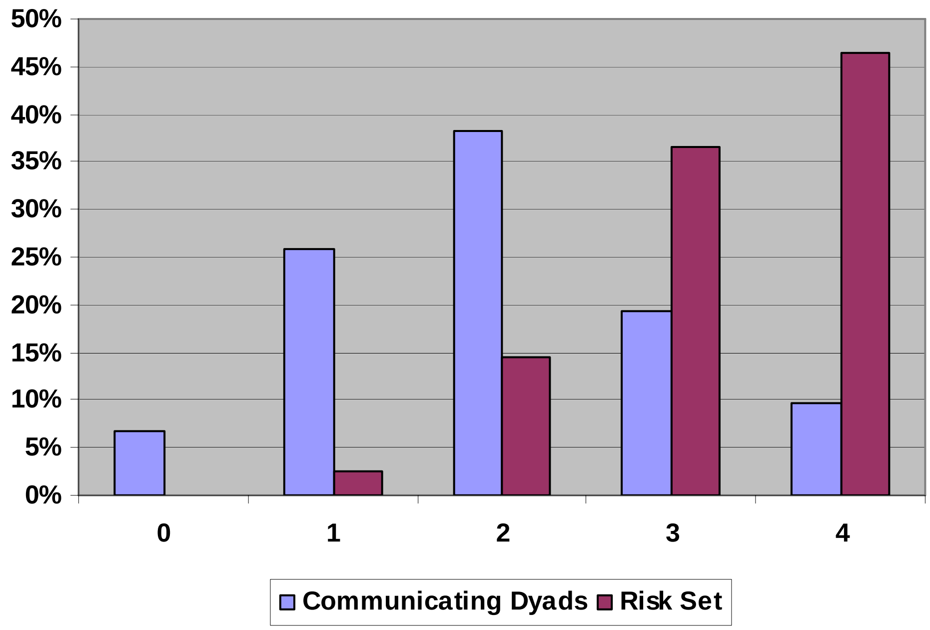

Figure 1: Risk Set versus Realized Ties.

Table 1: Overall summary statistics about the sub-sample; as in our primary analysis, these statistics are based on non-mass, non-BCC communications.

		Gender		Corporate Headquarters		Level		
		M	F	No	Yes	Exec	MiddleMgr	Rank&File
E-mail Data	Degree	90.0 [90.6]	107.6 [116.1]	92.6	96.0	190.8	86.5	35.7
	In-sample volume	1130.3 [1133.9]	1367.5 [1422.2]	1233.4 *	1193.1	2643.5	1065.2	378.7
	Out-of-sample volume	786.8 [792.3]	933.3 [930.3]	1089.5	760.3	1101.6	820.0	405.2
	Average Frequency	12.45 [12.42]	13.04 [12.70]	13.46	12.40	13.47	12.60	11.26
	Average Distance	816.8	854.2	771.5	843.6	801.6	844.7	584.7
	Average Bands Spanned	1.35	1.42	1.46	1.34	1.95	1.23	2.62
	Percent in BU	52.4%	51.5%	57.2%	50.8%	41.5%	53.1%	59.8%
	Percent in Function	57.1%	53.6%	52.2%	57.1%	43.3%	57.3%	62.1%
	Percent in Office	14.5%	13.0%	15.8%	13.5%	14.2%	13.4%	24.3%
	Percent in Band	27.3%	25.6%	25.4%	27.2%	16.4%	29.0%	10.2%
	Herfindahl (BU)	0.386 (n.s.)	0.383	0.431	0.372	0.284	0.391	0.509
	Herfindahl (Function)	0.412	0.385	0.396	0.406	0.256	0.416	0.529
	Herfindahl (Office)	0.107	0.095	0.119	0.099	0.082	0.100	0.225
	Herfindahl (Band)	0.205	0.195	0.193	0.204	0.138	0.208	0.229
Calendar Data	Degree	41.4 [41.6]	45.3 [56.9]	36.4	44.3	80.4	39.6	9.2
	In-sample volume	248.9 [250.9]	309.7 [386.3]	241.9	274.1	418.9	259.3	62.8
	Out-of-sample volume	188.5 [189.7]	209.3 [220.3]	242.6	181.7	249.6	195.4	55.7
	Average Frequency	6.19 [6.21]	7.02 [7.04]	7.09	6.26	5.18	6.73	4.08
	Average Distance	769.6	808.1	732.7	794.4	769.8	801.0	441.5
	Average Bands Spanned	1.41	1.52	1.60	1.40	2.86	1.25	1.76
	Percent in BU	49.0%	50.8%	59.0%	46.9%	41.6%	52.0%	22.1%
	Percent in Function	53.7%	52.5%	50.7%	54.1%	38.9%	56.7%	24.9%
	Percent in Office	11.8% (n.s.)	12.4%	15.4%	11.0%	16.5%	12.4%	-6.3%
	Percent in Band	23.2% (n.s.)	22.7%	22.6% (n.s.)	23.2%	13.9%	26.5%	-19.8%
	Herfindahl (BU)	0.388	0.405	0.473	0.372	0.335	0.411	0.207
	Herfindahl (Function)	0.431	0.416	0.424 (n.s.)	0.427	0.289	0.453	0.244
	Herfindahl (Office)	0.129 (n.s.)	0.131	0.161	0.121	0.112	0.138	0.022
	Herfindahl (Band)	0.221 (n.s.)	0.221	0.223 (n.s.)	0.220	0.165	0.238	0.030

