Organization Science
Vol. 24, No. 5, September–October 2013, pp. 1316–1336
ISSN 1047-7039 (print) | ISSN 1526-5455 (online)

informs

http://dx.doi.org/10.1287/orsc.1120.0804

Discretion Within Constraint: Homophily and Structure in a Formal Organization

Adam M. Kleinbaum
Tuck School of Business, Dartmouth College, Hanover, New Hampshire 03755, adam.m.kleinbaum@tuck.dartmouth.edu

Toby E. Stuart
Haas School of Business, University of California, Berkeley, Berkeley, California 94720, tstuart@haas.berkeley.edu

Michael L. Tushman
Harvard Business School, Boston, Massachusetts 02163, mtushman@hbs.edu

Homophily in social relations results from both individual preferences and selective opportunities for interaction, but how these two mechanisms interact in large, contemporary organizations is not well understood. We argue that organizational structures and geography delimit opportunities for interaction such that actors have a greater level of discretion to choose their interaction partners within business units, job functions, offices, and quasi-formal structures. This leads us to expect to find a higher proportion of homophilous interactions within these organizational structures than across their boundaries. We test our theory in an analysis of the rate of dyadic communication in an email data set comprising thousands of employees in a large information technology firm. These findings have implications for research on homophily, gender relations in organizations, and formal and informal organizational structure.

Key words: social networks; homophily; informal structure; organizational structure; gender
History: Published online in *Articles in Advance* February 12, 2013.

Introduction

Social interactions are notoriously homophilous (McPherson et al. 2001). Across a range of relationship types (e.g., Lazarsfeld and Merton 1954, Fischer 1982, Marsden 1988, Ruef et al. 2003, Goodreau et al. 2009) and diverse empirical contexts (e.g., South et al. 1982, Shrum et al. 1988, Reagans 2005, Marmaros and Sacerdote 2006, Kossinets and Watts 2009), research has demonstrated that people associate most often and most strongly (Reagans 2011) with others who are similar to themselves. Theory and empirical evidence point to two distinct mechanisms that promote homophily: choice homophily, or the preferences of actors to affiliate with similar others, and induced homophily, which results when people find themselves in situations in which they are surrounded disproportionately by others like themselves (McPherson and Smith-Lovin 1987).

In the organization theory literature, scholars have observed the uneven distribution of members of social groups across jobs and ranks in organizations and have argued that homophily tends to subtly reinforce social stratification by providing more beneficial social capital to members of majority groups (Baron et al. 1986, Bielby and Baron 1986, Ibarra 1993, Ridgeway 1997, McGuire 2002, Singh et al. 2010). This is particularly so because social homogeneity is thought to facilitate predictability and trust, and therefore it leads to homosocial reproduction of memberships in empowered groups (Kanter 1977). But at least two significant theoretical gaps remain in our understanding of homophily in organizations. First, although we know that homophily results both from the choices of individual actors and from the constraints imposed by the demographic makeup of organizational subunits, our understanding of how these two mechanisms interact in an organizational context, with its attendant task requirements, is not well understood (Reagans 2011). Second, we know little about how the interplay of homophily, organizational structure, and communication patterns differ between men and women in contemporary organizations.

In this paper, we argue that interaction patterns are strongly influenced by a firm's organizational structure and by its geography: unsurprisingly, people are far more likely to interact if they are assigned to the same business unit, job function, or office building, as well as if they share overlapping affiliations in work groups and other quasi-formal structures. But within the constraints established by organizational structures and physical locations, actors often have discretion to exercise choice homophily. We argue that much homophily arises from discretionary choice within the boundaries of the firm's formal and quasi-formal structures. These structures script the set of interactions that are necessary to carry out the business of the organization, and they determine which interactions are likely to arise out of the convenience of physical proximity. Therefore, the organization bridges the mechanisms of homophily by establishing the consideration sets within which choices are exercised.

Furthermore, we extend this theory in two ways. First, we argue that larger organizational structures afford greater freedom in selecting communication partners because they imply larger consideration sets, which will result in a higher level of choice homophily. Therefore, larger groups impose fewer constraints on who communicates with whom. As a result, we expect that the within-group homophily effect that we predict will be larger in magnitude in large groups than in small groups. Second, we argue that the role of geography differs theoretically from that of organizational structure: whereas organizational structure defines the boundaries of one's social sphere beyond a given geographic place, offices create convenience samples of local interaction partners. We argue that within-office interactions are much more likely to be infused with social content than are other communications within companies, and the preference for homophily will be strongest for socially laden exchanges.

To measure homophilous interaction, we study the incidence of same-gender communication in electronic mail data. The results reveal a consistent pattern of interaction effects in which same-gender communications occur at a significantly higher rate in dyads assigned to the same business unit, job function, or office, relative to those dyads that do not. Upon deeper inspection, we find these effects to be strongest in large business units, job functions, and offices. As an extension to our analysis, we explore gender-specific differences in communication behavior. We find that the networks of both men and women exhibit some patterns consistent with our theory of discretion within the constraints given by formal structures, but we also highlight significant gender differences that merit future research.

These findings have implications for research on homophily, gender in organizations, and organizational design. First, research has long distinguished opportunities for interaction from a preference for within-group ties as alternative mechanisms that generate homophily (McPherson and Smith-Lovin 1987). Echoing this, Lin (2001, p. xi) prefaces his book on social capital with the claim that "central to sociology is the analysis of both action and structure: choice behaviors in the context of structural opportunities and constraints." But the parameters of opportunity and choice in organizations have remained underspecified. In large, multibusiness firms, it is business unit, functional, and office boundaries that most strongly influence the opportunity set of potential interaction partners for organizational actors (Han 1996, Kleinbaum et al. 2009). Within the consideration sets defined by these constraints, discretionary choice gives rise to the highest incidence of homophilous interaction. Second, we contribute to the literature on gender in organizations by offering more current and detailed evidence concerning gender differences in network structures. Our findings cast some doubt on conventional wisdoms regarding gender differences in social network structure in current-day organizations. Third, we contribute to the literature on organizational design, offering evidence of how and by whom formal lateral structures bridge an otherwise siloed organization (e.g., Gulati and Puranam 2009, Soda and Zaheer 2012, Tzabbar et al. 2010, Yakubovich and Shekshnia 2010).

Homophily in Organizational Settings

In a range of relationship types and across a diverse set of contexts, researchers have demonstrated that people associate more with others who are similar to themselves. Why is this so? One possibility is that actors have an underlying psychological preference to interact with others who are like themselves. There is evidence of such choice homophily in friendship networks among children (Shrum et al. 1988), college students (Marmaros and Sacerdote 2006, Lewis et al. 2008), and adults (Lazarsfeld and Merton 1954); confiding networks among adults (Marsden 1988); social support networks in the government (South et al. 1982); interaction networks among coreligionists (Fischer 1982); and co-founding networks among entrepreneurs (Ruef et al. 2003), to cite a few among myriad examples (cf. Ingram and Morris 2007).

But a taste for homophilous interaction is not the only reason why we observe heightened interaction rates between members of the same social categories. Because they often share interests or possess comparable backgrounds, similar people often sort (or are sorted) into similar situations. As a result, they find themselves in places, groups, or positions, such as jobs (Bielby and Baron 1986, Kaufman 2010), college courses (Kossinets and Watts 2009), neighborhoods (Laumann 1966), or voluntary organizations (McPherson and Smith-Lovin 1987), that are disproportionately populated with others like themselves. Therefore, even if people choose interaction partners without regard to membership in social categories, we are still likely to observe elevated interaction rates among demographically similar people in consequence of the patterned distribution of individuals with similar demographic characteristics across time, space, social positions, and social roles. When interests are relatively homogeneous within groups and serve to focus social relations, they produce structurally induced homophily (Feld 1981, McPherson and Smith-Lovin 1987).

Although the notion that homophily is both chosen and induced is not novel, our contribution lies in teasing apart induction from choice in the context of a complex, formal organization. Much of the relevant literature has examined homophily in society, rather than within business organizations. For example, Bossard's (1932) classic study examined the propinquity effect on spouse selection in Philadelphia. More recently, Aral et al. (2009) examined the role of homophily in driving coadoption of technology services by friends, and a recent spate of studies has examined homophily in educational settings (Marmaros and Sacerdote 2006, Lewis et al. 2008, Goodreau et al. 2009, Kossinets and Watts 2009). This literature provides insight into the mechanisms and consequences of homophilous interaction, but findings may not generalize to business organizations, where the existence of elaborate task and authority structures often prescribe patterns of interaction.

There has been, of course, influential research on homophily in corporate settings (e.g., Ibarra 1993, Ibarra and Smith-Lovin 1997). However, the emphasis in this work has not been on how organizational structures interact with individual preference to contour the aggregate patterns of homophily within organizations. Illustrative, for instance, is Blau's (1994, p. 130) seminal work, in which he discusses "the internal structure of organizations, conceptualized as the distribution of their employees among official positions along various lines," but in which he offers no specificity about what positions or lines serve to structure interaction opportunities. Indeed, although we may have strong intuitions about how formal structures and individual choice may come together in an organizational context, we have few empirical results on which to rely. In consequence, this work tends not to consider the potential influence of individual preference and how it may operate in conjunction with organizational mandates to determine the level of homophilous interaction. In short, we lack an integrated theory of homophily in organizations that accounts for both opportunity structure and choice in determining communication patterns.

The Mechanisms of Homophily in Organizational Settings

For individual preference to be a significant source of homophily in communication relationships within organizations, actors must have latitude to choose their interaction partners. To what extent do they have this in present-day organizations? And, across what dimensions of organizational structure are individuals most likely to have the greatest discretion to choose their contacts? After all, the formal structure of the organization is designed to execute a set of tasks, and therefore the structure itself induces a great deal of interaction. The question we explore is, where within the organization is discretion, as manifest in homophilous interaction, most apparent, given the multitude of interactions that are induced by the organization's formal and quasi-formal structures? We rely on several, classic lines of organization theory to describe the conditions under which actors are more likely to be free to choose with whom to interact.

The earliest research on this question dates back to Taylor (1911). Taylor and fellow adherents to "scientific management" drew a distinction between workers and managers and proposed that, to maximize efficiency, managers should reduce the amount of discretion workers are able to exercise when performing their jobs. Subsequent perspectives, however, suggest that even industrial work must contain certain discretionary tasks (Turner and Lawrence 1965). More recent scholarship goes further, emphasizing that giving workers a greater level of discretion, even in highly scripted work, can lead to greater efficiency (Adler et al. 1999). Since McGregor (1960), research has found that worker autonomy is associated with positive outcomes such as motivation, job satisfaction, and individual performance (Hackman and Oldham 1980). And as we travel across the continuum from worker to manager and from the factory floor to knowledge work, we would expect that actors will have greater discretion in choosing with whom to communicate. Indeed, knowledge work often depends on the sharing of information and know-how that is distributed across an organization (Kogut and Zander 1992).

Within any organizational context, however, the discretion to select communication partners occurs within an organizationally defined choice set. The task structure of the firm, embodied in its formal organizational structure and geographic configuration (Galbraith 1973, Tushman and Nadler 1978), forms the backbone of the communication structure of the company. Employees often must interact with specific others to complete the task requirements of their jobs, but the degree to which this is true is likely to depend on the nature of the task and the costs of interaction. Allen (1977), building on the work of Thompson (1967), argues that this assertion is true by design. He writes, "The real goal of formal organization is the structuring of communication patterns" (Allen 1977, p. 211). In the typical, modern, complex organization, four types of structures will pattern communications: business units, functional units, geographic units, and quasi-formal structures, such as project teams and task forces. We discuss each in turn. Put in general terms, we posit that organizational members will have more latitude to choose with whom they interact *within* each of these four types of structures, relative to the communications that occur between two people who are in different organizational units. This will occur because people within units typically have greater knowledge of the alternatives with whom they could interact to complete a task, and any preference for homophilous interaction will manifest in the selection of a specific person from among the set of (relatively) redundant alternative options. Given the wider scope for choice in within-unit relative to cross-unit interactions, we expect greater homophily among the former set of interactions.

