Productivity Effects of Information Diffusion in Networks

Paper 234

July 2007

Sinan Aral
Erik Brynjolfsson
Marshall Van Alstyne

Productivity Effects of Information Diffusion in Networks

Sinan Aral
Stern School of Business, NYU & Sloan School of Management, MIT.
sinan@stern.nyu.edu

Erik Brynjolfsson
Sloan School of Management, MIT.
erikb@mit.edu

Marshall Van Alstyne
B.U. School of Management & Sloan School of Management, MIT.
mva@bu.edu

Abstract

We examine the drivers of diffusion of information through organizations and the effects on performance. In particular, we ask: What predicts the likelihood of an individual becoming aware of a strategic piece of information, or becoming aware of it sooner? Do different types of information exhibit different diffusion patterns, and do different characteristics of social structure, relationships and individuals in turn affect access to different kinds of information? Does better access to information predict an individual's ability to complete projects or generate revenue? We hypothesize that the dual effects of content and structure jointly predict the diffusion path of information, and ultimately performance. To test our hypotheses, we characterize the social network of a medium sized executive recruiting firm using accounting data on project co-work relationships and ten months of email traffic observed over two five month periods. We identify two distinct types of information diffusing over this network – 'event news' and 'discussion topics' – by their usage characteristics, and observe several thousand diffusion processes of each type of information from their original first use to their varied recipients over time. We then test the effects of network structure and functional and demographic characteristics of dyadic relationships on the likelihood of receiving each type of information and receiving it more quickly. Our results demonstrate that the diffusion of news, characterized by a spike in communication and rapid, pervasive diffusion through the organization, is influenced by demographic and network factors but not by functional relationships (e.g. prior co-work, authority) or the strength of ties. In contrast, diffusion of discussion topics, which exhibit more shallow diffusion characterized by 'back-and-forth' conversation, is heavily influenced by functional relationships and the strength of ties, as well as demographic and network factors. Discussion topics are more likely to diffuse vertically up and down the organizational hierarchy, across relationships with a prior working history, and across stronger ties, while news is more likely to diffuse laterally as well as vertically, and without regard to the strength or function of relationships. Furthermore, we find that access to information strongly predicts the number of projects completed by each individual and the amount of revenue that person generates. The effects are economically significant, with each additional "word seen" correlated with about $70 of additional revenue generated. Our findings highlight the importance of simultaneous considerations of structure and content in information diffusion studies and provide some of the first evidence on the economic importance of information diffusion in networks.

Keywords: Social Networks, Information Diffusion, Productivity, Information Workers.

We are grateful to the National Science Foundation (Career Award IIS-9876233 and grant IIS-0085725), Cisco Systems, France Telecom and the MIT Center for Digital Business for generous funding. We thank Tim Choe and Jun Zhang for their tireless research assistance.

1. Introduction

The process of information diffusion through social groups lies at the heart of numerous economic phenomena in industrial organization, strategy, productivity, finance, marketing, and innovation. Theories on subjects as wide ranging as the diffusion of innovations (e.g. Rogers 1995), dynamic trading behavior (e.g. Hirshleifer et. al. 1994), and the mechanics of word of mouth marketing (e.g. Dellarocas 2003), rely on information diffusion as a central theoretical building block, making important assumptions about how information spreads between individuals. Timely access to strategic information, innovative ideas, or current news can also highlight hidden opportunities, provide negotiating leverage (Burt 1992), promote innovation (Hargadon & Sutton 1997, Burt 2004), and ultimately drive economic performance (Reagans & Zuckerman 2001, Hansen 2002, Aral, Brynjolfsson & Van Alstyne 2006). But, while theories based on information diffusion proliferate, empirical evidence on how information spreads through social groups, and its ultimate economic effects, remains scarce.

Diffusion studies typically observe adoption or purchase decisions rather than the actual diffusion of information itself. The studies that do focus on information diffusion are typically theoretical, or derive results from computer simulations of information passing among a handful of actors. Existing theory focuses mainly on which global social structures maximize diffusion, and although we know that transfers of certain types of information are easier than others (Von Hippel 1998), diffusion studies typically treat information as a uniform concept, making variation in diffusion patterns across different information types and social structures difficult to theorize. The apparent value of timely access to strategic information, which rests on the likelihood of seeing the information and seeing it sooner than others, gives rise to a natural set of questions about the dynamic movement of information through populations: How does information diffuse through a given social group? What makes someone more likely to be exposed to an idea as it spreads? Do different types of information diffuse differently? Can we explicitly link access to novel information to changes in performance?

To complement studies of the economic value of information, and provide evidence on how and when information diffuses in organizations, we study the movement of different types of information through one organization over a period of two years. We argue that the social nature of information diffusion necessitates simultaneous examination of both the type of information and the type of social relationship or structure through which it diffuses. In organizations, the dual effect of content and structure jointly predict the diffusion path of a given piece of information. While one type of information may be more likely to diffuse upward through the organizational hierarchy or strictly across functional relationships, a different type of information may diffuse laterally or without regard to function or hierarchy.

To test our theory, we characterize the social network of a medium sized executive recruiting firm using ten months of email data observed over two five month periods and accounting data detailing project co-work relationships. We identify two distinct types of information diffusing over this network – 'event news' and 'discussion topics' – by their usage characteristics, and observe several thousand diffusion processes of each type of information from their original first use to their varied recipients over time. We then test the effects of network structure and functional and demographic characteristics of dyadic relationships and individuals on the likelihood of receiving each type of information and receiving it more quickly.

Our results demonstrate that the diffusion of news, characterized by a spike in communication and rapid, pervasive diffusion through the organization, is influenced by demographic and network factors but not by functional relationships (e.g. prior co-work, authority) or the strength of ties. In contrast, diffusion of discussion topics, which exhibit more shallow diffusion characterized by 'back-and-forth' conversation, is heavily influenced by functional relationships and the strength of ties, as well as demographic and network factors. Discussion topics are more likely to diffuse vertically up and down the organizational hierarchy, across relationships with a prior working history, and across stronger ties, while news is more likely to diffuse laterally as well as vertically, and without regard to the strength or function of relationships. These findings highlight the importance of simultaneous considerations of structure and content in information diffusion studies.

We also find that access to information strongly predicts employees' productivity. Timely access to more information predicts the number of projects completed by each individual and the amount of revenue each person generates. Each additional "word seen" is associated with about $70 of additional revenue generated, while the longer it takes for information to diffuse to an individual, the less productive they are.

2. Theory & Literature

2.1. The Central Role of Information in Diffusion Studies

Most diffusion studies take information diffusion as a central starting point from which to investigate the propagation of innovations or the potential success of targeted marketing campaigns. As a result, our understanding of underlying information dynamics that drive influence remains underdeveloped. Ironically, small variations in assumptions about how information spreads can drastically alter insights provided by diffusion models. For example, if news of an innovation spreads differently than knowledge about how to use it, or if status, authority or the functions of relationships moderate the flow of information from one individual or group to another, markedly different diffusion outcomes can be derived from the same basic propagation models. In this section we highlight the role of information in current diffusion studies and describe critical areas of knowledge that remain underdeveloped.

Theories of the diffusion of innovations (e.g. Rogers 1995), which since early studies of the diffusion of hybrid corn (Grilliches 1957, Ryan & Gross 1943) have been applied to phenomena as wide ranging as the spread of democracy (e.g. Wejnert 2002) and the adoption of technological innovations (e.g. Coleman, Katz & Menzel 1966), rely on information diffusion as a central mechanism driving adoption decisions. Potential adopters are exposed to new innovations and are convinced to adopt through "processes by which participants create and share information with one another in order to reach mutual understanding" (Rogers 1995: 17). As Rogers (1995: 17-18) describes, "the essence of the diffusion process is the information exchange through which an individual communicates a new idea to one or several others."

Information diffusion drives word of mouth contagion and viral marketing. Much of this research is concerned with maximizing the spread of influence through a social network by identifying influential nodes likely to "trigger" pervasive information cascades (e.g. Domingos & Richardson 2001, Kempe, Kleinberg, Tardos 2003), or enumerating characteristics of information cascades, such as the empirical distributions of their depth and structure (e.g. Leskovec, Singh, Kleinberg 2006). For example, Leskovec, Singh & Kleinberg (2006: 1) find that cascades in online recommendation networks "tend to be shallow, but occasionally large bursts of propagation appear" such that "the distribution of cascade sizes is approximately heavy-tailed." They also find different cascade properties across different products (DVDs, books and music), and identify various structural cascade patterns that recur in their data. Agent based models also apply notions of information cascades to revenue forecasts for various product categories such as movie box office receipts (De Vany & Lee 2001).