*** difference significant at 5%; n.s. not significantly different**

Except where noted otherwise, all differences are significant at p < 1%

For comparison purposes, we show several variables' averages <u>including</u> administrative assistants in square brackets

Table 2: Average number of dyadic e-mails between salary bands (Panel A), business groups (Panel B) and selected functions (Panel C) respectively. Herfindahls are normalized to range from 0 to 1 (instead of the standard 1/N to 1) for greater comparability across variables.

	Average E-mails Exchanged										Percent	Normalized
	5	6	7	8	9	10	11	12	13	14	Intra-Unit	Herfindahl
5	10.98	0.17	0.29	0.50	0.14	0.15	0.02	0.02	0.00	0.01	21.5%	78.3%
6	0.17	23.55	13.86	7.80	6.50	4.57	1.74	0.71	0.31	0.03	10.4%	16.6%
7	0.29	13.86	36.64	21.85	20.46	16.28	3.04	1.30	0.40	0.03	13.8%	12.0%
8	0.50	7.80	21.85	65.04	54.16	41.73	5.56	2.18	0.17	0.04	20.2%	15.5%
9	0.14	6.50	20.46	54.16	126.90	90.14	16.30	4.52	1.32	0.11	28.5%	19.1%
10	0.15	4.57	16.28	41.73	90.14	128.40	32.00	13.11	3.91	0.96	26.0%	17.1%
11	0.02	1.74	3.04	5.56	16.30	32.00	46.18	39.55	20.34	2.34	5.2%	10.6%
12	0.02	0.71	1.30	2.18	4.52	13.11	39.55	86.44	10.12	7.61	9.4%	27.0%
13	0.00	0.31	0.40	0.17	1.32	3.91	20.34	10.12	96.39	75.16	8.2%	28.6%
14	0.01	0.03	0.03	0.04	0.11	0.96	2.34	7.61	75.16	332.15	28.9%	62.5%

Shaded blue cells are above the 80th percentile

Panel A

	Average E-mails Exchanged						Percent	Normalized
	BizSvcs	TechSvcs	Software	Hardware	CHQ	Sales	Intra-Unit	Herfindahl
BizSvcs	**195.35**	10.60	4.39	0.89	**21.66**	10.52	71.0%	58.7%
TechSvcs	10.60	**204.53**	5.62	4.03	**27.71**	**20.97**	54.8%	49.3%
Software	4.39	5.62	**276.53**	8.70	**11.75**	**17.34**	71.2%	67.9%
Hardware	0.89	4.03	8.70	**156.09**	8.83	9.41	60.0%	63.6%
CHQ	**21.66**	**27.71**	**11.75**	8.83	**265.99**	**17.31**	64.9%	49.7%
Sales	10.52	**20.97**	**17.34**	9.41	**17.31**	**184.30**	51.3%	42.6%

Bold cells are above the 50th percentile; shaded blue cells are above the 80th percentile