Business Units. 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. In this view, individuals whose task requirements necessitate reciprocal interdependence are organizationally colocated within a task-oriented business unit, which minimizes the costs of coordination within the organization (Thompson 1967, Galbraith 1973). Therefore, the business units of a multidivisional firm are designed to be largely autonomous of one another, with interactions concentrated within, rather than between, them (Galbraith 1973, Williamson 1975).

When actors communicate across business units, we suspect that such communications are likely to be episodic and driven by a narrow, nonrecurrent set of task requirements. This suggests that individuals will be less likely to know a broad set of colleagues with the authority, responsibility, or expertise for the task at hand. Because these cross-unit interactions are more likely to be narrowly prescribed by the formal task responsibility, and because knowledge of the set of potential collaborators is limited, individuals often will communicate with a specific alter, rather than choose someone from among a set of possible, redundant exchange partners. Given that their choice set may be limited to those relatively few people whom they happen to know, they have less discretion in choosing their interaction partners in cross-unit communications. We therefore expect to observe less choice homophily across business unit boundaries than will occur within them. We hypothesize the following.

HYPOTHESIS 1. *The rate of homophilous interaction will be higher for dyads in which members are employed in the same* business unit *than for dyads in which members work in two different business units.*

Functional Units. Within multidivisional firms, business units are further subdivided along functional lines (Galbraith 1973, Hrebiniak and Joyce 1984). Functions serve two distinct purposes: lateral linking and sharpening the division of labor. The first objective of the organization of work into job functions is to provide a locus for interaction across business unit boundaries. Many organizations promote cross-business-unit, within-function sharing of best practices (Galbraith 1994). A second purpose of job functions is to create further specialization within each business unit, narrowing the range of tasks performed by each person and colocating the most reciprocally interdependent tasks within a smaller partition of a larger business unit. In the process, job functions also sharpen the set of relevant interaction partners for each person. Thus, like business units, job functions prioritize interactions within their boundaries, relative to cross-functional interactions.

Job functions, however, differ from business units in one important respect concerning their potential influence on the incidence of within- versus across-organizational-unit interaction. Like business units, job functions are a structural means to achieve a division of labor; however, unlike business units, job functions are explicitly designed to be interdependent (Thompson 1967, Williamson 1975). In a typical organization, it would be reasonable to expect a higher level of cross-function-unit than cross-business-unit interaction. Nevertheless, we believe that most cross-functional interactions remain formalized, with interfaces that are prescribed by the design of the organization (Galbraith 1973). Thus, despite a greater level of theoretical interdependence between job functions, we still expect that, as in business units, job functions will serve to focus interactions within their boundaries. We anticipate a larger set of potentially redundant interaction partners within, relative to across, functional boundaries. Therefore, as with business units, we expect that individuals will have greater discretion to select homophilous communication partners in their within-function, relative to their across-function, communications.

HYPOTHESIS 2. *The rate of homophilous interaction will be higher for dyads in which members are employed in the same* job function *than for dyads in which members work in two different job functions.*

Quasi-Formal Organizational Structures. Of course, business unit and job function are the major categories of formal structure that organize people and tasks within organizations, but in contemporary organizations, there are myriad less permanent and less formal subdivisions and structures that further shape interaction patterns. For example, project-based work groups, committees, task forces, and line reporting structures are a few of the many substructures that mold interactions within the suprastructures of business unit, function, and office (Galbraith 1973). Such quasi-formal structures have eluded much systematic research because of the fluid, organic way in which they are composed and the challenge this has posed for researchers to track and record their appearance, disappearance, and shifting rosters of members. Nevertheless, such structures play an important role in driving intraorganizational communications.

Within the context of the quasi-formal structure and its raison d'être, interactions may be prescribed. But participation in such structures allows individuals with little other basis for interaction to meet, thus creating opportunities for discretionary interaction in other contexts that are not available to otherwise similar dyad members who lack such quasi-formal affiliations (Reagans et al. 2004). Such discretionary interactions will, other things being equal, be subject to choice homophily. Therefore, we hypothesize the following.

HYPOTHESIS 3. *The rate of homophilous interaction will be higher for dyads in which members share affiliations in* quasi-formal organizational structures *than for dyads in which members do not.*

Geographic Units. In addition to formal and quasi-formal organizational structure, there is abundant evidence of spatial effects in the formation of social relationships. Studies have found that ties are much more likely when two individuals live or work near one another (e.g., Zipf 1949, Festinger et al. 1950, Blau and Schwartz 1984, Kono et al. 1998, Sorenson and Stuart 2001). This is true of geographic space, of functional spaces within physical structures, and of microspaces within buildings (Marmaros and Sacerdote 2006, Liu 2010). In fact, despite rampant speculation that the proliferation of electronic communication will herald "the death of distance" (Cairncross 2001), the evidence on the issue contradicts the view that modern communication technologies have dramatically reduced the impact of geographic proximity on the likelihood of interaction (Marmaros and Sacerdote 2006, Mok et al. 2010).

Whereas business units, job functions, and quasi-formal structures focus interaction by prescribing specific job tasks, geographic proximity operates differently: it creates a convenience sample of possible exchange partners. Indeed, Zipf (1949) referred to the mechanism behind the proximity effect as "the principle of least effort," which underscores that the lowest cost interactions tend to be among colocated individuals. In formal organizational contexts, geographic colocation is a residual category of social organization. It may coincide with business unit and functional memberships, as organizations often choose to geographically group individuals who share common structural units. After accounting for affiliations to particular organizational units, however, colocation captures the ease of interaction.[1] Net of common organizational affiliations, we expect that colocated individuals will have a high degree of discretion in selecting interaction partners.

HYPOTHESIS 4. *The rate of homophilous interaction will be higher for dyads in which members are employed in the same* office location *than for dyads in which members are employed in two different offices.*

Indeed, there is reason to expect that the *greatest* levels of homophily will occur within office boundaries. This is because of the nature of within-office ties: relative to other interactions that occur within organizations, we conjecture that intraoffice interactions are more likely to be informal and infused with social content. Why? Coffee and lunch breaks, casual banter, office and company gossip, and so forth all are forms of interaction that are greatly facilitated by physical proximity. Communications that are purely social in nature are indications of what Allen (1977) calls neutral social interactions: even if these interactions are not themselves generating productive output for the company, they indicate to the analyst—and reaffirm to the individuals themselves—an existing interpersonal relationship that makes each person a potential candidate to help the other person meet her discretionary informational needs (Kleinbaum 2012). Although the myriad incidental interactions that occur within offices begin as social ties, ultimately many of these connections become components of the productive effort of the enterprise.

Regardless of where they fall on the continuum between social and work communications, we suspect that, as a proportion of total communications, interactions of a social nature are more prevalent within geographic office spaces than across them. In establishing informal interactions of this nature, individuals are relatively unconstrained by the organization's formal task structure relative to when their interactions are strictly task based. This leads us to postulate the following.

HYPOTHESIS 5. *The rate of homophilous interaction for dyads in the* same office *will exceed the comparable rates for homophilous dyads in the same business unit, job function, or quasi-formal structures.*

Theoretical Extension: Group Size. In Hypotheses 1–4, we identified four organizational boundaries and argued that individuals will have greater discretion in the choice of partners when they are interacting within boundaries relative to when they are communicating across them. We extend this argument by positing that within each of these organizational units, the level of discretion in the choice of communication partners will be greater in large groups than in small groups. In each case, our argument is that, to a large extent, the formal and quasi-formal structures in large organizations determine the boundaries of individuals' social spheres: within these structures, people have many, often redundant, contacts; across them, interaction is less prevalent and less open to discretionary choice. If this line of reasoning is correct, we would further anticipate that one's ability to select into homophilous exchanges will be strongest within larger organizational units, for the simple reason that there is a greater availability of potential contacts from whom to select. Consequently, our theory suggests that the homophilous communication premium that we hypothesize to exist within business units, job functions, and offices should be more pronounced within large organizational units relative to small ones.[2]

HYPOTHESIS 6. *The effect of homophily on the rate of interaction will be greater in magnitude in large business units, job functions, and offices than in small business units, job functions, and offices.*

Data and Methods

Sample and Data Collection

Data for this study were collected from "BigCo," a large information technology and electronics company. BigCo has 29 product divisions, organized into four primary product groups: hardware, software, technology services, and business services. Overlaid across the business unit structure is a formal lateral organization, in which each person is also assigned a job function. Within this formal structure, the employees of BigCo are widely dispersed geographically, with a relatively loose coupling between formal structure and geography.

The data we analyzed include the complete record, as drawn from the firm's servers, of email communications among 30,328 employees. These data are well suited to test our hypotheses. First, because we can collect electronic communication data for large numbers of individuals at low cost, we can explore the determinants of homophilous interaction in a larger, more complex organization than those studied previously. Given our interest in the influence of organizational structure on shaping interaction, greater insights are possible from the study of a multibusiness, multifunction, multioffice organization (cf. Johnson et al. 2012). Second, by measuring actual communication, rather than self-reports of friendship, social, or instrumental ties, we are able to observe homophilous interactions directly. Thus, the data are not filtered by the subjective perceptions of survey respondents (Bernard et al. 1981, Quintane and Kleinbaum 2011).

All internal email information that was on the server at the time of data collection, spanning an observation period from September 2006 to December 2006, 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, BigCo stripped all messages of their content, leaving only the metadata (e.g., sender and recipient, both encrypted to protect individual privacy; time stamp). We consolidated these files and expanded each multiple-recipient message to include one entry for each unique dyad. The final file contains 114 million dyadic communications. In the core models, we excluded "blind carbon copy," or Bcc, recipients (but results are robust to their inclusion), mass mailings (defined as messages with more than 4 recipients; results are robust to alternative thresholds of 1, 6, or 10 recipients), and direct interactions with administrative assistants. Imposing these screens shrinks the data set by almost an order of magnitude to 13 million emails.[3] For confidentiality reasons, BigCo would not disclose many sociodemographic variables such as employees' race, ethnicity, or age, but the company did provide gender and human resource (HR) information about each employee, which we were able to link to the communication data through encrypted employee identifiers. The HR data include each employee's business unit, major job function and subfunction, salary band, and office location code.

Because it is available to us and because it has been the most studied dimension of homophilous interaction in the organizational theory literature, we used *same gender* as our measure of homophily. Numerous studies of homophily (e.g., Lincoln and Miller 1979, South et al. 1982, Brass 1985, Bielby and Baron 1986, Marsden 1987, McPherson and Smith-Lovin 1987, Shrum et al. 1988, Ibarra 1992, Kalleberg et al. 1996, Ibarra 1997, Ruef et al. 2003, Singh et al. 2010) have focused primarily or exclusively on gender. Similarly, there is an enormous related literature on gender differences in workplace behavior and career outcomes (e.g., Ridgeway 1997, Burt 1998, Fernandez-Mateo 2009, Gibson and Lawrence 2010, Adams and Funk 2012).