Two fundamental models have emerged to explain the diffusion of influence in contagion and innovation diffusion: threshold models and cascade models. Threshold models posit that individuals adopt innovations after reaching and surpassing their own private "threshold" of influence (e.g. Granovetter 1978, Schelling 1978). As others close to them adopt an innovation, each subsequent adoption[1] brings an actor closer to the threshold of influence required for their personal adoption. Individuals may be "close" in terms of their physical proximity, their direct interactions, or the equivalence of their social status or roles (Burt 1987). Cascade models on the other hand posit that each time a proximate individual adopts, the focal actor adopts with some probability that is a function of their relationship (e.g. Kempe, Kleinberg, Tardos 2003).

In both threshold models and cascade models, information sharing encourages adoption by spreading awareness (Meyer & Rowan 1977, Dewar & Dutton 1986), facilitating mimetic pressures (e.g.

[1] Often weighted by some factor that describes why one proximate node is more influential than another.

Coleman et. al. 1966), or through direct peer-to-peer influence (e.g. Rogers 1995). Each of these mechanisms involves significant information exchanges between adopters and non-adopters about the existence of a given innovation or product (awareness), its contextual uses and advantages (influence), or positive signals from the adoption of others (imitation, mimetic pressure). However, these communication processes, which are themselves complex, varied and subject to a host of underlying social norms and structural constraints, are rarely theorized. For instance, a great deal less information is necessary to communicate the existence of an innovation than its relative costs and benefits. In certain circumstances information may flow freely, spreading through social interaction without regard to status or competition, whereas in competitive or hierarchical environments the desire to maintain information advantages or authority may facilitate some information transfers while limiting others. Both threshold and cascade models assume an information transmission between adopters and non-adopters, but rarely specify the nature of the information or the conditions under which exchanges take place. Rather, the diffusion process is typically tested under various assumptions about the distribution of thresholds or dyadic adoption probabilities in the population. In fact, as Kempe, Kleinberg, Tardos (2003: 2) explain "the fact that [thresholds] are randomly selected is intended to model our lack of knowledge of their values."[2]

Information diffusion also underlies several well known theories of dynamical trading behavior in financial markets. Hirshleifer et. al. (1994) demonstrate that temporal asymmetries in the diffusion of information to various traders create abnormal profits for those that are informed (and informed early) and explain seemingly irrational trading equilibria, such as "herding" or outcomes based on "follow the leader" strategies, that seem to contradict rational private valuations. Since the consequences of financial trading decisions are actualized in relation to the collective decisions of the market, the receipt of market information and the exact timing of when investors become aware of a given piece of information "may be even more important than the accuracy of the information" in explaining profits and collective trading

[2] A related body of literature, stemming from Stanley Milgram's famous study of the "small-world phenomenon," studies network structures that support efficient search in local networks to explain the ability of actors to find short paths to targets through large networks with only local information (e.g. Watts 1999, Kleinberg 1999, 2001; Adamic et. al. 2001).

behaviors (Hirshleifer et. al. 1994: 1688). In this way, cascades of information that pulse through the population of financial traders can separate winners from losers, determine equilibria, and create collective tendencies that contradict privately held information. Yet, in these models temporal asymmetries in information acquisition are taken as given, and how and why these systematic asymmetries arise remains unknown.

Finally, there is a body of literature on knowledge transfers and performance, which explores relationships between network structure, knowledge sharing and the performance of teams (e.g. Reagans & Zuckerman 2001, Cummings 2004). These studies show that knowledge sharing and the network structure that guides information flows can impact the productivity and performance of work groups. However, most of this work remains "agnostic with respect to content" (Hansen 1999: 83) and only considers whether knowledge is flowing rather than the type of knowledge being transferred. A related literature examines the conditions under which knowledge and information flow efficiently between business units and individuals (e.g. Hansen 1999, 2002, Reagans and McEvily 2003), although this work focuses on dyadic transfers of information between business units or individuals rather than on the diffusion of information from an originator to all potential recipients in the organization.

While these related literatures highlight the importance of information dynamics and in particular information diffusion in organizations, there is scant research on the subject. Although processes and circumstances that govern the movement of information through populations form a core micro-level foundation on which diffusion models are based, little research examines enablers and constraints of the movement of information itself. In order to understand the underlying dynamics driving the diffusion of innovations, dynamical trading behavior, information cascades that empower viral marketing, and global properties of how information transfers in organizations affect individual and group performance, it is important to develop theory about organizational information dynamics – how information moves through social groups within and between organizations.

2.2. Information Dynamics in Organizations

In this section we develop theory about a limited subset of information dynamics: the combined effects of information content and social structure on the diffusion of information in organizations. More specifically, we focus on the diffusion of event news and discussion topics as a function of both dyadic and individual network, demographic, authority, hierarchy and co-work characteristics.

Although some information diffusion studies exist, they typically rely on computer simulations of a handful of agents (e.g. Yamaguchi 1994, Buskens & Yamaguchi 1999, Newman et. al. 2002, Reagans & Zuckerman 2006), treat information as a uniform, homogeneous concept (e.g. Buskens & Yamaguchi 1999, Wu et. al. 2004, Newman et. al. 2002, Reagans & Zuckerman 2006), and focus on global properties that maximize the diffusion of a given piece of information (e.g. Newman et. al. 2002). While these studies provide an important starting point for nascent explorations of information dynamics, significant theoretical underpinnings remain unexplored.

Most current conceptualizations of information dynamics assume information is homogeneous, without separately theorizing different types of information. For example, Bushkens & Yamaguchi (1999) improve on Yamaguchi (1994) by introducing the assumption that information exchange is non-rival (that information is retained by the sender in an exchange) into their agent based model, but maintain the tradition of assuming information homogeneity. Reagans & Zuckerman (2006) model the passage of "bits" through different network structures and assume the "uniqueness" of each bit, but theorize that each bit passes between individuals with uniform probability and without regard to how the uniqueness of a bit is related to its transmission probability. Wu et. al. (2004) test decay processes in information diffusion using email data from employees at HP Labs. However, by studying the diffusion of attachments and links, they too assume different types of information diffuse uniformly.

The assumption of information homogeneity is problematic in light of prior evidence on differences in information transfer effectiveness across different types of information. Some information is simply "stickier" (Von Hippel 1998) and more difficult to transfer (Hansen 1999) due to its specificity (Rosenberg 1982, Nelson 1990), complexity (Uzzi 1996, 1997, Hansen 1999), the amount of related knowledge of the receiver (Cohen & Levinthal 1990, Hansen 2002), and the degree to which the information is declarative or procedural (Cohen & Bacdayan 1994). These factors make it unlikely that all types of information exhibit uniform transfer rates or diffusion patterns across different relationships or social structures. As Wu et. al. (2004: 328) point out: “There are … differences between information flows and the spread of viruses. While viruses tend to be indiscriminate, infecting any susceptible individual, information is selective and passed by its host only to individuals the host thinks would be interested in it.” We take this departure from epidemic models of disease a step further. Many other factors can influence the diffusion of a given type of information beyond the senders’ perception of the receivers’ interest. We hypothesize that the strength and function of social relationships, geographic proximity, organizational boundaries, and hierarchy, authority and status differences across social groups effect the movement of information, and have different effects across different types of information.

Finally, current work focuses more on global social network properties that maximize the diffusion of a given piece of information through a population (e.g. Watts & Stogatz 1998), than on drivers of individual access to information cascades. There is also a lack of robust empirical evidence on different types of information diffusing through real social groups. Although Wu et. al. (2003) study the diffusion of attachments and links in email, which are content agnostic, and Newman et. al. (2002) model the spread of computer viruses over an observed sample of email address books, neither of these works nor any other work we found examines empirical evidence on the diffusion patterns of different types of information and the predictors of access to those diffusion processes among individuals in an organization.

We therefore propose three extensions to the current body of work on information dynamics. First, we propose that in addition to global structural properties of social groups, there exist hierarchal, demographic and functional enablers and constraints of information diffusion. For example, we hypothesize that information may diffuse more readily vertically (or laterally) through an organizational hierarchy due to authority or status differences, or more quickly through functional relationships than strong ties per se. Second, we hypothesize that different types of information content diffuse differently. We argue that characteristics of information are pertinent to how it diffuses. Third, we argue that content and structure jointly predict the diffusion path of a given piece of information, and that different social and structural factors will govern the diffusion of different types of information. We hypothesize that simple, declarative news will diffuse relatively indiscriminately without regard to the strength of ties, the function of relationships, or authority; in contrast, the diffusion of discussion topics, characterized by shallow cascades and back and forth conversations, will be influenced by the strength of ties, the function of relationships and by authority and hierarchy. We develop the theory behind each of these extensions in the next two sections, and then test our theory by studying several thousand diffusion processes observed over two years in our email data.

2.2.1. Social Drivers of Organizational Information Diffusion

Several factors could influence the flow of information in organizations. Theories on how status hierarchies, demography, social networks and formal organizational structures facilitate and hinder relationship building guide our thinking on how information may diffuse. We also utilize semi-structured interview data from employees of our research site to identify potential drivers of information diffusion in our setting. In the end, we hypothesize four categories of factors that may impact information dynamics in organizations: demography, organizational hierarchy, tie and network characteristics and functional task characteristics. Each of these categories includes individual and dyadic dimensions of interest.