Panel B

	Average E-mails Exchanged										Percent	Normalized
	CO	FI	GM	HR	LE	MK	RD	SC	SL	SV	Intra-Unit	Herfindahl
Communications	**331.82**	1.14	4.40	4.94	1.21	16.59	3.02	0.95	1.43	1.23	57.1%	80.1%
Finance	1.14	**315.03**	5.49	9.16	1.72	2.40	2.51	12.75	16.16	12.86	56.9%	66.1%
General Exec Mgmt	4.40	5.49	**386.94**	6.79	5.40	2.14	0.83	0.07	0.71	0.22	32.0%	86.5%
Human Resources	4.94	9.16	6.79	**320.68**	7.34	1.98	2.25	3.32	4.61	5.48	43.3%	74.2%
Legal	1.21	1.72	5.40	7.34	**135.66**	1.62	2.03	5.11	2.53	0.60	24.1%	66.2%
Marketing	16.59	2.40	2.14	1.98	1.62	**148.88**	7.32	0.64	8.70	2.58	39.4%	56.4%
R&D	3.02	2.51	0.83	2.25	2.03	7.32	**271.92**	2.01	13.35	14.70	72.4%	69.7%
Supply Chain	0.95	12.75	0.07	3.32	5.11	0.64	2.01	**269.97**	7.48	6.21	54.7%	74.3%
Sales	1.43	16.16	0.71	4.61	2.53	8.70	13.35	7.48	**189.70**	25.93	50.4%	45.4%
Services	1.23	12.86	0.22	5.48	0.60	2.58	14.70	6.21	25.93	**174.52**	67.1%	47.7%

Shaded blue cells are above the 80th percentile; bold blue cells are above the 90th percentile

Panel C

Table 3: Univariate assortativity coefficients for the major variables in our analysis. Assortativity (Newman, 2003) is defined as the degree to which actors in a network interact across levels of a variable, independent of other variables.

Variable	**Assortativity Coefficient**
BU	0.6841
Function	0.6500
Subfunction	0.4340
Office	0.2660
Band	0.1544
Gender	0.1756

Table 4: Primary results

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
SameBU	2.427		2.410	4.930	2.478	2.404
	(0.033)**		(0.033)**	(0.224)**	(0.033)**	(0.033)**
BothCHQ	-0.002		0.003	0.096	-0.084	-0.031
	(0.049)		(0.049)	(0.051)	(0.051)	(0.050)
OneCHQ	0.729		0.724	0.435	0.744	0.708
	(0.033)**		(0.033)**	(0.028)**	(0.034)**	(0.033)**
SameFunction	1.034		1.031	1.053	0.915	1.029
	(0.027)**		(0.027)**	(0.027)**	(0.027)**	(0.027)**
SameSubfunction	1.204		1.191	1.190	4.519	1.178
	(0.033)**		(0.034)**	(0.034)**	(0.215)**	(0.034)**
SameOffice	1.154		1.158	1.136	1.137	3.972
	(0.078)**		(0.082)**	(0.085)**	(0.090)**	(0.294)**
Distance (log)	-0.161		-0.161	-0.161	-0.162	-0.156
	(0.010)**		(0.010)**	(0.011)**	(0.012)**	(0.010)**
SameBand	0.269		0.264	0.269	0.210	0.258
	(0.029)**		(0.029)**	(0.029)**	(0.030)**	(0.028)**
SameGender		0.206	0.055	0.058	0.045	0.055
		(0.015)**	(0.028)	(0.029)*	(0.029)	(0.029)
BothFemale		0.186	0.285	0.278	0.268	0.287
		(0.023)**	(0.041)**	(0.042)**	(0.043)**	(0.041)**
TenureWithin5		0.501	0.270	0.278	0.279	0.268
		(0.015)**	(0.029)**	(0.029)**	(0.030)**	(0.029)**
InSample (log)	1.488	1.420	1.483	1.490	1.474	1.475
	(0.024)**	(0.016)**	(0.024)**	(0.024)**	(0.025)**	(0.024)**
OutOfSample (log)	0.046	0.082	0.050	0.042	0.043	0.044
	(0.024)	(0.015)**	(0.023)*	(0.024)	(0.024)	(0.024)
AvgBUSize (log)	-0.568	-0.262	-0.577	-0.303	-0.588	-0.567
	(0.020)**	(0.013)**	(0.020)**	(0.019)**	(0.020)**	(0.019)**
AvgFunctionSize (log)	-0.146	-0.404	-0.130	-0.130	-0.129	-0.146
	(0.025)**	(0.015)**	(0.025)**	(0.025)**	(0.027)**	(0.025)**
AvgSubfunctionSize (log)	-0.138	0.106	-0.150	-0.148	0.101	-0.144
	(0.023)**	(0.014)**	(0.023)**	(0.023)**	(0.021)**	(0.023)**
AvgOfficeSize (log)	-0.219	-0.106	-0.214	-0.206	-0.205	-0.092
	(0.017)**	(0.011)**	(0.018)**	(0.018)**	(0.019)**	(0.016)**
AvgBandSize (log)	-0.002	0.054	0.003	0.002	0.041	0.008
	(0.027)	(0.029)	(0.028)	(0.028)	(0.028)	(0.027)
AvgTenureSize (log)	-0.170	-0.193	-0.198	-0.193	-0.183	-0.210
	(0.023)**	(0.013)**	(0.024)**	(0.025)**	(0.025)**	(0.025)**
SameBU x BUSize (log)				-0.361		
				(0.029)**		
SameSubfunction x SubfunctionSize (log)					-0.493	
					(0.030)**	
SameOffice x OfficeSize (log)						-0.511
						(0.050)**
Constant	-7.214	-8.859	-7.167	-9.159	-9.042	-7.640
	(0.451)**	(0.353)**	(0.451)**	(0.473)**	(0.480)**	(0.472)**
Observations	665137	665137	665137	665137	665137	665137