Overall, BigCo's nonadministrative U.S. workforce is 69.9% male and 30.1% female.[4] The four major product groups of the company are similar in their gender composition, ranging from 25% to 28% female. The corporate sales organization and corporate headquarters have higher proportions of women than other units, at 32% and 39%, respectively. There is also some gendered sorting into job functions: in addition to administration, women are overrepresented in finance and form a majority of employees in the communications and human resource functions. Conversely, men are overrepresented in general executive management and in research and development. The two largest functions within the company, sales and services, have gender distributions similar to the company as a whole. The proportion of women at BigCo decreases with increasing rank.[5]

The full sample contains 24% of BigCo's total U.S. employee population[6] but differs from that population in several respects. Therefore, the possibility exists 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 sampling issues, we exploit our large sample size and our knowledge of the firm's population of U.S.-based employees to create a stratified random subsample of employees. Our subsampling approach, which maximizes the correspondence between the subsamples we draw and a set of population parameters, yields a subsample consisting of 15,240 employees. To assemble the representative subsample, we created a three-dimensional matrix of salary band (rank and file; middle managers; and each of band 11, 12, 13, and 14), function (general executive management, marketing, sales, services, and everyone else), and business unit (corporate headquarters and everyone else). For each of the 60 cells of this $6 \times 5 \times 2$ matrix, we calculated the sampling probability that would be needed to achieve a subsample rate of 12.1% of the U.S. population of the firm. We chose to make our subsample representative of only selected groups to maintain a large sample size; had we made our subsample representative across all variables, we would have diminished our sample to just 2.9% of the U.S. population of the firm. Once we had these probabilities, we randomly determined whether each person in the overall sample, given her salary band, job function, and business unit, would be included in the subsample. Importantly, the gender distribution of both the full sample and subsample are representative of the employee population. The analyses we will present are based on the more conservative, random subsample, but the findings do not substantively change from the full sample or in various random draws of the subsample.

Estimation Approach

Scholars have proposed a number of methods to infer social networks from email data (e.g., Eckmann et al. 2004, Moody et al. 2005, Wuchty and Uzzi 2011). We follow the recent work by Wuchty and Uzzi and employ the "total volume" method: 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 data set with counts of the number of communications within unordered pairs of individuals. Within the emerging literature on inferring social networks from email data, this approach has been shown to reliably reflect individuals' perceptions of relations while balancing low levels of false-positive and false-negative errors (Wuchty and Uzzi 2011). With the time axis compressed so that the data are structured as a single cross section, the communication matrix remains large and sparse. In fact, less than 0.5% of the approximately 116 million possible unordered cells in the email matrix are nonzero. Even given modern computing power, it is not expeditious to work with the full matrix of email communications among the 15,240 members of our quasi-random subsample of individuals.

Random sampling from the set of the 116 million dyads is one potential solution to this problem. However, this approach ignores the fact that the realized (nonzero) ties provide most of the information to identify the parameter estimates (Cosslett 1981, Imbens 1992, Lancaster and Imbens 1996). We therefore construct a "case cohort" data set by including in our dyad-level regression models all nonzero cells and a random sample of zero cells (King and Zeng 2001), which are then weighted according to their probability of being drawn into the analysis sample (e.g., Russell et al. 2001, French et al. 2008). We do not stratify on the sampling of zeros; we simply draw the zero cells at random, with a probability set to generate an approximately 1:1 ratio of zero to nonzero observations (results are robust to alternative samples that include 5:1 and 10:1 ratios of zero to nonzero observations). This weighted sample reflects the complete pseudopopulation of all dyads comprised by two members of our quasi-random subsample of 15,240 individuals.

Our dependent variable is *CommunicationFrequency*, a count of the number of emails exchanged within each dyad.[7] 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[8] (Gouriéroux et al. 1984, Wooldridge 1997, Silva and Tenreyro 2006). Thus, we estimate the likelihood that dyad-level covariates affect the frequency of interaction using models of the form

$$E[Y_{ij} \mid X_{ij}] = \exp((X_{ij} + Z_{ij})\beta), \tag{1}$$

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

We analyze the data as a cross section, forming a dependent variable by pooling email correspondences over a three-month period (October–December 2006). We then use the previous month's data (from September 2006) to construct lagged values for covariates that cannot reasonably be assumed to be exogenous. Given this cross-sectional design, we analyze the incidence of interaction among pairs of employees who are at risk of communicating. Therefore, the regressions estimate frequencies of communication, one dimension of tie strength (Reagans 2011), rather than the rate of formation of ties in as-yet-non-communicating dyads.

In dyad regressions, there is a well-known estimation problem: observations are likely to be nonindependent. In particular, dyad models are prone to two types of nonindependence that potentially can yield misleading results. First, interactions within a dyad are not independent: the number of emails actor i sends to actor j is dependent on the number of emails i receives from j (Quintane and Kleinbaum 2011). In our handling of the data, this problem is avoided because we analyze the total number of messages exchanged within the dyad; that is, the value Y_{ij} includes messages sent from i to j and from j to i; Y_{ji} does not appear in our analysis as a separate observation. The second concern is that each individual in a dyad appears in numerous other dyads, which introduces a common person effect (Kenny et al. 2006); that is, Y_{ij} may be correlated with Y_{ik} because some unobserved attribute of person i affects both values. This problem should not affect the parameter estimates, but it can cause standard errors to be underestimated (Kenny et al. 2006).

We address the nonindependence problem by estimating robust standard errors that are simultaneously clustered on *both* members of a dyad. Cameron et al. (2011) developed this approach theoretically but only implemented it for ordinary least squares and logit regression. Because their approach was more generally robust (Lindgren 2010, Cameron et al. 2011), we develop clus_nway.ado,[9] an implementation of it in Stata that is suitable for other estimators, including the Poisson quasi-maximum likelihood models we employ. As in the work of Cameron et al., standard errors are calculated in three separate, cluster-robust covariance matrices: one by clustering according to i, one by clustering according to j, and one by clustering according to their intersection. Standard errors in the regressions we report, which cluster on both dyad members, are estimated based on the matrix formed by adding the first two covariance matrices and subtracting the third. This approach is functionally similar to using the quadratic assignment procedure's bootstrap approach to adjusting standard errors in multiple regression (MR-QAP), but it can be implemented more quickly in large data sets (Cameron et al. 2011). Likewise, relative to exponential random graph modeling, this approach is feasible on much larger data sets and on networks with nondichotomous ties. Finally, we note that group memberships are nonnested, so employing multilevel models was not necessary.

Independent Variables

The independent variables in our dyad-level regressions are all properties of the ijth 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 1 when i and j are in the same strategic business unit. BigCo has 29 business units that are organized into four business groups; additionally, corporate headquarters and the corporate sales force are treated as business units by the company and in our data. 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 and development, supply chain, sales, services, and a catchall "other" category. 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.

To measure comembership in quasi-formal organizational structures that are not directly observable, we follow Kossinets and Watts's (2009) approach to identifying "implicit social foci." A social focus is any structure that creates interaction opportunities between two people (Feld 1981). Kossinets and Watts (2009, p. 419) defined an implicit social focus to be "social groups, sporting and cultural organizations, shared housing, and so forth," all of which likely play a significant role in shaping interaction patterns, but none of which was directly observable in their data. Similarly, we cannot directly observe membership in quasi-formal organizational structures at BigCo, such as work groups, committees, task forces, and formal reporting structures, but we know that membership in these groups or reporting lines shapes interactions within the company.

To infer the existence of these quasi-structures, we make use of the mass emails in the data. When two people receive the same mass email, we assume it indicates the presence of a common group affiliation (Kossinets and Watts 2009). Coreceipt of a large number of mass emails by two people may suggest that the two are comembers of a particularly active work group, reporting line, or task force, or it may indicate joint affiliation with multiple quasi-formal groups. The *SharedImplicitFoci* variable is measured as the count of mass emails (i.e., emails with more than four recipients) that the two members of each dyad coreceive.

In operationalizing *SharedImplicitFoci* this way, we note two potential problems, which our approach is designed to avoid. First, it would be problematic to estimate the effect of shared social foci if the same emails created both a direct tie between the sender and recipient and a shared implicit focus between corecipients. For this reason, we construct the data so that direct ties are based on non-mass emails, and shared implicit foci are based on the disjoint set of mass emails. A related, conceptual problem of inference would result if a coreceived email (a covariate in our models) occurred chronologically after a direct exchange of email (our dependent variable). To avoid this situation, our *SharedImplicitFoci* variable is based on data from September 2006, the month preceding the observation period of our dependent variable (October–December 2006).

We use two different specifications for gender: in our primary models, we include *SameGender*, set to 1 when i and j either are both male or are both female, and 0 otherwise. Therefore, the regressions we emphasize compare same-sex pairs to female–male dyads. In an extension of the core analysis, we then split the same-gender effect by including separate covariates for *BothMale* and *BothFemale*.

Our first four hypotheses anticipate that homophily will be stronger within formal and quasi-formal organizational units than across those boundaries. To test these predictions, we add to our baseline model interaction terms in successive models of *SameBU*, *SameFunction*, *SameOffice*, and *SharedImplicitFoci*, each by *SameGender*. These specifications test the hypotheses that the difference in communication frequency between same-gender and mixed-gender dyads is equal within, relative to across, units. A positive, significant interaction term is sufficient to reject the null hypothesis and to conclude that gender homophily is stronger within units than between them (for the effect of a similar interaction on dyadic tie strength, see Reagans 2011).

Hypothesis 6 posits that larger groups should afford more discretion, and therefore more latitude for choice homophily, than smaller groups. To test this hypothesis for each of our observable groups—business units, job functions, and office locations—we define variables of the type *LargeFunction* and *SmallFunction*; that is, we specify the group memberships of the dyad members into three mutually exclusive categories: pairs of people who are not in the same group, same-group dyads where the group is small, and same-group dyads where the group is large. Taking the example of job functions, one approach would be to estimate separate models on the small-function and large-function subsamples to see whether the *SameGender* interaction differs in small groups compared with large groups. We provide these results for intuition, but this approach offers no definitive test for whether the difference in effect sizes across models is statistically significant. To test the significance of the difference, we run a third model for each type of group in which the subsample analyzed includes dyads whose members are coemployed in either small or large functions. We estimate the difference in the magnitude of the *SameFunction* effect between small and large functions by including a covariate *LargeFunction*, set to 1 when the job function to which the dyad members are both assigned is large and to 0 when the job function to which both dyad members are assigned is small. Additionally, we include in the model the interaction *LargeFunction* × *SameGender*. If Hypothesis 6 is correct, the coefficients of the *LargeFunction* × *SameGender* interactions will be positive and statistically significant. Analogous variables are calculated for business units and offices; together, these represent the test statistics for Hypothesis 6.

Finally, in addition to examining gender homophily overall, our empirical extension section looks separately at homophily between men and homophily between women. To do this, we drop the covariate *SameGender* and replace it with covariates for the main effects of *BothMale* and *BothFemale* as well as interaction variables of the form *BothMale* × *SameBU* and *BothFemale* × *SameBU*. The former interaction variable tests the hypothesis that the difference in communication frequency between male–male dyads and mixed-gender dyads is greater within business units than across business units. Conversely, the latter interaction variable tests the hypothesis that the difference in communication frequency between female–female dyads and mixed-gender dyads is greater within business units than across business units.