Demography. Individuals' demographic characteristics and dissimilarity are likely to affect social choices about information seeking and information transmission. Similar individuals tend to flock together in social relationships – a phenomenon known as homophilly (McPherson, Smith-Loving, & Cook 2001), creating parity in perspectives, information and resources across demographically similar individuals in organizations (Burt 1992, Reagans & Zuckerman 2001). Demographic diversity has also been shown to introduce social divisions and create tension in organizational work groups (Pfeffer 1983), reducing the likelihood that individuals of dissimilar demographic backgrounds will go to each other for advice or pass information to one another. We therefore measure the demographic characteristics of individuals and the demographic dissimilarity of pairs of individuals at our research site, focusing on age, gender, and education, three of the most important variables in organizational demography.[3]

Organizational Hierarchy. There is good reason to suspect that information flows are affected by formal organizational structures. Formal structures define reporting relationships and work dependencies that necessitate communication and coordination (Mintzberg 1979). Mangers and employees frequently communicate to manage administrative tasks even when they are not working on the same projects, and the importance of notification for accountability, and recognition for upward mobility encourages dialogue and information exchange along hierarchical lines. Embedded within formal organizational hierarchies are gradients of status and authority that may also guide information flows. Task level knowledge and familiarity with customers, specialized technology, competitors and new market opportunities often resides with mid-level managers rather than upper management, leading some firms to decentralize decision rights to take advantage of local knowledge (Dessein 2002). If this is the case, we might expect information to flow most quickly to employees in mid-level management positions. In our organization, teams are composed hierarchically, generally consisting of one partner, one consultant, and one researcher. As project teams are organized hierarchically, task related information is likely to flow vertically rather than laterally across individuals of the same organizational rank.

Tie & Network Characteristics. Informal networks are also likely to impact information diffusion in organizations. A vast literature treats the relationship between social network structure and organizational performance (e.g. Burt 1992, Gargiulo and Benassi 2000, Ahuja 2000, Sparrowe et al. 2001, Cummings & Cross 2003). Although most of this work does not measure information flows explicitly, evidence of a relationship between performance and network structure is typically assumed to be due in part to the information flowing between connected actors (Burt 1992, Reagans & Zuckerman 2001). Some network research does explicitly treat information transmission. For example, several studies have shown that the strength of ties increases the transfer of information and knowledge in dyadic

[3] Unfortunately, we do not have access to race or organizational tenure variables, although we do have measures of industry tenure which we consider a functional dimension of social distance rather than a demographic one.

relationships and in particular improves the effectiveness of transfers of tacit knowledge (Uzzi 1996, 1997). As individuals interact more frequently, they are likely to pass information to one another. We therefore measure the *strength of communication ties* by the total volume of email passing between each pair of individuals in our network. Other studies demonstrate that '*betweenness centrality*' $B(n_i)$ (Freeman 1979),[4] which measures the probability that the individual will fall on the shortest path between any two other individuals linked by email communication, predicts the total amount of knowledge acquired from other parts of the network (Hansen 1999), and that actors with high network *constraint* C_i (Burt 1992: 55),[5] which measures the degree to which an individual's contacts are connected to each other (a proxy for the redundancy of contacts), are less privy to new information (Burt 1992). We therefore measure individuals' betweenness centrality and their constraint as follows:

$$B(n_i) = \sum_{j<k} g_{jk}(n_i) \Big/ g_{jk} ;$$

$$C_i = \sum_j \left(p_{ij} + \sum_q p_{iq} p_{qj} \right)^2, \quad q \neq i, j.$$

The total amount of email communication that individuals engage in is also likely to drive their access to information and how quickly they see it (Cummings & Cross 2003). Finally, a great deal of evidence links physical proximity to communication between actors (e.g. Allen 1977), however, in the case of email communication, it could be that greater geographic distance is associated with more email communication between actors who find it costly to meet and communicate face to face. We therefore measure physical proximity by whether or not two people work together in one of the firm's fourteen offices.

Task Characteristics. Working relationships are conduit of communication and information flow. As individuals work together, they develop stronger bonds of trust and experience that encourage them to

---

[4] Where g_{jk} is the number of geodesic paths linking *j* and *k* and $g_{jk}(n_i)$ is the number of geodesic paths linking *j* and *k* involving *i*.

[5] Where $p_{ij} + \sum p_{iq}p_{qj}$ measures the proportion of *i*'s network contacts that directly or indirectly involve *j* and C_i sums this across all of *i*'s contacts.

exchange information. Working relationships also necessitate exchange of task related information and create relatively stable and enduring ties that individuals rely on for advice on future projects. However, relationships can decay over time when they are not active (Burt 2002), and repeated relationships are more likely to create long term conduits through which information diffuses. We therefore measure the strength of project co-work relationships by the number of projects employees have worked on together. We also know from the literature on absorptive capacity (Cohen & Levinthal 1990) that related knowledge helps individuals consume new information, and individuals in related fields and of related expertise are more likely to swim in the same pools of information. We therefore also measure whether or not employees work in the same expertise area in the firm. For instance, some recruiters focus on health care positions, or technology positions, as well as in certain regions. We measure specialization using a dummy variable whose value is one for employees working in the same expertise area. Finally, as with demographic distance, we expect information to diffuse more easily between employees in the same industry cohort. Pfeffer (1983) has noted the importance of organizational cohorts in maintaining lines of communication and organizational relationships. The same logic can be applied to industry cohorts. Employees with the same industry tenure have likely been through similar work related milestones and may already be familiar with each other through industry relationships. In addition, more experienced workers may rely on other experienced workers for information, while status and authority prevent the less experienced from sharing as much information across industry tenure gradients.

2.2.2. Dimensions of Information Content that Affect Diffusion

Apart from social context, characteristics of information itself are likely to affect diffusion patterns and the probability that a given individual sees a piece of information. Previous research has shown that certain types of information are "stickier" and have higher transfer costs (Von Hippel 1998), and several dimensions of information may influence when it is shared and how it diffuses. In what follows we describe two contrasting types of information we call 'event news' and 'discussion topics,' which serve as vignettes for comparison across information types. These vignettes are intended as archetypes, not mutually exclusive categories. Information is contextual – one may have lengthy discussions about event news or hear news about an important discussion. The point of illustrating these two archetypes is to evoke the underlying characteristics of information that are most likely correlated with what we eventually measure and test: the diffusion patterns and usage characteristics of particular words in email. Our contention is that information of the types described here are likely to diffuse in a certain way and that words that exhibit these diffusion patterns proxy for information content with the characteristics we describe. However, the relationship between the types of information described and the diffusion patterns observed is not critical to the argument. In the end, our goal is to demonstrate that different characteristics of people, relationships and social structure affect access to different types of information with different aggregate diffusion patterns.

Event News. We define 'event news' as simple, declarative, factual information that is likely triggered by an external event and is of general interest to many people in the organization. In the context of our research site, employees may learn of forthcoming layoffs at a source company, a forthcoming change in company policy or a significant change in top management through a rapid pervasive information cascade that travels quickly and pervasively throughout the organization. Such information is likely to be simple, declarative and factual, informing recipients of an event that has or will soon take place. Such information is of general interest to all employees in the firm and is likely to be widely shared amongst many people and across organizational and hierarchical boundaries.

Discussion Topics. We define 'discussion topics' as more specific, complex, and procedural, characterized by back and forth discussion of interest to limited and specialized groups of people. At this firm, work groups discuss particular projects, and most frequently have back and forth discussion about particular candidates or clients. So, for instance, a particular candidate's name may be discussed back and forth as their merits for a particular job are being considered. Teams specializing in filling nursing job vacancies in the south eastern United States may circulate names amongst other recruiters who specialize in the same type of job in the same region. When we presented this typology to an IT employee of a different firm, they immediately related to this category of information by giving the example of "ITIL"

(Information Technology Infrastructure Library), a framework of best practice approaches for delivering IT services that has since received ISO accreditation. This informant indicated that the IT staff in her firm would frequently discuss "ITIL," and that we would likely find reference to it in back and forth email discussions in her group, but that it was unlikely that anyone else in the firm would discuss it.