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%

Table 5: Regression results by level.

	Model 1 Middle Mgrs		Model 2 Execs
SameBU	2.401		2.110
	(0.036)**		(0.263)**
BothCHQ	0.016		0.159
	(0.054)		(0.435)
OneCHQ	0.698		0.397
	(0.036)**		(0.330)
SameFunction	1.005		1.724
	(0.029)**		(0.220)**
SameSubfunction	1.196		
	(0.036)**		
SameOffice	1.132	<<>>	-0.212
	(0.080)**		(0.393)
Distance (log)	-0.164	<<>>	-0.329
	(0.010)**		(0.077)**
SameBand	0.224	<<>>	-0.500
	(0.029)**		(0.233)*
SameGender	0.052		0.117
	(0.031)		(0.251)
BothFemale	0.294		0.988
	(0.044)**		(0.448)*
TenureWithin5	0.246	<<>>	0.794
	(0.033)**		(0.208)**
InSample (log)	1.480		0.964
	(0.028)**		(0.249)**
OutOfSample (log)	0.038		0.340
	(0.026)		(0.235)
AvgBUSize (log)	-0.592		-0.849
	(0.022)**		(0.152)**
AvgFunctionSize (log)	-0.133	<<>>	0.045
	(0.028)**		(0.141)
AvgSubfunctionSize (log)	-0.183		
	(0.025)**		
AvgOfficeSize (log)	-0.214	<<>>	-0.214
	(0.018)**		(0.142)
AvgBandSize (log)	0.265	<<>>	-0.784
	(0.085)**		(0.130)**
AvgTenureSize (log)	-0.205		-0.572
	(0.046)**		(0.439)
Constant	-8.954	<<>>	-0.870
	(0.864)**		(2.459)
Observations	553,816		4,308

Robust standard errors in parentheses
*** significant at 5%; ** significant at 1%**
<<>> differences between models significant at 1%
Group size controls account for the different populations

Table 6: Results of individual-level regressions on propensity to span organizational boundaries. Note that the Average Improbability column assembles coefficients from the four previous columns.