Control Variables

Because the unit of analysis is a potential pair of communicators, all covariates are specified as properties of dyads. The 13 job functions at BigCo are subdivided into 60 subfunctions, which we account for in the regressions with a *SameSubfunction* dummy variable. We also include *Distance* (logged), the natural logarithm of the estimated door-to-door (driving) distance in miles between employee i and employee j's office buildings, plus one mile. The company has a 15-band salary hierarchy ranging from 0 (for employees in training) to 14. We include a *SameSalaryBand* dummy variable to indicate that both members of a dyad are in the same salary band. Finally, we control for the gender distribution of the environment surrounding the dyad by including, for example, *BU%Women*, the cumulative percentage of women in i's and j's business units, as well as similar controls for the gender distributions of i's and j's job functions and home offices. We chose to specify these control variables as percentages rather than as levels to capture effects of gender composition independent of group size. However, results are substantively similar when these controls are respecified as either raw counts or log-transformed counts of the number of women.

In addition to clustering by each dyad member, we also control for communication volumes directly. We include *EmailVolume* (logged), the natural logarithm of 1 plus the number of emails the two actors exchanged with all other (non–$i-j$) partners in our sample to adjust for the fact that the individuals within the sample have differential propensities to communicate. If we excluded these volume controls, our results would indicate additional communications in which men or women engage on the margin, as opposed to shifts in the distribution of a fixed number of communications across potential recipients. (As we report below, there are, in fact, significant gender differences in the volume of communication: women at BigCo send and receive more emails than men do.) Unreported results reveal the same pattern of results in models that exclude communication volume controls. Although we take comfort in the similar findings, we prefer to include volume controls to better account for unobserved heterogeneity in communication behavior.

Results and Discussion

Descriptive statistics appear in Tables 1(a) and 1(b). Table 1(a) describes individual-level email activity, broken down by gender. Surprisingly, we find that women, on average, have a higher total email volume than men. The average woman at the company exchanged 2,850 messages during the observation period, compared with 2,564 for the average man ($p < 0.01$). This higher total results from both a larger number of contacts and a larger average frequency of interaction within each dyad ($p < 0.01$). We also find that women, on average, have a higher proportion of female contacts than

Table 1(a) Descriptive Statistics for Individual-Level Email Patterns by Gender

Gender	Total volume	Average degree	Avg. dyadic frequency	Gender distribution of contacts (%): Overall	Same BU	Across BUs	Same office	Across offices	Same function	Across functions
Male	2,564	38.2	12.4	M: 69.9 F: 30.1	M: 70.6 F: 29.4	M: 67.1 F: 32.9	M: 72.0 F: 28.0	M: 68.8 F: 31.2	M: 71.7 F: 28.3	M: 62.0 F: 38.0
Female	2,850	46.4	13.3	M: 57.8 F: 42.2	M: 57.3 F: 42.7	M: 59.4 F: 40.6	M: 58.8 F: 41.2	M: 57.7 F: 42.3	M: 57.1 F: 42.9	M: 55.2 F: 44.8

Notes. "Total volume" includes both the mass emails used to measure implicit social foci as well as the non-mass emails used to measure direct communication. The rest of the table is based only on the non-mass, non-Bcc emails that comprise our dependent variable. The columns under "Gender distribution of contacts (%)" indicate the gender distribution of the average man's and the average woman's contacts, both within and across business unit (BU), office, and functional boundaries. Univariate *t*-tests indicate that total volume, average degree, average dyadic frequency, and the overall gender distribution of contacts differ significantly between men and women ($p < 0.01$). Additionally, both men's contacts and women's contacts have different gender distributions within (compared to across) business units, job functions, and offices ($p < 0.05$). M, male; F, female.

Table 1(b) Descriptive Statistics for Dyad-Level Variables

	Weighted pseudopopulation		Communicating dyads only		Noncommunicating dyads only	
	Mean	SD	Mean	SD	Mean	SD
CommunicationFrequency	0.0186	1.2832	10.9370	29.0979	0	0
SameBU (%)	17.23	37.76	59.00	49.18	17.15	37.70
SameFunction (%)	31.67	46.52	60.63	48.86	31.62	46.50
SameOffice (%)	1.43	11.88	14.38	35.09	1.41	11.79
SharedImplicitFoci (time 0)	0.0245	0.5854	2.7829	7.7087	0.0197	0.4784
SameGender (%)	57.77	49.39	58.62	49.25	57.77	49.39
BothMale (%)	48.57	49.98	44.50	49.70	48.58	49.98
BothFemale (%)	9.20	28.90	14.12	34.82	9.19	28.89
SameSalaryBand (%)	23.02	42.10	32.47	46.83	23.01	42.09
Distance (logged)	6.373	1.383	5.212	2.568	6.374	1.379
EmailVolume (logged)	6.258	0.814	6.848	0.713	6.257	0.813
BU%Women	36.94	8.61	36.73	9.49	36.94	8.61
JobFunction%Women	13.99	1.96	15.09	4.26	13.98	1.95
Office%Women	36.14	6.73	36.92	7.47	36.14	6.73
No. of observations	113,108,944		198,081		221,854	

Note. Here, statistics are reported separately for the weighted pseudopopulation, which approximates the entire population of dyads; the subsample of communicating dyads; and the subsample of noncommunicating dyads.

men do ($p < 0.01$). Table 1(a) shows gender differences in patterns of homophilous interaction. The average male in BigCo maintains an overall contact distribution that reflects the gender composition of the company almost precisely: 69.9% of his contacts are other men, and 30.1% are women. Conversely, women are overrepresented in the communication networks of women: 42.2% of a typical female's contact network are female, whereas only 30% of BigCo employees are female. Said differently, the average female employee of BigCo has *almost exactly* the same number of men in her network as does the average male employee (difference is not statistically significant): women have 26.6 male contacts on average (57.8% of 46.4), whereas men have 26.7 male contacts (69.9% of 38.2). But the typical female employee also has an average of 19.6 women in her network, whereas the typical male has only 11.5 women in his ($p < 0.01$). Thus, the data reveal that the additional conversations that women initiate are *far* more likely to be with other women.

Table 1(a) also reports the gender composition of interactions within versus across organizational units. The numbers in this table are unadjusted and purely descriptive. The table indicates that the average man in the sample has a gender distribution of contacts that is shifted slightly in favor of men for interactions within his own business unit (70.6%), office location (72.0%), and job function (71.7%) relative to the overall distribution of 69.9% male. Conversely, the average man's interactions across business units, offices, and functions include a slightly lower proportion of same-gender ties. Similarly, the average woman interacts with other women in her own business unit at a slightly higher rate compared with across business units. But when we examine office and function boundaries, the pattern of results reverses. Women have a higher proportion of their cross-office (42.3%) and cross-functional (44.8%)

ties with women relative to their within-office (41.2%) and within-function (42.9%) contacts, though all numbers in this table are unadjusted for risk set differences in the within- versus across-organizational-unit gender composition.

Descriptive statistics of dyad-level variables appear in Table 1(b). The first set of columns shows means and standard deviations for dyad-level variables in the weighted sample, which reflects the entire pseudopopulation of dyads. For comparison, we also show descriptive statistics for the sample of communicating dyads and the random sample of noncommunicating dyads. A correlation matrix for the weighted pseudopopulation appears in Table 2.

Table 3 presents dyad-level Poisson quasi-maximum likelihood regression models of the frequency of email exchange in BigCo. Model 1 provides baseline estimates and shows positive, significant (both statistically and substantively) main effects on communication of comembership in formal structural groups (business units and job functions), geographic units (offices), quasi-formal organizational structures (indicated by shared implicit foci), and social categories (gender). 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 $\exp(1.822) = 6.18$ times the rate of otherwise similar dyads that span different business units. The effects of being in the same function and subfunction are large as well: two individuals in the same function communicate at $\exp(0.755) = 2.13$ times the rate of those who are in different functions, ceteris paribus. Two individuals who also are in the same subfunction communicate at 3.22 times the rate of those who are in the same function but not the same subfunction (equal to $\exp(1.169)$); combining these effects indicates that those in the same function and subfunction communicate at 6.84 times the baseline rate (equal to $\exp(0.755 + 1.169)$).

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 $\exp(0.836 \times \mathit{SameOffice} - 0.190 \times \ln(\text{distance}+1\text{ mile}))$ times the rate as otherwise identical cross-office pairs. Compared to two employees separated by just 100 miles, two people in the same office communicate 5.53 times more frequently. Relative to a dyad separated by the mean geographic distance in the sample (1,026 miles), two people in the same office communicate at 8.61 times that rate.

Shared implicit foci also exert an effect on interaction, but their magnitude is significantly weaker. The median number of coreceived emails during the month of September, conditional on a nonzero value, is three; two people who coreceive just three emails in September exchange 14% ($1.14 = \exp(0.043 \times 3)$) more messages during the subsequent quarter than an otherwise similar dyad with no shared implicit foci. Dyad members in the 90th percentile of coreceived messages (again, conditional on a nonzero value) are predicted to exchange 2.2 times more emails than otherwise similar dyad members who do not share an implicit focus. We note that when Kossinets and Watts (2009) included a similar covariate for shared implicit foci in their regressions, the demographic similarity variables in their models lost significance. Their interpretation of this pattern of results is that because university extracurricular activities are fully discretionary, endogenous selection into social foci explains all the variance in homophilous communication. Once they controlled for social foci, the effects of homophily disappear. By contrast, we find a positive effect of gender homophily, even when controlling for shared implicit foci: overall, same-gender dyad members exchange emails at a rate 24% higher ($1.24 = \exp(0.220)$) than that of mixed-gender pairs. When we split this effect by gender (see Model 1 in

Table 2 Correlation Matrix for Dyad-Level Variables in the Pseudopopulation ($N = 113{,}108{,}944$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) *CommunicationFrequency*	1												
(2) *SameBU*	0.022	1											
(3) *SameFunction*	0.013	0.264	1										
(4) *SameOffice*	0.025	0.042	0.025	1									
(5) *SharedImplicitFoci* (time 0)	0.191	0.048	0.034	0.050	1								
(6) *SameGender*	0.001	−0.004	0.032	−0.004	0.002	1							
(7) *BothMale*	−0.001	−0.015	0.052	−0.008	−0.001	0.832	1						
(8) *BothFemale*	0.003	0.019	−0.036	0.006	0.004	0.271	−0.309	1					
(9) *SameSalaryBand*	0.003	0.001	−0.004	0.005	0.013	0.013	0.017	−0.007	1				
(10) *Distance* (logged)	−0.018	−0.050	−0.007	−0.563	−0.034	−0.001	−0.002	0.002	−0.005	1			
(11) *EmailVolume* (logged)	0.013	−0.002	−0.127	0.006	0.019	−0.062	−0.106	0.077	0.006	−0.034	1		
(12) *BU%Women*	−0.001	0.049	−0.156	0.009	−0.012	−0.059	−0.097	0.066	−0.008	−0.047	0.085	1	
(13) *JobFunction%Women*	0.011	0.159	−0.071	0.019	0.009	−0.052	−0.115	0.111	−0.003	−0.035	0.140	0.233	1
(14) *Office%Women*	0.001	0.045	−0.035	0.018	0.001	−0.058	−0.098	0.070	−0.012	−0.081	0.122	0.106	0.198