Theories of information transfer support our distinctions between event news and discussion topics. Hansen (1999) demonstrates that complex knowledge is more difficult and costly to transfer, and shows that in dyadic relationships, strong ties are necessary to effectively transfer complex knowledge. There is also a theoretical distinction made between declarative and procedural information (Cohen & Bacdayan 1994: 557), with the former consisting of "facts, propositions and events," and the later of information about how to accomplish tasks, activities or routines. We argue that event news is more likely to be simple and declarative, and thus more easily transferred widely amongst different types of people. Rosenberg (1976), Nelson (1990) and Von Hippel (1998) also make the distinction between "specific" and "generic" information and knowledge, arguing that, in contrast to the specific, "generic knowledge not only tends to be germane to a wide variety of uses and users. Such knowledge is the stock in trade of professionals in a field … so that when new generic knowledge is created anywhere, it is relatively costless to communicate to other professionals" (Nelson 1990: 11-12, as quoted in Von Hippel 1998: 431).[6] Finally, transfers of information and knowledge are more effective among individuals with related knowledge (Cohen & Levinthal 1990, Hansen 2002). Those with similar expertise or specialization are more likely to share information more often and more effectively due not only to their shared common interest in certain information, but also their ability to more effectively communicate ideas based on their "common ground" (Cramton 1991). We therefore hypothesize that diffusion of event news, which is likely to be simple, declarative, factual and relevant to a wide variety of users, will be driven by the

[6] As pointed out by Orlikowski (2002), there is an important distinction to be made between knowledge and information; and in her case between knowledge and knowing. Without exploring the vast theoretical details of this distinction, we assume that the characteristics that make knowledge complex (and therefore costly to transfer) in turn influence characteristics of the information employees in this firm send and receive. Specifically, we argue that they are likely to *send* generic information to "a wide variety of users" (Nelson 1990: 11-12), without making the deeper assumption that what makes knowledge complex also makes information that communicates that knowledge complex.

demographic and network factors theorized to constrain interactions due to homophilly and network constraints.

H1: *Access to event news is driven by demographic similarity, and structural characteristics of network position such as betweeness centrality, constraint and path length.*

On the other hand, we argue that information passed back and forth amongst small groups is likely to be task specific and reflect information relevant to those socially and organizationally proximate to the originator. At our research site, since work groups are organized vertically along the organizational hierarchy, with teams composed of one member from each organizational level, we expect task related information to be passed vertically up and down the organizational hierarchy, rather than laterally between members of the same organizational level. We therefore hypothesize that diffusion of discussion topics, which are likely to be complex, specific, procedural and discussed in small groups, will be driven not only by demographic and network factors, but also by project co-work relationships and organizational hierarchy.

H2: *Access to discussion topics is driven by demographic similarity, and structural characteristics of network position such as betweeness centrality, constraint and path length, as well as by task characteristics and organizational hierarchy.*

Finally, better information should improve performance. Individuals who learn of novel information are in a position to take actions on that information which ultimately speed the completion of projects and the generation of revenue. In particular, executive recruiters can be thought of as human information processors – they match information about job candidates with information about positions available. Better information improves the timeliness and quality of these matches.

H3: *Project completion and revenue generation by individuals will be correlated with the amount and timeliness of novel information observed by those same individuals.*

3. Methods

3.1. Data

Data for this study come from three sources: (i) accounting data detailing project co-work relationships, organizational positions, physical locations, projects completed and revenues generated, (ii) email data captured from the firm's corporate email server, and (iii) survey data that capture demographic characteristics, education, and industry tenure.

Internal accounting data describe: (i) revenues generated by individual recruiters, (ii) contract start and stop dates, (iii) projects handled simultaneously by each recruiter, (iv) project team composition, and (v) job levels of recruiters and placed candidates. These provide excellent performance measures that can be normalized for quality.

Email data cover 10 months of complete email history at the firm. The data were captured from the corporate mail server during two equal periods from October 1, 2002 to March 1, 2003 and from October 1, 2003 to March 1, 2004. We wrote and developed capture software specific to this project and took multiple steps to maximize data integrity and levels of participation. New code was tested at Microsoft Research Labs for server load, accuracy and completeness of message capture, and security exposure. To account for differences in user deletion patterns, we set administrative controls to prevent data expunging for 24 hours. The project went through nine months of human subjects review prior to launch and content was masked using cryptographic techniques to preserve individual privacy. Spam messages were excluded by eliminating external contacts who did not receive at least one message from someone inside the firm.[7] Participants received $100 in exchange for permitting use of their data, resulting in 87% coverage of recruiters eligible to participate and more than 125,000 email messages captured. Details of data collection are described by Aral, Brynjolfsson & Van Alstyne (2006). Since cryptographic techniques were used to protect privacy, we observe unique tokens for every word in the email data and construct diffusion metrics based on the movement of words through the organization in email. Methods for analyzing the diffusion of email content are described in greater detail in § 3.2.

[7] In this study we focus on email sent to and from members of the firm due the difficulty of estimating accurate social network structures without access to whole network data (see Marsden 1990).

Survey questions were generated from a review of relevant literature and interviews with recruiters. Experts in survey methods at the Inter-University Consortium for Political and Social Science Research vetted the survey instrument, which was then pre-tested for comprehension and ease-of-use. Individual participants received $25 for completed surveys and participation exceeded 85%.

Table 1: Descriptive Statistics

Variable	Obs.	Mean	SD	Min	Max
Gender (Male = 1)	832419	.50	.49	0	1
Age Difference	562650	12.22	8.81	0	39
Gender Difference	832419	.50	.49	0	1
Education Difference	562650	1.38	1.26	0	6
Email Volume	809613	1474.65	1129.95	0	4496
Strength of Tie	832419	11.71	36.90	0	464
Path Length	832419	2.61	2.68	0	10
Geographic Proximity (Same Office = 1)	832419	.30	.46	0	1
Friends in Common	832419	6.70	5.75	0	35
Betweeness Centrality	809613	36.77	36.81	0	165.73
Constraint	809613	.213	.09	0	.51
Prior Project Co-Work	832419	.26	1.33	0	19
Industry Tenure Difference	562650	10.08	8.32	0	38
Same Area Specialty	832419	.10	.30	0	1
Managerial Level Difference	832419	.86	.71	0	2
Partner	832419	.36	.48	0	1
Consultant	832419	.40	.48	0	1
Researcher	832419	.22	.41	0	1

Our data collection avoids several important limitations of data used in previous studies of networks and information transfer (e.g. Hansen 1999, 2002), and information diffusion in particular. First, as noted in Essay 2, traditional network studies tradeoff exists comprehensive observation of whole networks and the accuracy of respondents' recall. Respondents are shown to have difficulty recalling their networks (e.g. Bernard et. al 1981), especially when assessing network connections among individuals socially distant to themselves (Krackhardt & Kilduff 1999). The inaccuracy of respondent recall and the bias associated with recall at social distance creates inaccurate estimates of networks (Kumbasar, Romney & Batchelder 1994). The estimation challenge posed by this approach is that network metrics are incredibly sensitive to the completeness of data (Marsden 1990). Observation of whole networks in their entirety is important for creating unbiased estimates of actors' topological positions. By capturing network data in email we not only address the issue of respondent recall, we are able to capture almost every employee in the firm with a high degree of accuracy. 87% of the recruiters eligible for our study agreed to participate in email data collection, and given that our inability to observe the remaining 13% is limited to messages between two employees who both opted out of the study, we have a relatively unbiased view of the communication network with nearly full coverage of the firm.

Second, estimation of diffusion processes from incomplete data can be especially problematic (Greve, Tuma & Strang 2001). As we record the time that emails were sent and received employing time stamps in email data, we avoid time aggregation bias – where event times are recorded imprecisely or with systematic error. Time aggregation occurs when observations of sequential events are recorded as having occurred simultaneously due to coarse grained observations over large blocks of time like days, weeks or months. Previous research has shown that time aggregation creates severe bias when time intervals are large relative to the average time to an event (Petersen 1991, Petersen & Koput 1992, Greve, Tuma & Strang 2001). In contagion models, time aggregation also creates bias in contagion updating. As previous adopters can influence future adoption, imprecise time aggregation can inaccurately characterize time dependent influences on actors. Also, without relatively complete coverage of the population, sparse sampling can create significant bias in estimates of the impact of relational variables such as social proximity. When the sampling probability is small, coefficient estimates can be bias downward to zero, and the extent of the bias increases as the sampling probability decreases (Greve, Tuma & Strang 2001). Our near complete coverage of the firm and the precision with which we record event times in email data help us avoid these sources of bias.