	Average Improbability Measure				Average Improbability	Structural Constraint
	ex-Gender	ex-BU	ex-Function	ex-Band		
Female	0.096	0.168	0.143	0.160	0.096	-0.011
	(0.010)**	(0.008)**	(0.008)**	(0.010)**	(0.010)**	(0.003)**
CHQ	0.033	0.131	-0.010	0.030	0.131	-0.002
	(0.011)**	(0.010)**	(0.009)	(0.011)**	(0.010)**	(0.004)
GenExecMgmt	1.097	0.975	0.618	1.193	0.618	-0.024
	(0.108)**	(0.075)**	(0.120)**	(0.110)**	(0.120)**	(0.026)
Mktg	0.841	0.683	0.556	0.898	0.556	-0.023
	(0.025)**	(0.019)**	(0.023)**	(0.026)**	(0.023)**	(0.008)**
Sales	0.526	0.472	0.435	0.574	0.435	-0.049
	(0.013)**	(0.011)**	(0.011)**	(0.014)**	(0.011)**	(0.004)**
Services	0.315	0.359	0.327	0.300	0.327	0.022
	(0.011)**	(0.010)**	(0.010)**	(0.012)**	(0.010)**	(0.004)**
MiddleMgr	-0.355	-0.411	-0.371	0.152	0.152	-0.090
	(0.021)**	(0.019)**	(0.019)**	(0.021)**	(0.021)**	(0.009)**
Band 11	0.070	-0.038	0.015	0.282	0.282	-0.166
	(0.036)	(0.030)	(0.031)	(0.038)**	(0.038)**	(0.011)**
Band 12	0.216	0.064	0.138	0.290	0.290	-0.184
	(0.039)**	(0.029)*	(0.035)**	(0.040)**	(0.040)**	(0.013)**
Band 13	0.625	0.479	0.340	0.481	0.481	-0.163
	(0.085)**	(0.079)**	(0.086)**	(0.093)**	(0.093)**	(0.015)**
Band 14	-0.693	-0.772	-0.694	-0.434	-0.434	-0.044
	(0.110)**	(0.077)**	(0.124)**	(0.121)**	(0.121)**	(0.040)
InSample (000s)	0.003	0.009	-0.002	0.010		-0.028
	(0.002)	(0.002)**	(0.002)	(0.002)**		(0.001)**
OutOfSample (000s)	-0.037	-0.044	-0.034	-0.029		-0.012
	(0.004)**	(0.003)**	(0.003)**	(0.004)**		(0.002)**
Constant	1.780	1.668	1.752	0.959		0.646
	(0.024)**	(0.022)**	(0.022)**	(0.025)**		(0.019)**
Observations	14445	14445	14445	14445	14445	14445
R-squared						0.15

Robust standard errors in parentheses
*** significant at 5%; ** significant at 1%**
"Average Improbability" models estimated using fractional logit

Table A1: Results are robust to numerous alternate assumptions regarding the data.