Table 3 Poisson QML Models of Frequency of Dyadic Communication

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
SameBU	1.822 (0.032)**	1.703 (0.055)**	1.821 (0.032)**	1.825 (0.031)**	1.819 (0.032)**	1.764 (0.040)**
SameFunction	0.755 (0.046)**	0.757 (0.046)**	0.675 (0.071)**	0.759 (0.042)**	0.766 (0.043)**	0.758 (0.047)**
SharedImplicitFoci (time 0)	0.043 (0.002)**	0.043 (0.002)**	0.043 (0.002)**	0.040 (0.004)**	0.044 (0.002)**	0.042 (0.004)**
SameOffice	0.836 (0.095)**	0.838 (0.095)**	0.839 (0.095)**	0.860 (0.091)**	0.505 (0.176)**	0.568 (0.133)**
SameGender	0.220 (0.046)**	0.076 (0.027)**	0.122 (0.029)**	0.187 (0.031)**	0.103 (0.029)**	0.021 (0.035)
SameBU × *SameGender*		0.197 (0.067)**				0.095 (0.043)*
SameFunction × *SameGender*			0.137 (0.068)*			0.016 (0.045)
SharedImplicitFoci × *SameGender*				0.004 (0.005)		0.002 (0.004)
SameOffice × *SameGender*					0.538 (0.183)**	0.459 (0.130)**
SameSubfunction	1.169 (0.066)**	1.167 (0.065)**	1.168 (0.065)**	1.161 (0.052)**	1.154 (0.057)**	1.152 (0.051)**
Distance (logged)	−0.190 (0.011)**	−0.190 (0.011)**	−0.190 (0.011)**	−0.190 (0.011)**	−0.190 (0.011)**	−0.190 (0.011)**
SameSalaryBand	0.358 (0.041)**	0.357 (0.040)**	0.357 (0.040)**	0.345 (0.036)**	0.356 (0.039)**	0.350 (0.037)**
EmailVolume (logged)	1.505 (0.034)**	1.506 (0.034)**	1.506 (0.034)**	1.516 (0.036)**	1.502 (0.035)**	1.508 (0.035)**
BU%Women	−0.007 (0.002)**	−0.008 (0.002)**	−0.008 (0.002)**	−0.008 (0.002)**	−0.008 (0.002)**	−0.008 (0.002)**
JobFunction%Women	0.044 (0.006)**	0.045 (0.006)**	0.045 (0.006)**	0.044 (0.006)**	0.048 (0.006)**	0.047 (0.006)**
Office%Women	−0.014 (0.002)**	−0.014 (0.002)**	−0.014 (0.002)**	−0.014 (0.003)**	−0.014 (0.002)**	−0.014 (0.002)**
Constant	−14.475 (0.286)**	−14.395 (0.281)**	−14.426 (0.282)**	−14.498 (0.283)**	−14.421 (0.272)**	−14.396 (0.278)**
No. of observations	408,622	408,622	408,622	408,622	408,622	408,622

Note. Multiway, cluster-robust standard errors are in parentheses.
*Significant at 5%; **significant at 1%.

Table 5), we find that male–male dyads exhibit no significant homophily overall ($p > 0.7$), whereas female–female dyads communicate 62% more than an otherwise similar mixed-gender pair ($1.62 = \exp(0.482)$). We infer from this that selection into shared implicit foci is more task-directed and less self-selected in firms than in universities.

Hypothesis Tests

To test our hypotheses that actors are more likely to choose same-gender communication partners within opportunity structures, we introduce to the baseline models interactions between comembership in structural units and comembership in social categories. Results of Model 2 in Table 3 indicate that relative to same-gender dyad members who are *not* in the same business unit, same-gender dyad members who *are* in the same business unit communicate 6.7 times as much ($\exp(\beta_{SameBU} + \beta_{SameBU \times SameGender}) = \exp(1.703 + 0.197)$). This ratio is significantly higher than that implied by the main effect of *SameBU* alone ($6.2 = \exp(1.703)$). Said differently, the *SameGender* effect is amplified by 22% ($1.22 = \exp(\beta_{SameBU \times SameGender}) = \exp(0.197)$) in *SameBU* dyads compared with dyads where members are not employed in the same business unit, as indicated by the positive and significant ($p < 0.01$) coefficient of *SameBU* × *SameGender*. This finding provides support for Hypothesis 1.

Likewise, Model 3 indicates that for dyads with members employed in the same job function, the *SameGender* effect is 15% larger ($1.15 = \exp(\beta_{SameFunction \times SameGender}) = \exp(0.137)$) than for dyads with members *not* in the same job function ($p < 0.05$). This finding provides support for Hypothesis 2. In

Model 4, we test the hypothesis that quasi-formal organizational structure plays a similar role in creating opportunity for and constraint on interaction. The coefficient on the *SharedImplicitFoci* × *SameGender* interaction is positive, as predicted, but statistically insignificant. We find no overall evidence to support Hypothesis 3, but we will return to it in an extension to our hypothesis tests. In Model 5, we find a positive, significant coefficient on the *SameGender* × *SameOffice* interaction. Relative to dyad members who are *not* in the same office, dyad members who *are* in the same office experience a *SameGender* effect that is 71% larger ($1.71 = \exp(\beta_{SameOffice \times SameGender}) = \exp(0.538)$ with $p < 0.01$). This finding provides support for Hypothesis 4.

Hypothesis 5 argues that the effect of homophily on rates of interaction would be greater for dyads in which both members are in the same office, compared with its effect on dyads with both members in the same business units or functions. The logic for this prediction is that within-office communications are those most likely to have a heavy social component, because these are the most discretionary interactions. Inspection of Models 2–5 indicates that the coefficient magnitude of the *SameOffice* × *SameGender* interaction term in Model 5 is far larger than the corresponding interactions in Models 2–4. We also ran a full panel of interactions in Model 6, including each of the four interactions entered severally in Models 2–5. Results indicate that when estimated jointly, only the interactions of *SameBU* and *SameOffice* with *SameGender* remain significant. Whereas the *SameBU* × *SameGender* interaction is diminished by half, relative to Model 2, the *SameOffice* × *SameGender* interaction is diminished by just 15% relative to Model 5. And the *SameOffice* × *SameGender* interaction is by far the largest in magnitude in Model 6. We interpret these results as evidence in support of Hypothesis 5.

We test Hypothesis 6, that larger groups afford more discretion in selecting communication partners, and therefore more homophily, than smaller groups, in Table 4. We compare small business units with fewer than 275 people against large business units with over 3,000 people, small functions with fewer than 210 people against larger functions with over 400 people, and small offices with as few as 21 people against larger offices with more than 200 people. Table 3 established that the base rate of homophilous interaction is higher within groups than across group boundaries by focusing on the interaction between *SameGender* and, for example, *SameBU*. In Table 4, we compare the magnitude of the *SameGender* effect in large business units against its effect in small business units through a subsample analysis of dyads located in the same business unit. Because all dyads in the analysis share group affiliations by construction, we test Hypothesis 6 by comparing the main effect of *SameGender* in large groups against that in small groups.

Models 1 and 2 in Table 4 show separate estimates of the *SameGender* effect for dyads with both members in large business units or with both members in small business units, respectively. Consistent with Hypothesis 6, the same gender effect is positive and significant in the large business unit subsample (Model 1), but not in the small business unit subsample (Model 2). To test the statistical significance of this difference, we estimate these effects jointly in Model 3, which includes both same-large-business-unit dyads and same-small-business-unit dyads in the same regression. We differentiate between these two groups by including the covariate *LargeBU*, a binary indicator that equals 1 for the former group and 0 for the latter group.[10]

We test Hypothesis 6 by examining interactions of *SameGender* with covariates indicating that dyad members are coemployed in large (versus small) groups in Table 4. The positive, significant ($p < 0.05$) interaction coefficient of *LargeBU* × *SameGender* indicates that relative to dyads in small business units, dyads in larger business units experience a larger interaction effect between *SameBU* and *SameGender*. This result is not only statistically significant, it is practically significant as well: the *SameGender* premium is more than twice as large ($\exp(0.738) = 2.09$) in large business units compared with small business units. Analogous results for job function appear in Models 4–6 and for office location in Models 7–9. Across all three types of groups, the same-gender interaction is stronger in large groups than in small groups, providing support for Hypothesis 6.[11]

Empirical Extension: Male–Female Differences. To tease apart any differences in gender homophily among males from gender homophily among females, we turn to Table 5, in which we replace the *SameGender* variable with two variables: *BothMale* and *BothFemale*, each of which is estimated relative to the baseline of a mixed-gender dyad.[12] Additionally, we drop the volume control from Table 5 so that our findings can reflect the observed gender differences in total communication volume. Model 1 replicates the baseline regression from Table 3, but with the homophily covariate split by gender. It shows that the baseline rate of homophilous interaction is driven by women (62% more than male–female dyads: $\exp[0.482] = 1.62$). By contrast, male–male communications are no more likely than male–female interactions. Other results of Model 1 are substantively similar to those reported in Table 3.

Across the rest of Table 5, however, we find that the gender-comingled results of Table 3 mask significant gender-based differences in communication patterns. Starting with men, we find in Models 3 and 4

Table 4 Poisson QML Models of Frequency of Dyadic Communication Within Groups, Comparing the Magnitude of the Effect in Small vs. Large Groups

	Business units			Job functions			Offices		
	Large (Model 1)	Small (Model 2)	Both (Model 3)	Large (Model 4)	Small (Model 5)	Both (Model 6)	Large (Model 7)	Small (Model 8)	Both (Model 9)
SameGender	0.234	−0.269	−0.503	0.241	0.490	−0.662	0.399	−0.149	−0.168
	(0.052)**	(0.253)	(0.342)	(0.055)**	(0.389)	(0.450)	(0.118)**	(0.243)	(0.293)
SameBU				1.990	1.312	1.987	1.687	3.357	1.727
				(0.038)**	(0.626)*	(0.038)**	(0.111)**	(0.562)**	(0.107)**
SameFunction	1.383	0.336	1.284				0.822	−0.381	0.753
	(0.077)**	(0.333)	(0.075)**				(0.142)**	(0.439)	(0.139)**
SameOffice	1.847	1.110	1.802	2.053	0.883	2.049			
	(0.089)**	(0.317)**	(0.089)**	(0.083)**	(0.442)*	(0.083)**			
LargeBU			−3.520						
			(0.212)**						
LargeBU × *SameGender*			0.738						
			(0.344)*						
LargeFunction						−2.237			
						(0.291)**			
LargeFunction × *SameGender*						0.902			
						(0.450)*			
LargeOffice									−2.826
									(0.271)**
LargeOffice × *SameGender*									0.567
									(0.313)+
SharedImplicitFoci (time 0)	0.047	0.049	0.047	0.042	0.023	0.042	0.037	0.032	0.036
	(0.002)**	(0.007)**	(0.002)**	(0.002)**	(0.014)+	(0.002)**	(0.004)**	(0.008)**	(0.004)**
SameSubfunction	1.177	1.273	1.194	1.125	2.583	1.125	1.385	0.794	1.383
	(0.057)**	(0.332)**	(0.058)**	(0.055)**	(1.001)**	(0.055)**	(0.152)**	(0.383)*	(0.147)**
SameSalaryBand	0.230	−0.090	0.224	0.345	0.533	0.344	0.272	−0.026	0.264
	(0.056)**	(0.296)	(0.057)**	(0.050)**	(1.905)	(0.050)**	(0.122)*	(0.207)	(0.115)*
EmailVolume (logged)	1.678	1.408	1.659	1.476	1.361	1.472	1.345	0.827	1.311
	(0.051)**	(0.213)**	(0.050)**	(0.038)**	(0.554)*	(0.037)**	(0.088)**	(0.159)**	(0.084)**
BU%Women	0.023	−0.025	0.021	0.017	−0.052	0.017	0.014	−0.084	0.012
	(0.003)**	(0.014)+	(0.003)**	(0.003)**	(0.037)	(0.003)**	(0.007)*	(0.039)*	(0.007)+
JobFunction%Women	0.031	0.053	0.034	0.028	0.129	0.030	−0.011	0.071	−0.007
	(0.008)**	(0.035)	(0.007)**	(0.009)**	(0.140)	(0.009)**	(0.016)	(0.048)	(0.016)
Office%Women	−0.019	0.030	−0.017	−0.011	−0.002	−0.012	−0.029	−0.018	−0.028
	(0.004)**	(0.016)+	(0.004)**	(0.004)**	(0.024)	(0.004)**	(0.008)**	(0.012)	(0.007)**
Constant	−16.611	−10.472	−12.873	−15.583	−11.239	−13.300	−12.134	−3.871	−9.046
	(0.401)**	(1.468)**	(0.426)**	(0.300)**	(4.319)**	(0.423)**	(0.672)**	(1.810)*	(0.692)**