Table 2: Correlation Matrix

Measure	1	2	3	4	5	6	7	8	9	10	11	12
1. Gender (Male = 1)	1.00											
2. Age Difference	.06	1.00										
3. Gender Difference	.27	.06	1.00									
4. Education Difference	.08	.09	.06	1.00								
5. Email Volume	-.11	-.00	-.04	-.04	1.00							
6. Strength of Tie	-.04	-.06	-.01	.01	.30	1.00						
7. Path Length	-.11	.00	.00	-.04	-.37	-.18	1.00					
8. Geographic Proximity	-.06	.09	-.01	-.03	.06	.17	-.06	1.00				
9. Friends in Common	.04	-.02	.03	.02	.50	.38	-.40	.08	1.00			
10. Betweeness Centrality	.06	.01	.01	-.07	.66	.22	-.31	.03	.48	1.00		
11. Constraint	-.18	-.05	-.05	.01	-.26	-.06	.54	-.05	-.34	-.33	1.00	
12. Prior Project Co-Work	.02	-.01	.01	-.01	.01	.37	-.09	.08	.15	.02	-.06	1.00
13. Industry Tenure Difference	.11	.50	.06	-.05	-.09	-.08	.03	.07	-.07	-.08	-.12	.05
14. Same Area Specialty	-.01	-.16	-.00	.01	.12	.40	-.12	.27	.19	.05	-.04	.33
15. Managerial Level Difference	.05	.52	.05	.11	.03	-.10	.01	-.03	.01	.02	-.07	.03
16. Partner	.21	.06	.06	.02	-.06	-.05	-.14	-.06	.07	-.03	-.31	.09
17. Consultant	-.12	-.07	-.03	-.04	-.31	-.09	.29	-.26	-.19	-.22	.19	-.01
18. Researcher	-.09	.01	-.03	.03	.40	.15	-.17	.35	.13	.27	.12	-.08

	13	14	15	16	17	18
13. Industry Tenure Difference	1.00					
14. Same Area Specialty	-.12	1.00				
15. Managerial Level Difference	.50	-.21	1.00			
16. Partner	.26	-.05	.23	1.00		
17. Consultant	-.13	-.07	-.21	-.56	1.00	
18. Researcher	-.13	.14	-.01	-.44	-.51	1.00

3.2. Identifying Heterogeneous Information Types

Our goal is to identify samples of words that exhibit different diffusion patterns and whose usage characteristics reflect those one would expect to find exhibited by event news and discussion topics. We defined 'event news' as simple, declarative, factual information that is likely triggered by an external event and of general interest to many people in the firm. Given these criteria, we assume event news is characterized by a spike in activity and a rapid pervasive diffusion to members of the organization, followed by a decline in use. More accurately, we assume that words that exhibit these usage characteristics are likely to be event news or information whose theoretical characteristics are similar to those described for event news. At the same time we are interested in identifying a sample of 'discussion topics,' which we define as more complex, specific to a group of people, containing more procedural information and in Von Hippel's (1998) parlance "sticky." We expect this information to exhibit more shallow diffusion, characterized by 'back-and-forth' conversation among smaller groups for more extended periods. In this section we describe our analytical method for identifying these two types of information.[8]

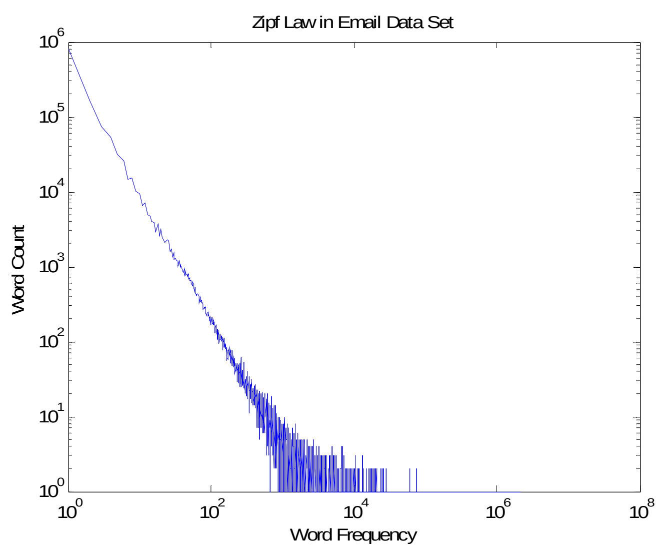

Figure 1. Distribution of Word Frequencies in Email Data

[8] We thank Tim Choe for his tireless coding efforts that extracted and manipulated the email data described in this section, and created the graphics depicting email discussions. This work was done as part of his research assistantship and Master's thesis work under my supervision as well as that of Erik Brynjolfsson and Marshall Van Alstyne.

We began with a dataset consisting of approximately 1.5 million words whose frequencies were distributed according to the standard Zipf's Law distribution, with many highly infrequent words and smaller and smaller numbers of more frequent words, as shown in Figure 1. We initially eliminated words unlikely to cascade through the firm by culling the most infrequent words (term frequency < 11), words that are commonly used every week and are likely to be common language (words in at least one email in every week of the observation period), and words with low term frequency- cumulative inverse document frequency (tf-cidf), a common metric used to identify spikes in usage (Gruhl et. al. 2004).[9] The tf-cidf constraint chooses words that record a spike in weekly usage greater than three times the previous weekly average, retaining words likely to cascade or diffuse. Together, these three methods reduced the number of emails under consideration to approximately 500,000, 495,000, and 120,000 words respectively. From these 120,000 candidates we sampled words likely to be event news and discussion topics.

In selecting event news, we sought words whose usage was characterized by a spike in activity and a rapid, pervasive diffusion to members of the organization, followed by a decline in use. To chose such a sample we chose words seen by more than 30 people with a coefficient of variation one standard deviation above the mean. We chose the 30 person threshold by first determining the percentage of the firm that used common words.

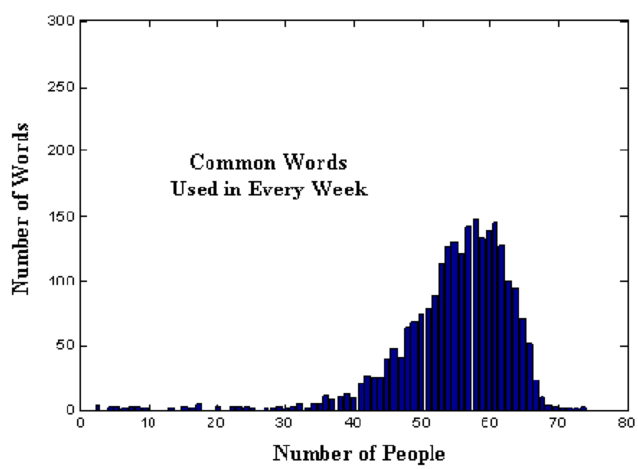

Figure 2. Distribution of Common Words over Employees

[9] Our use of the cutoff of 11 produced similar results as cutoffs in the neighborhood of 11.

The distribution of employees using common words provides a robust contextual proxy for information that is 'widely used' in the firm. By examining a histogram of the distribution of the number of common words over the number of people who used those words, we determined that most common words were used by between 30 and 70 people. To be conservative, we selected any word seen by more than 30 people as a potential observation of event news.

In order to select words likely to display rapid propagation, of those words that reached 30 people or more, we selected words with a high coefficient of variation of activity - words with bursts of activity in some weeks relative to others. The coefficient of variation has been used in previous work to identify spikes in topic frequency in blog posts (Gruhl et. al. 2004) and is a good measure of dispersion across data with heterogeneous mean values (Ancona & Caldwell 1992).[10] Observations of a large number of people suddenly using a word much more frequently than usual are likely to indicate information triggered by some external event that is diffusing through the organization.[11] We therefore select words with a coefficient of variation one standard deviation above the mean. The result is a sample of 3275 words that are at first rarely used, and then suddenly are used much more frequently and by more than 30 people in the firm, followed by a decline in use. An example of an event news item is shown in Figure 3. The blue line represents the cumulative number of people to have seen the word in email, while the green line represents the frequency of use in email. As the figure shows, the word is rarely used and seen by less than 5 people in the first 80 days of the observation period, after which there is spike in activity accompanied by diffusion to nearly 60 people, followed by a decline in use. This example is descriptive of the words in our event news sample.

[10] The coefficient of variation is simply the standard deviation of the number of emails per week that contain a given word divided by the mean number of emails per week that contain that word.

[11] While the argument could be made that a spike of activity is no guarantee of diffusion, occurrences of significant numbers of event driven spikes in usage that are not part of diffusion processes are only likely to downward bias estimates of the influence of relationship based metrics on the likelihood of seeing a given piece of information, making our estimates more conservative.

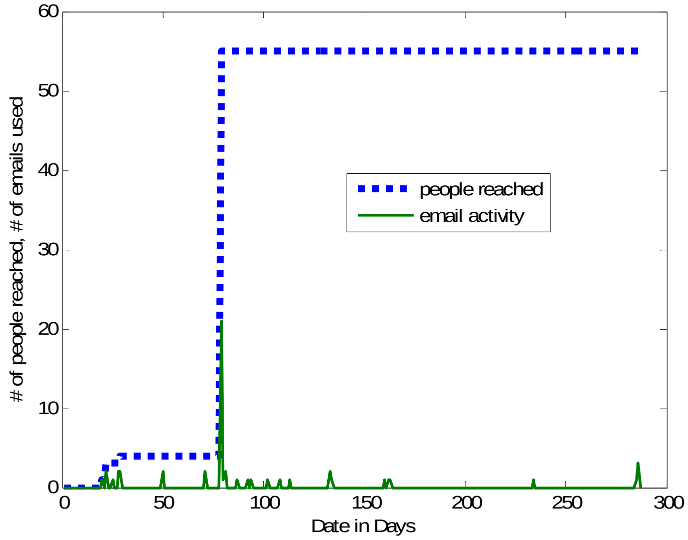

Figure 3. An Example Event News Item

We then selected a sample of discussion topic words. We began by identifying individual information cascades by tracing the flow of words through the organization. By ordering email time stamps and recording emails as they are sent and received, we constructed subgraphs of the network that traced a given word. As we were looking for discussion, we looked for words that were used in back and forth discussions via emails. We recorded a link if the receiver sent the word on within a small time window after receiving it to exclude words that were simply used later rather than received and sent on. This process created tree like subgraphs that recorded the path of a word diffusing from person to person. We characterized these subgraphs by their depth and their average depth – depth recording the number of people and average depth recording the sum of the depths of all connected subgraphs for a word divided by the total number of connected subgraphs. This process yielded approximately 3000 words with an average depth greater than 1.5. While this process identified words likely to be discussion topics, we were concerned that selecting our sample based on the same data that generated some of our social network graphs, would introduce endogeneity into our statistical specifications by 'selecting on the dependent variable.' We therefore chose a more parsimonious approach that did not use links between people in email as a selection criterion. We simply selected words where users both received and sent the word in email. This simple criterion selected approximately 4100 words from the original candidate set. An example of the usage characteristics of discussion topic words is shown in Figure 4.