	Baseline	Weaker Ties	Stronger Ties		Work Ties		Full
	Model	10 recips	1-to-1	12+ messages	Attachments	Co-Recip	Sample
SameBU	2.410	2.460	2.403	2.590	2.466	1.656	2.245
	(0.033)**	(0.031)**	(0.051)**	(0.083)**	(0.057)**	(0.076)**	(0.019)**
BothCHQ	0.003	0.160	0.080	0.565	0.046	0.982	-0.310
	(0.049)	(0.044)**	(0.073)	(0.098)**	(0.093)	(0.097)**	(0.031)**
OneCHQ	0.724	0.862	0.752	1.016	1.002	0.909	0.410
	(0.033)**	(0.034)**	(0.058)**	(0.092)**	(0.064)**	(0.093)**	(0.018)**
SameGender	0.055	0.077	0.091	0.078	0.094	0.004	0.087
	(0.028)	(0.026)**	(0.041)*	(0.072)	(0.055)	(0.051)	(0.015)**
BothFemale	0.285	0.250	0.207	0.230	0.374	0.226	0.237
	(0.041)**	(0.032)**	(0.065)**	(0.097)*	(0.065)**	(0.082)**	(0.021)**
SameFunction	1.031	0.976	1.086	1.148	1.065	1.044	1.046
	(0.027)**	(0.024)**	(0.035)**	(0.065)**	(0.049)**	(0.070)**	(0.015)**
SameSubfunction	1.191	1.191	1.188	1.395	1.310	1.320	1.159
	(0.034)**	(0.030)**	(0.043)**	(0.080)**	(0.059)**	(0.063)**	(0.018)**
SameOffice	1.158	1.064	0.931	1.063	0.948	0.879	0.808
	(0.082)**	(0.062)**	(0.101)**	(0.159)**	(0.136)**	(0.158)**	(0.040)**
Distance (log)	-0.161	-0.173	-0.191	-0.170	-0.189	-0.149	-0.201
	(0.010)**	(0.007)**	(0.012)**	(0.018)**	(0.016)**	(0.024)**	(0.005)**
SameBand	0.264	0.274	0.202	0.259	0.070	0.382	0.261
	(0.029)**	(0.025)**	(0.043)**	(0.074)**	(0.059)	(0.052)**	(0.016)**
TenureWithin5	0.270	0.267	0.329	0.333	0.292	0.119	0.280
	(0.029)**	(0.025)**	(0.037)**	(0.061)**	(0.053)**	(0.051)*	(0.015)**
InSample (log)	1.483	1.662	1.497	1.785	1.686	1.362	1.571
	(0.024)**	(0.024)**	(0.036)**	(0.055)**	(0.040)**	(0.056)**	(0.015)**
OutOfSample (log)	0.050	0.013	0.050	0.017	-0.009	-0.059	0.024
	(0.023)*	(0.022)	(0.027)	(0.051)	(0.037)	(0.042)	(0.013)
AvgBUSize (log)	-0.577	-0.615	-0.561	-0.733	-0.711	-0.648	-0.564
	(0.020)**	(0.019)**	(0.038)**	(0.054)**	(0.039)**	(0.028)**	(0.012)**
AvgFunctionSize (log)	-0.130	-0.109	-0.191	0.016	-0.139	-0.320	-0.130
	(0.025)**	(0.021)**	(0.036)**	(0.041)	(0.043)**	(0.046)**	(0.014)**
AvgSubfunctionSize (log)	-0.150	-0.135	-0.143	-0.276	-0.199	0.080	-0.237
	(0.023)**	(0.020)**	(0.032)**	(0.048)**	(0.037)**	(0.040)*	(0.014)**
AvgOfficeSize (log)	-0.214	-0.210	-0.218	-0.239	-0.228	-0.112	-0.195
	(0.018)**	(0.016)**	(0.022)**	(0.049)**	(0.032)**	(0.031)**	(0.010)**
AvgBandSize (log)	0.003	0.064	-0.002	0.002	0.063	0.397	-0.037
	(0.028)	(0.025)**	(0.045)	(0.046)	(0.067)	(0.130)**	(0.015)*
AvgTenureSize (log)	-0.198	-0.174	-0.149	-0.019	-0.161	-0.177	-0.154
	(0.024)**	(0.023)**	(0.042)**	(0.066)	(0.051)**	(0.051)**	(0.018)**
Constant	-7.167	-9.077	-6.526	-10.396	-6.144	-9.815	-6.272
	(0.451)**	(0.453)**	(0.697)**	(1.187)**	(0.827)**	(1.217)**	(0.292)**
Observations	665,137	857,283	382,076	135,612	211,171	665,137	2,115,211

Robust standard errors in parentheses
* significant at 5%; ** significant at 1%
Group size controls are different for each model, to account for the different populations

Table A2: Baseline models estimated across both e-mail and calendar data. There are virtually no substantive differences across the two media.