Note. Multiway, cluster-robust standard errors are in parentheses.
+Significant at 10%; *significant at 5%; **significant at 1%.

that *BothMale* is insignificant in its main effect on communication and in its interactions with *SameFunction* and *SharedImplicitFoci*. Models 2 and 5 show significant interactions of *BothMale* with *SameBU* and *SameOffice*, respectively. However, these effects partially offset negative and significant main effects of *BothMale* in those models. Within business units, homophily among men is insignificant ($\exp[-0.116+0.180]=1.07$; $p>0.38$); across business units, men actually communicate 11% less with other men than they do with women ($\exp[-0.116]=0.89$). It appears that the only place where gender homophily among men is positive and statistically significant is within office locations, where men interact with other men 60% more ($\exp[0.594-0.121]=1.60$) than they do with women (and even within offices, gender homophily is weaker among men than it is among women; $p<0.055$). This fact buttresses support for Hypothesis 5, that comembership in offices will play the most significant role in fostering homophily. In contrast, gender homophily among women appears to be pervasive at BigCo. All six models in Table 5 include positive, statistically significant main effects of *BothFemale*, suggesting that communication among women is amplified by as much as 60% relative to male–female communication rates, both within and across the boundaries we study (the magnitude of this

Table 5 Poisson QML Models of Frequency of Dyadic Communication, Examining Interactions Separately for Male and Female Same-Gender Dyads

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
SameBU	1.843	1.754	1.850	1.835	1.839	1.771
	(0.030)**	(0.048)**	(0.031)**	(0.030)**	(0.030)**	(0.038)**
SameFunction	0.559	0.559	0.493	0.540	0.567	0.526
	(0.039)**	(0.039)**	(0.057)**	(0.039)**	(0.038)**	(0.046)**
SharedImplicitFoci (time 0)	0.036	0.036	0.036	0.041	0.036	0.044
	(0.002)**	(0.002)**	(0.002)**	(0.006)**	(0.002)**	(0.006)**
SameOffice	0.883	0.882	0.888	0.825	0.585	0.452
	(0.082)**	(0.081)**	(0.082)**	(0.076)**	(0.127)**	(0.103)**
BothMale	0.014	−0.116	0.009	0.057	−0.121	−0.169
	(0.042)	(0.031)**	(0.035)	(0.030)+	(0.029)**	(0.039)**
BothFemale	0.482	0.442	0.237	0.432	0.436	0.299
	(0.045)**	(0.041)**	(0.046)**	(0.040)**	(0.037)**	(0.051)**
SameBU × *BothMale*		0.180				0.178
		(0.061)**				(0.046)**
SameBU × *BothFemale*		0.055				−0.118
		(0.069)				(0.060)*
SameFunction × *BothMale*			0.011			−0.064
			(0.063)			(0.049)
SameFunction × *BothFemale*			0.365			0.300
			(0.073)**			(0.061)**
SharedImplicitFoci × *BothMale*				−0.007		−0.009
				(0.006)		(0.006)
SharedImplicitFoci × *BothFemale*				0.022		0.020
				(0.004)**		(0.004)**
SameOffice × *BothMale*					0.594	0.734
					(0.126)**	(0.102)**
SameOffice × *BothFemale*					0.174	0.062
					(0.160)	(0.135)
SameSubfunction	1.242	1.241	1.238	1.258	1.233	1.245
	(0.041)**	(0.041)**	(0.041)**	(0.045)**	(0.038)**	(0.046)**
Distance (logged)	−0.225	−0.225	−0.224	−0.220	−0.224	−0.220
	(0.010)**	(0.010)**	(0.010)**	(0.010)**	(0.010)**	(0.010)**
SameSalaryBand	0.324	0.324	0.323	0.338	0.325	0.341
	(0.036)**	(0.035)**	(0.036)**	(0.029)**	(0.034)**	(0.029)**
Office%Women	−0.004	−0.004	−0.004	−0.003	−0.004	−0.006
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)+	(0.003)*
BU%Women	0.086	0.086	0.086	0.087	0.088	−0.003
	(0.006)**	(0.006)**	(0.006)**	(0.006)**	(0.006)**	(0.002)
JobFunction%Women	−0.007	−0.007	−0.006	−0.006	−0.006	0.089
	(0.003)*	(0.003)*	(0.003)*	(0.003)*	(0.003)*	(0.006)**
Constant	−5.130	−5.068	−5.092	−5.210	−5.097	−5.137
	(0.166)**	(0.160)**	(0.160)**	(0.150)**	(0.162)**	(0.155)**
Observations	408,622	408,622	408,622	408,622	408,622	408,622

Note. Multiway, cluster-robust standard errors are in parentheses.
+Significant at 10%; *significant at 5%; **significant at 1%.

effect is greater in magnitude than the *BothMale* effect with $p < 0.01$ in all models). These effects are further heightened within job functions ($\beta_{SameFunction \times BothFemale} = 0.365$ with $p < 0.01$ in Model 3) and quasi-formal structures ($\beta_{SharedImplicitFoci \times BothFemale} = 0.022$ with $p < 0.01$ in Model 4). These findings are robust to specifying interactions severally (Models 2–5) or as a full model (Model 6). The latter finding provides partial support for Hypothesis 3. We will return to the implications of these findings in the Conclusion.

To summarize our empirical results, comembership in business units, job functions, quasi-formal organizational structures, and office buildings all amplify the effect of gender homophily on dyadic communication. These effects are especially pronounced in large groups, which afford more discretionary choice compared with small groups, and in offices, where communication is laden with social content and more likely to be discretionary. We interpret these results as evidence in support of our theory that business units, job functions, quasi-formal structures, and, most of all, office locations each create both opportunity for and constraint on interactions. Within the constraints of the opportunity structure defined by these boundaries, individuals engage in discretionary interactions that are guided by choice homophily.

Conclusion

We have long known that homophily is a twin-engine phenomenon. The motors are individual preference to interact with similar others and differential opportunities to associate based on how people sort, self-select, or are selected into physical and social locations. Within organizations, we argue that much of this sorting transpires through the formal organizational structure, which delimits individuals' intraorganizational social worlds by determining with whom they are professionally interdependent or physically colocated. Within organizationally given sets of potential communication partners, we argue that individuals have more discretion to select specific alters, and therefore they will engage in more homophilous exchanges in the thick interactions that occur within organizational units, relative to the more episodic communications that span the boundaries of formal and quasi-formal organizational structures. This is especially true in large groups, which afford more choice in interaction partners than do small ones. Our empirical analysis of email communications among employees of a large information technology firm generally confirms these hypotheses.

We also explored male–female differences in interaction patterns in BigCo's email network. Here, we found that women have a greater number of contacts than men. We cannot know the broader context of this finding because we do not possess data on all means of communication. The result could indicate systematic gender differences in the use of email relative to other communication media, although we found no suggestion of this in our exploratory interviews at BigCo, nor have prior studies observed gender differences in the use of email in business organizations (e.g., Gefen and Straub 1997). Therefore, we suspect this result indicates substantive differences in network structure: women in BigCo communicate more than men, and they do so by adding other women to their networks. We also found sex differences in how gender interacts with organizational and geographic boundaries to influence the level of homophily in communications. The higher baseline rate of homophily we observe in women is insensitive to business unit or office boundaries; that is, women communicate at an elevated rate with other women both inside their business units and offices and across those boundaries. Job functions and quasi-formal structures further amplify homophily among women. Men, in contrast, only seem to experience any significant gender homophily within office locations. We find the gender difference in both communication rates and interaction patterns to be striking and a promising avenue for future research.

There are several possible, admittedly speculative, explanations for these results. The first is an inherent gender difference, in which women in modern-day organizations have networks that are more collaborative and less focused on parochial interests or within-unit loyalties than those that men have (Borgatti and Cross 2003). Recent research suggests that culturally influenced implicit attitudes toward collaboration may indeed affect the structure of social networks (Srivastava and Banaji 2011), and there may be a systematic gender difference in implicit attitudes. In fact, there is work to suggest that women tend to be more collaborative than men (Eagly and Carli 2003). The possibility therefore exists that women have a greater propensity to collaborate, which leads them to transcend the business unit and geographic boundaries that appear to constrain men. When traversing these boundaries, women appear more likely to communicate with other women.

A second possible explanation for this result is that women communicate with one another more often because they are effectively excluded from the male power structure of the firm. The classic literature on gender and managerial networks argued that women are largely excluded from the "shadow structure...of power" in organizations (Kanter 1977, p. 164). Although the situation women encounter in the labor market has clearly improved in the years since Kanter's analysis of "Indsco" (Kaufman 2010), such arguments are not yet relegated to the annals of business history. In a more recent incarnation, Groysberg (2010) argued that women face "institutional barriers" to creating strong working relations. If women respond to all-male cliques by creating ties with other women (Casciaro and Piskorski 2005)—even those outside their business units and offices—that too could explain our results. A third possibility is that women are connecting with one another as a result of unobservable (to us) programs specifically designed to enhance women's networks in the company. BigCo does have some such programs, although the communication differences between the sexes are too substantial to be fully explained by such initiatives. Furthermore, to the extent that they are effective, we would expect that they would be at least partially accounted for by the measure of shared implicit foci. Thus, the differences in gender homophily within versus across offices is net of the effect of such programs.

To begin to adjudicate between these explanations, we note that women are both more central and more broadly connected in the internal BigCo email network, as shown in the descriptive statistics of Table 1(a). These findings are consistent with the possibility that lateral communication among women serves to reinforce, not undermine, their positions in the organization. Thus, our results suggest—though we do not claim that they prove—that homophilous interaction can actually help to span formal organizational boundaries that are otherwise difficult to traverse (Simmel 1902, Blau and Schwartz 1984, Alderfer 1987). Consequently, a more complete picture of both formal and informal structure reveals a situation in which homophily with respect to one category (in our study, gender) can actually serve to broaden a person's network with respect to another category (office and business unit).[13]

We conclude with several notes of caution. First, and most significantly, our data describe just a single company. Although we analyze a vast data set, we have no basis for any claim of generalizability beyond the single organization we study. This limitation is particularly important because we cannot know the degree to which our findings depend on the particular composition and organizational structure of the firm we study. We do not believe that our results are an artifact of an idiosyncratic structure—indeed, a multidivisional structure is typical of large, complex firms—but we cannot claim that our results apply to any firm but BigCo. Furthermore, as Ely (1994) has demonstrated, gender dynamics are played out on a stage in which the overall gender composition of an organization matters, in contrast to purely person-centered views in which gender roles are assumed to be intrinsic. Like its formal structure, BigCo's gender composition is fairly typical of U.S.-headquartered companies in its sector. Nevertheless, we really cannot know whether or how the findings on the differing network structures of men and women may be influenced by the company's overall gender composition.