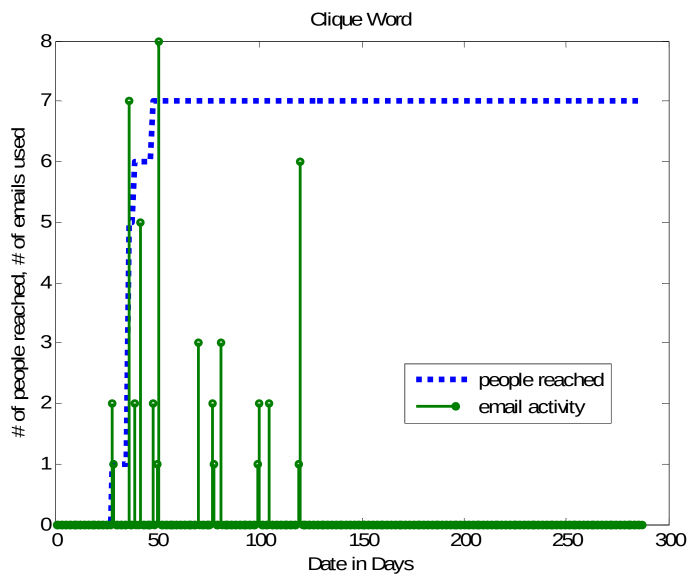

Figure 4. An Example Discussion Topic Item

Words in this sample display a lack of use, followed by a shallow diffusion to a limited number of people accompanied by an extended back and forth discussion, which in the case of the word show in Figure 4 lasts close to 3 months. These words are shared in back and forth conversation as demonstrated by the subgraphs of discussion topics shown in Figure 5.

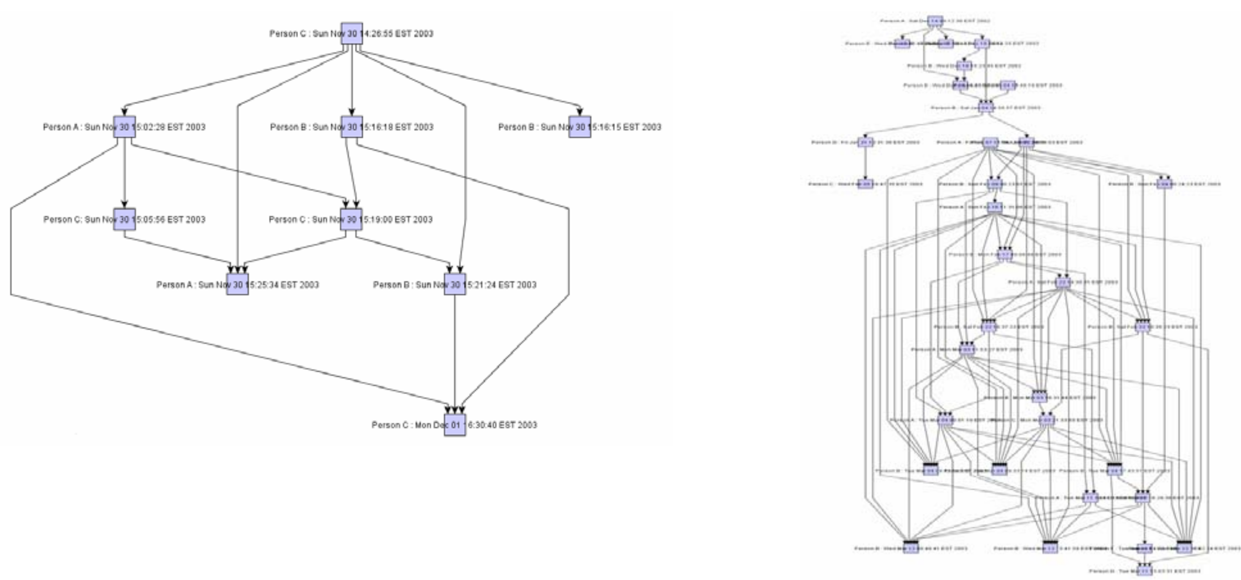

Figure 5. Discussion Paths in Discussion Topic Items

In the subgraph on the left, three people share a word back and forth in discussion. Person C sends the word to person A and person B, who each reply to person C, who in turn replies to them and subsequently receives a reply back from both B and C. The subgraph on the right shows a more complex back and forth discussion. This type of discussion, rather than exemplifying information that individuals disseminate widely among members of the firm, is likely specific to the individuals in this exchange. It is less likely to be declarative than procedural in the sense that those who receive the information react to it with a reply that triggers further ongoing discussion. In contrast to information that diffuses widely across the organization without back and forth discussion, we expect access to this type of information to be driven by functional relationships, strong ties and to move up and down the organizational hierarchy in line with the composition of teams which themselves are organized hierarchically.

After selecting these words based on their usage characteristics, we tested whether our information types exhibited significantly different usage characteristics and diffusion properties. As Leskovec, Singh & Kleinberg (2006) have noted, information cascades are typically shallow, but sometimes characterized by large bursts of wide propagation. We wanted to make sure we captured both these phenomena in our data. We therefore summarized the usage characteristics of words along several dimensions including the number of emails containing the word, the number of people who used the word, the coefficient of variation of use, the number of emails per person that contain the word, the total diffusion time divided by the total time in use (as a proxy for use beyond the diffusion to new users), and the maximum number of people who see the word for the first time in a given day (a proxy for the maximum spike in activity). We then tested whether words in each category differed significantly across these dimensions by conducting t-tests of the differences of means across information types and dimensions. Table 3 lists the mean usage characteristics and diffusion properties of both event news and discussion topics. Our t-tests demonstrate that these information types differ significantly across all dimensions of interest related to their use and diffusion. Results of t-tests of the difference of sample means for relevant dimensions are shown in Column 3.

Table 3: Mean Usage Characteristics and Diffusion Properties of Information Types

Information Type	News	Discussion	t-statistic
Usage Characteristics & Diffusion Properties			
Number of Words	3235	4168	-
Potential Diffusion Events	245280	320470	-
Realized Diffusion Events	65145	9344	-
Number of Emails	236.21	17.69	27.69***
Mean Diffusion Depth	36.31	2.48	213.28***
Coefficient of Variation	1.46	4.11	90.53***
Emails Per Person	6.10	7.47	1.105***
Diffusion Time / Total Use Time	.97	.48	66.36***
Maximum New Users Per Day	9.38	1.60	61.51***

Note: * $p < .05$, ** $p < .01$, *** $p < .001$

3.3. Data Structure

We observe the diffusion of several thousand observations of each type of information from the original first use, which we define as the first occurrence of a given word in our data, to all employees in our sample. For each piece of information we observe whether a given employee received the word, the rank order in which they received the word relative to other employees, and the time between the first use of the word and the receipt of the word by each employee, constructed using the time stamps in email traffic. An observation, therefore is indexed by a word-recipient pair (one for each possible recipient in the firm) in which the first user is suppressed due to the one to one relationship between words and first users. For each word, our data log dyadic characteristics of each first user-recipient pair, such as the difference in their ages or industry tenures, for all potential recipients. Each observation also records individual characteristics of potential recipients, such as their gender, network position, or managerial level.

3.4. Statistical Specifications

We are interested in estimating the impact of hypothesized factors on the likelihood of seeing a strategic piece of information and seeing it sooner. Linear regression models are problematic when estimating temporal processes or likelihood outcomes for several reasons. Linear estimates of probabilistic outcomes create bias due to non-linearity near the upper and lower bounds of the likelihoods of discrete events, are not well suited to temporal processes in which outcome variables can be conditioned on previous events (Strang & Tuma 1993), and produce biased estimates of longitudinal data in which right censoring is present and pervasive (Tuma & Hannan 1984). For these reasons we specify logistic regression and hazard rate models of the diffusion of different kinds of information in our data.