	Middle Managers		Execs	
	E-mail	Calendar	E-mail	Calendar
SameBU	2.402	2.320	2.080	2.763
	(0.037)**	(0.074)**	(0.264)**	(0.598)**
BothCHQ	0.053	0.245	0.231	-1.194
	(0.057)	(0.081)**	(0.432)	(0.859)
OneCHQ	0.727	0.580	0.422	0.037
	(0.037)**	(0.069)**	(0.328)	(0.591)
SameGender	0.061	-0.037	0.140	-0.441
	(0.032)	(0.050)	(0.255)	(0.424)
BothFemale	0.273	0.187	0.957	0.858
	(0.045)**	(0.080)*	(0.445)*	(0.628)
SameFunction	1.017	1.080	1.731	0.952
	(0.030)**	(0.049)**	(0.221)**	(0.367)**
SameSubfunction	1.191	1.354		
	(0.037)**	(0.057)**		
SameOffice	1.131	1.100	-0.217	5.380
	(0.083)**	(0.150)**	(0.392)	(1.391)**
Distance (log)	-0.163	-0.168	-0.330	0.456
	(0.010)**	(0.017)**	(0.077)**	(0.209)*
SameBand	0.214	0.267	-0.477	0.754
	(0.030)**	(0.059)**	(0.232)*	(0.366)*
TenureWithin5	0.246	0.357	0.793	-0.473
	(0.034)**	(0.049)**	(0.208)**	(0.466)
InSample (log)	1.661	1.716	1.185	2.151
	(0.029)**	(0.044)**	(0.250)**	(0.445)**
OutOfSample (log)	-0.028	0.049	0.252	0.166
	(0.026)	(0.045)	(0.224)	(0.429)
AvgBUSize (log)	-0.615	-0.585	-0.867	-0.453
	(0.022)**	(0.043)**	(0.151)**	(0.298)
AvgFunctionSize (log)	-0.117	-0.018	0.046	-0.071
	(0.029)**	(0.044)	(0.141)	(0.300)
AvgSubfunctionSize (log)	-0.175	-0.226		
	(0.025)**	(0.041)**		
AvgOfficeSize (log)	-0.210	-0.208	-0.230	-0.272
	(0.019)**	(0.031)**	(0.142)	(0.210)
AvgBandSize (log)	0.221	0.459	-0.734	-0.668
	(0.088)*	(0.244)	(0.130)**	(0.316)*
AvgTenureSize (log)	-0.194	-0.118	-0.575	-2.562
	(0.047)**	(0.085)	(0.441)	(0.790)**
Constant	-9.688	-12.979	-2.282	-3.936
	(0.891)**	(2.382)**	(2.513)	(4.824)
Observations	553,816	213,411	4,308	1,497
Correlation	0.9963		0.5630	

Robust standard errors in parentheses
*** significant at 5%; ** significant at 1%**
Group size controls are different for each pair of models, to account for the different populations

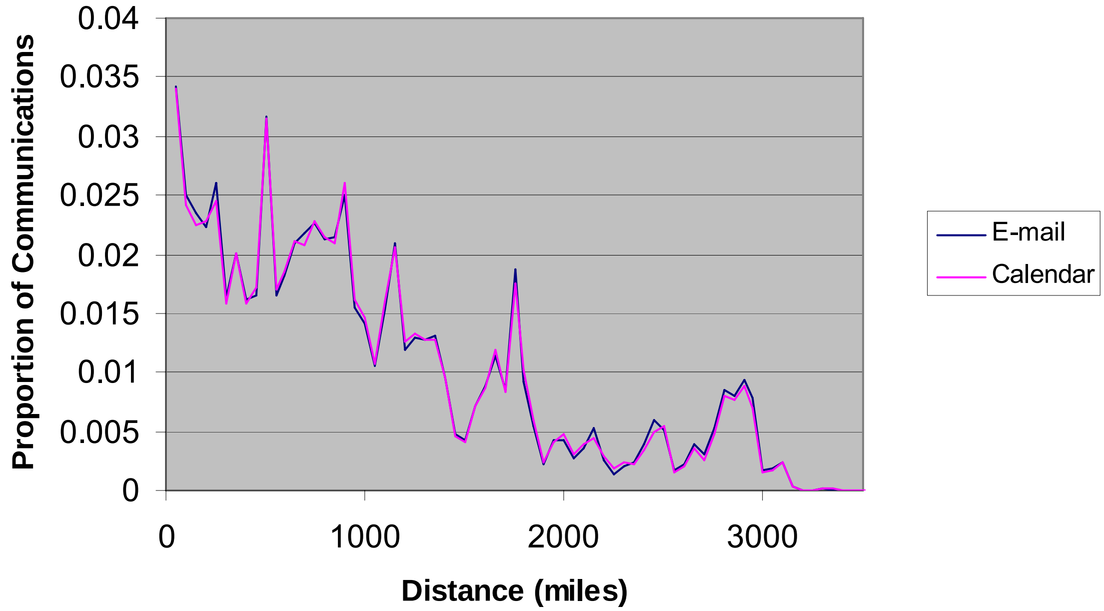

Figure A1: Effect of distance on frequency of dyadic communication by e-mail and meetings.