This latter point also raises a second limitation of our study: of necessity, we have treated the sorting of individuals into organizational locations to be exogenous to the incidence of communication in dyads. Although we believe this assumption to be a reasonable first approximation, to the extent that it is violated, the implications for the regression results are unknowable. For instance, individuals may select (or be selected) into business units, job functions, and offices to be near to their close contacts, and these alters may be disproportionately of the same gender. If an employee chooses a position in a (relatively) gender-skewed organizational unit and subsequently finds many organizationally proximate colleagues of the same gender, choice homophily in terms of what position to take will drive relative levels of opportunity for within-unit homophilous interaction. Therefore, future analyses must consider the endogenous selection of individuals into organizational roles (Bielby and Baron 1986, Kossinets and Watts 2009, Kleinbaum and Stuart 2013).

Third, we found strong evidence of gender differences in interaction patterns, but we cannot know whether these results generalize to homophily in other social dimensions or ascriptive characteristics. Unfortunately, gender is the only sociodemographic variable available to us. We do believe that our theory of discretionary choice within the constraints of an opportunity structure posed by formal organization and geography would extend to preferences for homophilous interaction along other dimensions such as race, ethnicity, and age, but we have no empirical evidence with which to address this question.

Last, we must note that there are some drawbacks with the use of electronic mail data, in addition to their many benefits. Most importantly, the email network does not measure social relations per se, but communication. Furthermore, we have no data on interactions that take place outside of email, such as phone calls or the proverbial watercooler conversations. Although we attribute the findings—and the gender differences findings, in particular—to differences in underlying communication patterns, we cannot rule out the possibility that we may be observing differences in the use of communication media. Consistent with prior research (e.g., Gefen and Straub 1997), our interviews with a dozen BigCo employees about their use of email raised no hint of this possibility. However, future research should explore the possibility of gender-based differences in how different communication media are used. For example, if the text of emails—or even just their subject lines—were available, it would create opportunities to better understand whether men and women differ in their use of email and to elucidate the nature of the additional ties possessed by women at BigCo, relative to men. Do women really have broader work-related networks? Do they reach out to seek more social interactions within the organization (e.g., Ibarra 1992)? Because content data were not available to us, these questions remain for future research to answer.

Conversely, the data have real strengths. First, the sample is large; it includes millions of interactions among tens of thousands of employees, enabling analyses of larger, more complex organizations than those studied previously and permitting unprecedented views into how formal and informal structures interact in large, complex business organizations. Second, the data are unaffected by problems of nonresponse, recall, or bias in survey response (Killworth and Bernard 1976, Quintane and Kleinbaum 2011), problems that network analysts have long acknowledged but ignored (Marsden 2005). Third, for many of the questions that will interest students of organizations, we believe that the availability of what may well be the majority of interactions in an organization is as much a benefit as it is a source of concern.

Acknowledgments

The authors are grateful to Pino Audia, Pierre Azoulay, Matt Bothner, Connie Helfat, Andy King, James Kitts, Gail McGuire, Ray Reagans, Marc-David Seidel, Judith White, and Valery Yakubovich; seminar participants at the Harvard Business School and the Dartmouth Sociology Department; conference attendees at the Academy of Management and American Sociological Association meetings for helpful comments and suggestions; and, of course, the many employees of BigCo who spent time with them. They especially acknowledge Paul Wolfson for helpful suggestions and valuable programming support. A. Kleinbaum gratefully acknowledges financial support from the Ewing Marion Kauffman Foundation. Please direct correspondence to the first author.

Endnotes

[1]In the organization we study, formal structure and geographic space are only loosely coupled. People in the same business unit and in the same function are often assigned to the same geographic office, but very often this is not the case. For example, the correlation between *SameBU* and *SameOffice* is less than 5% (see Table 2).

[2]Theoretically, the same logic applies to quasi-formal structures; if we could observe them directly, we would hypothesize the same effect. However, because we must infer the existence of quasi-formal structures from the pattern of coreceipt of mass emails, as described in the Independent Variables section, we cannot distinguish small quasi-formal structures from large ones.

[3]Of the original 114 million dyadic emails, 31 million involved a person outside the United States or otherwise not included in the sample and about whom we have no demographic data; 3.5 million involved administrative assistants; 1.2 million were Bccs (in these instances, we retain the message for "To" and "Cc" recipients but do not include the sender-to-Bcc interaction). In addition, 64 million were mass mailings (i.e., they included more than four recipients). Mass mailings represent just 17% of total emails sent, but they are just over 50% of pairwise exchanges based on the expansion of the message to include all sender–recipient ties.

[4]As in many U.S. corporations, the administrative staff is overwhelmingly female. Throughout this paper, however, we focus exclusively on nonadministrative employees.

[5]To protect the privacy of BigCo and its employees, we cannot more precisely disclose the gender distribution in specific parts of the company.

[6]Although BigCo is global in scope, privacy laws in Europe and Asia limited our data collection to the 289 U.S. offices.

[7]In unreported results, we used two alternative specifications of the dependent variable. In one, the dependent variable was a binary indicator of whether or not the dyad members communicated during the observation window, and logit models were estimated. In the second, zero-inflated Poisson models were used to separately estimate the probability of dyadic communication and its frequency, conditional on a nonzero value. Across these approaches, the results were substantively similar.

[8]Unlike the maximum likelihood Poisson, quasi-maximum likelihood estimation of the Poisson 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 have a negative binomial distribution, Poisson QML estimation is preferable because it is consistent even if the data are, in fact, distributed negative binomial. The only assumption of Poisson QML estimation concerns the distribution of the conditional mean of the data (Gouriéroux et al. 1984, Wooldridge 1997, Silva and Tenreyro 2006).

[9]clus_nway.ado will be made available for public use at http://bit.ly/clus_nway.

[10]Note that the main effect of *LargeBU* is large, negative, and statistically significant ($p < 0.01$). This indicates that, all else equal, dyads coemployed in a large business unit interact less, on average, than dyads coemployed in a small business unit. This finding is consistent with the general observation that network density typically diminishes with group size (Wasserman and Faust 1994).

[11]The designations of small and large groups are inherently arbitrary, so we tested them extensively and found them to be robust to numerous alternative specifications. Additionally, in the models in Table 3, a control variable was included for the logged distance between dyad members' work locations. In the subsample analyses in Table 4, several of the small job functions were scattered across different buildings within a corporate campus (which has the peculiar property that $Distance = 0$ and $SameOffice = 0$). As a result, when the distance control was included, the coefficient of *SameOffice* in Model 5 was negative, indicating extensive communication across buildings within the corporate campus. We therefore dropped the distance control from all models in Table 4 to yield more sensible results. Results on our test statistics for Hypothesis 6 were unaffected by this change.

[12]We use this contrast coding system (Kaufman and Sweet 1974) to highlight differences between both-male and male–female dyads and between both-female and male–female dyads, respectively. In unreported results, we replicate our analyses, respecifying our dummy variables as *SameGender* and *BothFemale* to get explicit significance tests for differences between both-male and both-female dyads. The results indicate that the $SameBU \times BothFemale$ coefficient is significantly larger than the $SameBU \times BothMale$ coefficient with $p < 0.10$; all other gender differences were significant with $p < 0.01$.

[13]Although our data are limited to gender homophily, our finding echoes similar results by Thomas (1990) and by Ibarra (1995), who showed that racial homophily led minority managers to have more contacts outside their departments or groups than white managers in a purposeful effort to establish relationships with others of their own race.

References

Adams RB, Funk P (2012) Beyond the glass ceiling: Does gender matter? *Management Sci.* 58(2):219–235.

Adler PS, Goldoftas B, Levine DI (1999) Flexibility versus efficiency? A case study of model changeovers in the Toyota production system. *Organ. Sci.* 10(1):43–68.

Alderfer CP (1987) An intergroup perspective on group dynamics. Lorsch JW, ed. *Handbook of Organizational Behavior* (Prentice-Hall, Englewood Cliffs, NJ), 190–222.

Allen TJ (1977) *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information Within the R&D Organization* (MIT Press, Cambridge, MA).

Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci. USA* 106(51):21544–21549.

Baron JN, Davis-Blake A, Bielby WT (1986) The structure of opportunity: How promotion ladders vary within and among organizations. *Admin. Sci. Quart.* 31(2):248–273.

Bernard HR, Killworth PD, Sailer L (1981) Summary of research on informant accuracy in network data and the reverse small world problem. *Connections* 4(2):11–25.

Bielby WT, Baron JN (1986) Men and women at work: Sex segregation and statistical discrimination. *Amer. J. Sociol.* 91(4):759–799.

Blau PM (1994) *Structural Contexts of Opportunities* (University of Chicago Press, Chicago).

Blau PM, Schwartz JE (1984) *Crosscutting Social Circles: Testing a Macrostructural Theory of Intergroup Relations* (Academic Press, Orlando, FL).

Borgatti SP, Cross R (2003) A relational view of information seeking and learning in social networks. *Management Sci.* 49(4):432–445.

Bossard JHS (1932) Residential propinquity as a factor in marriage selection. *Amer. J. Sociol.* 38(2):219–224.

Brass DJ (1985) Men's and women's networks: A study of interaction patterns and influence in an organization. *Acad. Management J.* 28(2):327–343.

Burt RS (1998) The gender of social capital. *Rationality Soc.* 10(1):5–46.

Cairncross F (2001) *The Death of Distance: How the Communications Revolution Is Changing Our Lives*, new ed. (Harvard Business School Press, Boston).

Cameron AC, Gelbach JB, Miller DL (2011) Robust inference with multi-way clustering. *J. Bus. Econom. Statist.* 29(2):238–249.

Casciaro T, Piskorski MJ (2005) Power imbalance, mutual dependence, and constraint absorption: A closer look at resource dependence theory. *Admin. Sci. Quart.* 50(2):167–199.

Chandler AD (1962) *Strategy and Structure: Chapters in the History of the Industrial Enterprise* (MIT Press, Cambridge, MA).

Cosslett SR (1981) Maximum likelihood estimator for choice-based samples. *Econometrica* 49(5):1289–1316.

Eagly AH, Carli LL (2003) The female leadership advantage: An evaluation of the evidence. *Leadership Quart.* 14(6):807–834.

Eckmann J-P, Moses E, Sergi D (2004) Entropy of dialogues creates coherent structures in e-mail traffic. *Proc. Natl. Acad. Sci. USA* 101(40):14333–14337.

Ely RJ (1994) The effects of organizational demographics and social identity on relationships among professional women. *Admin. Sci. Quart.* 39(2):203–238.

Feld SL (1981) The focused organization of social ties. *Amer. J. Sociol.* 86(5):1015–1035.

Fernandez-Mateo I (2009) Cumulative gender disadvantage in contract employment. *Amer. J. Sociol.* 114(4):871–923.

Festinger L, Schachter S, Back K (1950) *Social Pressures in Informal Groups: A Study of Human Factors in Housing* (Harper, New York).

Fischer CS (1982) *To Dwell Among Friends: Personal Networks in Town and City* (University of Chicago Press, Chicago).

French MT, Maclean JC, Ettner SL (2008) Drinkers and bettors: Investigating the complementarity of alcohol consumption and problem gambling. *Drug Alcohol Dependence* 96(1–2):155–164.

Galbraith JR (1973) *Designing Complex Organizations* (Addison-Wesley, Reading, MA).

Galbraith JR (1994) *Competing with Flexible Lateral Organizations*, 2nd ed. (Addison-Wesley, Reading, MA).

Gefen D, Straub DW (1997) Gender differences in the perception and use of E-Mail: An extension to the technology acceptance model. *MIS Quart.* 21(4):389–400.

Gibson DE, Lawrence BS (2010) Women's and men's career referents: How gender composition and comparison level shape career expectations. *Organ. Sci.* 21(6):1159–1175.

Goodreau SM, Kitts JA, Morris M (2009) Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography* 46(1):103–125.

Gouriéroux C, Monfort A, Trognon A (1984) Pseudo maximum likelihood methods: Theory. *Econometrica* 52(3):681–700.