We first estimate the influence of independent variables on the likelihood of receiving a given piece of information using a standard logistic regression model formalized in equation 1.

$$\ln\left(\frac{P(Y_i = 1)}{1 - P(Y_i = 1)}\right) = \alpha_i + \sum \beta_j X_i + \varepsilon_i \qquad [1].$$

In this model, we estimate the impact of independent variables of interest on the likelihood of receiving a given piece of information by email. Since many of the variables of interest are time invariant characteristics of individuals and relationships and because our email data overlap project accounting data by ten months, we group observations in email and in accounting data into a pooled logistic regression. The coefficient estimates describe the impact of a given variable on the likelihood of receiving the word during the ten months of email observation.

The logistic regression model estimates the likelihood of receiving information; however, one limitation of this type of model is that they are not well suited to the estimation of dynamic processes in which the ordered timing of events matters. In particular, in diffusion studies cross sectional estimates may wash away temporal variation and allow later events to influence the estimates of earlier diffusion (Strang & Tuma 1993, Burt 1987). We therefore estimate the rate of receipt of different types of information conditional on having received the information, using a Cox proportional hazard rate model of the speed with which employees receive information:

$$R(t) = r(t)^b e^{\beta X} \qquad [2],$$

where *R(t)* represents the project completion rate, *t* is project time in the risk set, and $r(t)^b$ the baseline completion rate. The effects of independent variables are specified in the exponential power, where β is a vector of estimated coefficients and *X* is a vector of independent variables. The coefficients in this model have a straightforward interpretation: β represents the percent increase or decrease in the rate at which information is seen associated with a one unit increase in the independent variable. Coefficients greater than 1 represent an increase in the rate of information diffusing to the receiver (equal to β - 1); coefficients less than 1 represent a decrease (equal to 1- β). In testing the proportional hazards assumption we found no compelling evidence of duration dependence in any variables and proceeded with traditional estimations of the Cox model.

There has been some debate regarding appropriate estimation strategies for non-linear models with group specific parameters (e.g. Chamberlain 1980, Hausman et. al. 1984, Greene 2001, Green, Kim & Yoon 2001, Beck & Katz 2001). Group specific parameters arise in hierarchical data when multiple observations are sampled from the same group, or in longitudinal or time-series cross-section panel data with repeated observations of the same individual. For example, samples of students drawn randomly from several school districts will exhibit group specific effects at the district level, just as repeated observations of individuals will exhibit time invariant characteristics specific to the individual in time-series cross-section panel data or longitudinal event analyses. Our goal is to estimate a parameter vector common to all groups, while controlling for omitted variables that are consistent with a particular group (Chamberlain 1980).

In linear models, a conservative approach utilizes the fixed effects within estimator to control for time invariant characteristics of individuals or groups. However, fixed effects estimation is problematic in non-linear specifications such as ours. In linear models, using the group mean as a sufficient statistic to estimate the fixed effect allows consistent estimation of the parameter vector holding the group effect constant. In non-linear models, since the density of the observed random dependent variable is assumed to be fully defined, maximum likelihood is the most appropriate estimation strategy. However, in this framework, since the parameter vector β is estimated as a function of the group or fixed effect, and since the variance of the group effect does not converge to zero, "the maximum likelihood estimator of β is a function of a random variable which does not converge to a constant as $N \to \infty$" (Greene 2001: 7), making maximum likelihood estimation problematic. There are also issues of small sample bias. Hsiao

(1996) demonstrates that in short panels and small samples, the bias created by fixed effects specifications of non-linear models can reach 100%. Even in larger samples, the bias persists if the panel is short (Greene 2001), and remains significant in longer panels and samples as large as 100 unique observations (Heckman 1981, Greene 2001). In addition, in the case of binary outcome variables, such as the likelihood of receiving information, fixed effect models ignore characteristics of observations that never achieve the observed outcome, focusing estimation toward a comparison of independent variables among observations that do achieve the outcome (Beck & Katz 2001). This not only creates downward bias in estimates of infrequent events, it prevents estimation of variables of interest that do not change during the observation period, which make up a significant portion of the variation of interest in our models. Finally, the computational difficulty of maximizing functions with a large numbers of fixed parameters makes non-linear fixed effects models difficult in practice.[12]

For these reasons, rather than employing fixed effects estimates, we address the group effect in our data by attempting to fully specify our models to control for the theoretically justifiable factors that could affect the flow of information. In doing so, we isolate estimates of our variables of interest holding constant most of the alternate explanations that may exist. We also cluster standard errors around individual recipients in order to maintain conservative estimates of the confidence intervals in light of the repeated observation of individuals. Although a host of hypotheses could be postulated about unobservable characteristics of individuals that make them more likely to receive information, our specifications include many of these possibilities. Given that we estimate many interesting independent variables that are constant (or nearly constant) over time and dyads, our conservative approach is likely to produce results robust to most alternate explanations of group effects without "throwing out the baby with the bath water" (Beck & Katz 2001).

[12] Although brute force methods are becoming more feasible (see Greene (2001) for a discussion).

Finally, we tested the performance implications of access to information diffusing through the network. We tested the relationship between measures of access to information (D_i) and productivity (P_i), controlling for traditional demographic and human capital factors (HC_{ji}) .

$$P_i = \gamma_i + \beta_1 D_i + \sum_j B_j HC_{ji} + \varepsilon_{it} \quad [3].$$

In these specifications productivity (P_i) is measured by projects completed and revenues generated during the period of email observation, and access to information (D_i) is measured by the number of words that were seen by the recruiter, the mean rank order in which they received words relative to their colleagues, the mean time it took for them to receive words, the number of words for which they were in the top 10% and the top 50% of recipients by time, and the number of words they saw in the first week and the first month. Human capital and demographic measures (HC_{ji}) include age, gender, education, industry experience, and organizational position (e.g. partner, consultant, and researcher). Each measure of access to information is entered separately into OLS regressions testing their association with project completions and revenues.

4. Results

Our results demonstrate that demography, organizational hierarchy, tie and network characteristics and functional task characteristics all significantly influence the diffusion of information at our research site. However, different factors impact the diffusion of different types of information differently. The diffusion of event news is influenced by demographic and network factors but not by functional relationships (e.g. prior co-work, authority) or the strength of ties. In contrast, diffusion of discussion topics is heavily influenced by functional relationships and the strength of ties, as well as demographic and network factors. Discussion topics are more likely to diffuse vertically up and down the organizational hierarchy, across relationships with a prior working history, and across stronger ties, while news is more likely to diffuse laterally as well as vertically, and without regard to the strength or function of relationships.

4.1. Estimation of the Diffusion of Information

We first tested the diffusion of all types of information through the firm. Table 3 presents the results of logistic regression and hazard rate model estimates of the likelihood of receiving information and the rate at which different types of information diffuse to different users. Model 1 presents the results of the logistic regression estimating factors that influence the likelihood of receiving information. Although employment at the firm is balanced along gender lines, and controlling for the potential effect of an uneven distribution of men and women in higher level positions in the firm (partner and consultant dummies), men are 55% more likely than women to receive information of all types. Demographic dissimilarity between the originator of the information and the eventual recipient reduces the likelihood of receiving information by between 1% and 13%, with gender differences recording the largest magnitude impact and age differences the smallest. Higher communication volume makes one more likely to see information, although we use this as a control variable rather than a variable of interest simply because we only measure information diffusion in email. The strength of ties between originator and recipient increases the likelihood of receiving information. Ten additional emails sent between originator and recipient increases the likelihood that a diffusion process started by the originator reaches the recipient by 2%. Path length reduces the likelihood of receiving information, with each additional hop reducing the average likelihood of being involved diffusion by 29%. Having friends in common with the first user of a word seems to reduce the likelihood of receiving an information cascade that originates from that user – a strange result. However, having friends in common is positively correlated with email volume and the strength of ties. Holding these other variables constant, the positive effects of friends in common reduce and reverse. Betweeness centrality has a strong positive effect on the likelihood of receiving information, as do stronger project co-work relationships.

Table 3. Drivers of Access to Information

	Model 1	**Model 2**
Dependent Variable:	**Word Received**	**Rate of Receipt**
Specification (Coefficient Reported)	*Logistic (Odds Ratio)*	*Hazard Model (Hazard Ratio)*
Demography[1]		
Gender Dummy (Male = 1)	1.551 (.219)***	1.236 (.167)
Demographic Distance		
Age Difference	.986 (.004)***	.996 (.004)
Gender Difference	.869 (.014)***	1.009 (.010)
Education Difference	.906 (.023)***	.971 (.020)
Tie & Network Characteristics		
Communication Volume (Total Email)	1.0002 (.0002)**	1.000 (.000)
Strength of Tie	1.002 (.001)***	1.000 (.000)
Path Length	.711 (.047)***	.828 (.033)***
Geographic Proximity (Same Office = 1)	.857 (.088)	.865 (.078)
Friends in Common	.954 (.007)***	.992 (.005)
Betweeness Centrality	1.005 (.002)**	1.004 (.002)**
Constraint	.212 (.225)	.326 (.389)
Task Characteristics		
Prior Project Co-Work	1.042 (.016)***	1.031 (.012)**
Industry Tenure Difference	.996 (.006)	1.002 (.006)
Same Area Specialty	.883 (.080)	.983 (.067)
Managerial Level Difference	.951 (.038)	.997 (.033)
Partner Dummy	.933 (.188)	1.062 (.168)
Consultant Dummy	.870 (.184)	1.118 (.207)
Word Type		
Common Information	3.209 (.056)***	2.292 (.065)***
Discussion Topics	.081 (.008)***	.025 (.002)***
Log Pseudolikelihood	-234204.48	-1694852.4
Wald χ^2 (d.f.)	6264.80 (19)***	8878.76 (19)***
Pseudo R^2	.28	-
Observations	543308	462422
Notes: 1. Age, Edu, Industry Tenure N.S.		