Groysberg B (2010) *Chasing Stars: The Myth of Talent and the Portability of Performance* (Princeton University Press, Princeton, NJ).

Gulati R, Puranam P (2009) Renewal through reorganization: The value of inconsistencies between formal and informal organization. *Organ. Sci.* 20(2):422–440.

Hackman JR, Oldham GR (1980) *Work Redesign* (Addison-Wesley, Reading, MA).

Han S-K (1996) Structuring relations in on-the-job networks. *Soc. Networks* 18(1):47–67.

Hrebiniak LG, Joyce WF (1984) *Implementing Strategy* (Macmillan, New York).

Ibarra H (1992) Homophily and differential returns: Sex differences in network structure and access in an advertising firm. *Admin. Sci. Quart.* 37(3):422–447.

Ibarra H (1993) Personal networks of women and minorities in management: A conceptual framework. *Acad. Management Rev.* 18(1):56–87.

Ibarra H (1995) Race, opportunity, and diversity of social circles in managerial networks. *Acad. Management J.* 38(3):673–703.

Ibarra H (1997) Paving an alternative route: Gender differences in managerial networks. *Soc. Psych. Quart.* 60(1):91–102.

Ibarra H, Smith-Lovin L (1997) New directions in social network research on gender and organizational careers. Cooper CL, Jackson SE, eds. *Creating Tomorrow's Organizations: A Handbook for Future Research in Organizational Behavior* (John Wiley & Sons, New York), 359–383.

Imbens GW (1992) An efficient method of moments estimator for discrete choice models with choice-based sampling. *Econometrica* 60(5):1187–1214.

Ingram P, Morris MW (2007) Do people mix at mixers? Structure, homophily, and the "life of the party." *Admin. Sci. Quart.* 52(4):558–585.

Johnson R, Kovács B, Vicsek A (2012) A comparison of email networks and off-line social networks: A study of a medium-sized bank. *Soc. Networks* 34(4):462–469.

Kalleberg AL, Knoke D, Marsden PV, Spaeth JL (1996) *Organizations in America: Analyzing Their Structures and Human Resource Practices* (Sage, Thousand Oaks, CA).

Kanter RM (1977) *Men and Women of the Corporation* (Basic Books, New York).

Kaufman RL (2010) *Race, Gender, and the Labor Market: Inequalities at Work* (Lynne Rienner Publishers, Boulder, CO).

Kaufman D, Sweet R (1974) Contrast coding in least squares regression analysis. *Amer. Ed. Res. J.* 11(4):359–377.

Kenny DA, Kashy DA, Cook WL (2006) *Dyadic Data Analysis* (Guilford Press, New York).

Killworth P, Bernard H (1976) Informant accuracy in social network data. *Human Organ.* 35(3):269–286.

King G, Zeng L (2001) Logistic regression in rare events data. *Political Anal.* 9(2):137–163.

Kleinbaum AM (2012) Organizational misfits and the origins of brokerage in intra-firm networks. *Admin. Sci. Quart.* 57(4):1–45.

Kleinbaum AM, Stuart TE (2013) Inside the black box of the corporate staff: Social networks and the implementation of corporate strategy. *Strategic Management J*., ePub ahead of print April 26.

Kleinbaum AM, Stuart TE, Tushman ML (2009) Communication (and coordination?) in a modern, complex organization. Working paper, Tuck School of Business, Dartmouth College, Hanover, NH.

Kogut B, Zander U (1992) Knowledge of the firm, combinative capabilities, and the replication of technology. *Organ. Sci.* 3(3):383–397.

Kono C, Palmer D, Friedland R, Zafonte M (1998) Lost in space: The geography of corporate interlocking directorates. *Amer. J. Sociol.* 103(4):863–911.

Kossinets G, Watts DJ (2009) Origins of homophily in an evolving social network. *Amer. J. Sociol.* 115(2):405–450.

Lancaster T, Imbens G (1996) Case-control studies with contaminated controls. *J. Econometrics* 71(1–2):145–160.

Laumann EO (1966) *Prestige and Association in an Urban Community; an Analysis of an Urban Stratification System* (Bobbs-Merrill, Indianapolis).

Lazarsfeld PF, Merton RK (1954) Friendship as social process: A substantive and methodological analysis. Berger M, Abel T, eds. *Freedom and Control in Modern Society* (Van Nostrand, New York), 18–66.

Lewis K, Kaufman J, Gonzalez M, Wimmer A, Christakis N (2008) Tastes, ties, and time: A new social network data set using Facebook.com. *Soc. Networks* 30(4):330–342.

Lin N (2001) *Social Capital: A Theory of Social Structure and Action* (Cambridge University Press, London).

Lincoln JR, Miller J (1979) Work and friendship ties in organizations: A comparative analysis of relation networks. *Admin. Sci. Quart.* 24(2):181–199.

Lindgren K-O (2010) Dyadic regression in the presence of heteroscedasticity—An assessment of alternative approaches. *Soc. Networks* 32(4):279–289.

Liu CC (2010) *Essays on Network Antecedents in a Knowledge Production Context* (Harvard University, Boston).

Marmaros D, Sacerdote B (2006) How do friendships form? *Quart. J. Econom.* 121(1):79–119.

Marsden PV (1987) Core discussion networks of Americans. *Amer. Sociol. Rev.* 52(1):122–131.

Marsden PV (1988) Homogeneity in confiding relations. *Soc. Networks* 10(1):57–76.

Marsden PV (2005) Recent developments in network measurement. Carrington PJ, Scott J, Wasserman S, eds. *Models and Methods in Social Network Analysis* (Cambridge University Press, New York), 8–30.

McGregor D (1960) *The Human Side of the Enterprise* (McGraw-Hill, New York).

McGuire GM (2002) Gender, race, and the shadow structure. *Gender Soc.* 16(3):303–322.

McPherson JM, Smith-Lovin L (1987) Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *Amer. Sociol. Rev.* 52(3):370–379.

McPherson JM, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annual Rev. Sociol.* 27(1):415–444.

Mok D, Carrasco J-A, Wellman B (2010) Does distance still matter in the age of the internet? *Urban Stud.* 47(13):2747–2783.

Moody J, McFarland D, Bender-deMoll S (2005) Dynamic network visualization. *Amer. J. Sociol.* 110(4):1206–1241.

Quintane E, Kleinbaum AM (2011) Matter over mind? E-mail data and the measurement of social networks. *Connections* 31(1):22–46.

Reagans R (2005) Preferences, identity, and competition: Predicting tie strength from demographic data. *Management Sci.* 51(9):1374–1383.

Reagans R (2011) Close encounters: Analyzing how social similarity and propinquity contribute to strong network connections. *Organ. Sci.* 22(4):835–849.

Reagans R, Zuckerman E, McEvily B (2004) How to make the team: Social networks vs. demography as criteria for designing effective teams. *Admin. Sci. Quart.* 49(1):101–133.

Ridgeway CL (1997) Interaction and the conservation of gender inequality: Considering employment. *Amer. Sociol. Rev.* 62(2):218–235.

Ruef M, Aldrich HE, Carter NM (2003) The structure of founding teams: Homophily, strong ties, and isolation among U.S. entrepreneurs. *Amer. Sociol. Rev.* 68(2):195–222.

Russell LB, Teutsch SM, Kumar R, Dey A, Milan E (2001) Preventable smoking and exercise-related hospital admissions: A model based on the NHEFS. *Amer. J. Preventive Medicine* 20(1):26–34.

Shrum W, Cheek NH Jr, Hunter SM (1988) Friendship in school: Gender and racial homophily. *Sociol. Ed.* 61(4):227–239.

Silva JMCS, Tenreyro S (2006) The log of gravity. *Rev. Econom. Statist.* 88(4):641–658.

Simmel G (1902) *The Sociology of Georg Simmel* [Reprinted 1950. Translated by Wolff KH. Free Press, Glencoe, IL.]

Singh J, Hansen MT, Podolny JM (2010) The world is not small for everyone: Inequity in searching for knowledge in organizations. *Management Sci.* 56(9):1415–1438.

Soda G, Zaheer A (2012) A network perspective on organizational architecture: Performance effects of the interplay of formal and informal organization. *Stragegic Management J.* 33(6):751–771.

Sorenson O, Stuart TE (2001) Syndication networks and the spatial distribution of venture capital investments. *Amer. J. Sociol.* 106(6):1546–1588.

South SJ, Bonjean CM, Markham WT, Corder J (1982) Social structure and intergroup interaction: Men and women of the federal bureaucracy. *Amer. Sociol. Rev.* 47(5):587–599.

Srivastava SB, Banaji MR (2011) Culture, cognition, and collaborative networks in organizations. *Amer. Sociol. Rev.* 76(2):207–233.

Taylor FW (1911) *Scientific Management*, 2nd ed. (Harper & Brothers, New York).

Thomas DA (1990) The impact of race on managers' experiences of developmental relationships (mentoring and sponsorship): An intra-organizational study. *J. Organ. Behav.* 11(6):479–492.

Thompson JD (1967) *Organizations in Action: Social Science Bases of Administrative Theory*, 2nd ed. (Transaction Publishers, New Brunswick, NJ).

Turner AN, Lawrence PR (1965) *Industrial Jobs and the Worker: An Investigation of Response to Task Attributes* (Division of Research, Graduate School of Business Administration, Harvard University, Boston).

Tushman ML, Nadler DA (1978) Information-processing as an integrating concept in organizational design. *Acad. Management Rev.* 3(3):613–624.

Tzabbar D, Vestal A, Porter R, McMahon S (2010) The interdependencies of formal and informal structure and the exploration of new technological opportunities among geographically dispersed firms. Presentation, SMS 30th Annual International Conference, September 13, Strategic Management Society, Chicago.

Wasserman S, Faust K (1994) *Social Network Analysis: Methods and Applications* (Cambridge University Press, New York).

Williamson OE (1975) *Markets and Hierarchies, Analysis and Antitrust Implications: A Study in the Economics of Internal Organization* (Free Press, New York).

Wooldridge JM (1997) Quasi-likelihood methods for count data. Pesaran MH, Schmidt P, eds. *Handbook of Applied Econometrics*, Vol. 2 (Blackwell, Oxford, UK), 352–406.

Wuchty S, Uzzi B (2011) Human communication dynamics in digital footsteps: A study of the agreement between self-reported ties and email networks. *PLoS ONE* 6(11):e26972.

Yakubovich V, Shekshnia S (2010) Networks of task interdependence and promotion: A case of the Russian corporate bureaucracy. Wharton School working paper, University of Pennsylvania, Philadelphia.

Zipf GK (1949) *Human Behavior and the Principle of Least Effort; An Introduction to Human Ecology* (Addison-Wesley Press, Cambridge, MA).

Adam M. Kleinbaum is an assistant professor of business administration at the Tuck School of Business at Dartmouth College. He received his doctorate in business administration at Harvard Business School. His research examines the antecedents of network structure within large organizations and its consequences for firm-level outcomes, such as innovation and firm performance.

Toby E. Stuart is the Leo Helzel Professor of Entrepreneurship, Strategy, and Innovation at the Haas School of Business, University of California, Berkeley. He received his Ph.D. from the Graduate School of Business at Stanford. His research focuses on social networks, venture capital networks, and their role in the creation of new firms.

Michael L. Tushman is the Paul R. Lawrence Class of 1942 Professor at the Harvard Business School. He received his Ph.D. from the MIT Sloan School of Management. His research interests include innovation streams and organization architecture, senior teams and attention to paradox, social networks and interdependent innovation, and the impact of distributed innovation on incumbents.