The hazard rate model estimates of the drivers of the rate of information receipt reveal positive effects for project co-work and betweeness centrality, and a negative relationship between path length and the rate at which information is received.[13] These results demonstrate the importance of demographic distance, network structure and project based working relationships on the likelihood of receiving information and the rate at which it is received.

4.2. Estimation of the Diffusion of Discussion Topics & Event News

[13] The dummy variables for word type, which control for common information, news, and discussion topics show that common information is more pervasive and appears at a faster rate among employees than news (the omitted category), while discussion topics are much less likely to be seen and diffuse at a much slower rate.

Table 4 presents the results of logistic regression and hazard rate model estimates of the drivers of event news and discussion topic diffusion. Models 1 and 2 report the results of logistic regressions. The results demonstrate that demographic distance reduces the likelihood of receiving both news and discussion topics although with a slightly larger impact for news. The coefficient on the education difference parameter for instance indicates that one additional year of education difference between two individuals reduces the likelihood that news will defuse between them by 7.5%, while the same one year difference in education reduces the likelihood of discussion topics diffusing between them by nearly 17%. Interestingly, men are over 50% more likely to see news information than women although gender has no effect on the likelihood of the diffusion of discussion topics. Overall demographic distance slows information diffusion.

Coefficients of tie and network characteristics also tell an interesting story. Strong ties are important predictors of the diffusion of discussion topics but not of news. News seems to diffuse pervasively throughout the organization without regard to the strength of ties – information of general interest is passed through relatively weak ties as well. The parameter estimate of the strength of ties on discussion diffusion in Model 2 indicate that ten additional emails exchanged between two people increases the likelihood that discussion topics will diffuse between them by 7% on average. Path length, which measures the number of nodes separating employees in the email network, reduces the likelihood of information diffusion, although the impact is much larger for discussion topics than for news. An additional hop between individuals reduces the likelihood of information diffusion by 97%, indicating discussion topics almost always only diffuse to people who know each other directly, whereas news may travel across multiple hops to reach a receiver. Geographic proximity has no effect on diffusion although the parameter estimates are less than one, supporting the proposition that exchanges of information over email are less likely for co-located than geographically distant employees in our firm. Having friends in common with the first user of a word again reduces the likelihood of receiving information that originates from that user. Betweenness centrality is also positively associated with the likelihood of seeing both news and discussion topics, while constraint is not significant.

Table 4. Drivers of Access to Discussion Topics & Event News

	NEWS	DISCUSSION	NEWS	DISCUSSION
	Model 1	**Model 2**	**Model 3**	**Model 4**
Dependent Variable:	**Word Received**	**Word Received**	**Rate of Receipt**	**Rate of Receipt**
Specification (Coefficient)	*Logistic (Odds Ratio)*	*Logistic (Odds Ratio)*	*Hazard Model (Hazard Ratio)*	*Hazard Model (Hazard Ratio)*
Demography[1]				
Gender (Male=1)	1.544 (.227)***	1.073 (.137)	1.332 (.228)*	1.075 (.162)
Demographic Distance				
Age Difference	.992 (.004)**	.981 (.007)***	.998 (.004)	.994 (.007)
Gender Difference	.902 (.017)***	.814 (.069)**	1.007 (.012)	1.092 (.110)
Education Difference	.925 (.022)***	.832 (.034)***	.966 (.024)	1.013 (.037)
Tie & Network Characteristics				
Email Volume	1.0001 (.00007)*	1.0001 (.0001)*	1.0001 (.000)	1.0001 (.000)**
Strength of Tie	1.000 (.000)	1.007 (.001)***	.999 (.000)	1.006 (.001)***
Path Length	.732 (.041)***	.029 (.005)***	.814 (.044)***	.310 (.045)***
Geographic Proximity	.883 (.090)	.929 (.106)	.879 (.097)	.993 (.115)
Friends in Common	.972 (.005)***	.877 (.012)***	.992 (.007)	.969 (.012)**
Betweeness Centrality	1.004 (.002)*	1.007 (.002)**	1.006 (.002)**	1.002 (.002)
Constraint	.186 (.213)	2.243 (2.651)	.282 (.410)	1.664 (1.698)
Task Characteristics				
Prior Project Co-Work	1.010 (.014)	1.080 (.0185)***	1.018 (.016)	1.066 (.018)***
Industry Tenure Difference	.996 (.006)	.978 (.008)**	.999 (.008)	.999 (.008)
Same Area Specialty	.933 (.073)	1.038 (.139)	.981 (.078)	1.795 (.252)***
Organizational Hierarchy				
Managerial Level Difference	.963 (.035)	1.138 (.079)*	.992 (.037)	1.097 (.089)
Partner Dummy	.856 (.186)	1.515 (.271)**	1.084 (.216)	1.411 (.232)**
Consultant Dummy	.798 (.177)	1.659 (.262)***	1.221 (.289)	1.749 (.288)***
Log Pseudolikelihood	-93273.148	-15167.79	-508288.77	-28166.432
Wald χ^2 (d.f.)	204.39 (17) ***	2816.61 (17)***	92.80 (17)***	762.33 (17)***
Pseudo R^2	.06	.54	-	-
Observations	163135	202500	120197	196541

Notes: [1] We controlled for age and education, but these variables were never significant and did not contribute to the explanatory power of the model. Geographic Proximity: Same Office = 1; * p < .05; ** p < .01; *** p < .001.

Perhaps most interesting are the effects of task characteristics and organizational hierarchy. Strong working relationships and similarity in industry tenure both have strong positive impacts on the likelihood of receiving discussion topics, but not on the diffusion of news. Each additional project that two people work on together increases the likelihood that discussion diffuses between them by 8%. The coefficient on managerial level difference indicates that discussion topics are more likely to diffuse vertically, up and down the organizational hierarchy rather than laterally across individuals of the same organizational rank. In our firm, since project teams and work flow are organized vertically, with one member from each organizational level represented on most teams and no teams of either three partners, three consultants or three researchers, discussion topics, which we propose represent more specific, procedural and complex information, typically move between organizational levels rather than across them. The coefficients on the partner and consultant dummy add more clarity to this result. As researcher is the omitted category, these strong positive estimates demonstrate that discussion topics are more likely to be seen by those higher in the organizational hierarchy, indicating that discussion is more likely to diffuse up than down the hierarchical structure of the firm. Our interviews revealed that consultants were responsible for taking care of most day to day work at the firm which is reflected in the fact that the parameter estimate for 'consultant' is larger and more significant than for partners. This result also provides some evidence for the argument that mid-level managers 'have their ear to the ground' and are possibly the most aware of informal information circulating in the firm.

Hazard rate analyses mirror the logistic regression results to a large extent. Men see news at a higher rate than women, although demographic differences do not seem to predict the rate at which individuals see either news or discussion topics. The strength of ties again has a strong positive impact on the hazard rate for discussion topics but has no effect for news, while greater path lengths consistently reduce the hazard rate across both types of information. We again see increases in the rate at which employees see discussion news with greater project co-work (6.6% increase per additional project), but this time we see that having the same area of expertise increases the rate while industry tenure differences have no effect. Although managerial level differences are insignificant, the partner and consult dummy variables show that employees in the top two levels of the organization see information at a higher rate than researchers, while consultants again seem to be the most well informed.

4.2. Access to Information and Productivity

Table 5 presents the results of our estimates of the impact of access to information on the productivity of individual recruiters as measured by the number of projects completed. Each measure of access to information captures a particular dimension of the degree to which recruiters are privy to information diffusing through the email network. 'Words seen' is a count of the number of words each recruiter received in email. 'Mean rank' measures the rank order of receipt for each word relative to other recruiters. If a recruiter is the first person to receive a word they are ranked 1, if they are the fifth person to receive a word they are ranked 5 and so on. The mean rank is simply the recruiter's average rank across all words. 'Mean time' measures the average time in days it takes recruiters to see words. 'Rank 10% (50%)' measures the number of words for which recruiters were in the first 10% (50%) of employees to see the word. 'Words seen in one week (month)' measures how many words the recruiter sees within one week (month).

Table 5. Information Diffusion & Project Completions

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Dependent Variable:	Number of Completed Projects	Number of Completed Projects	Number of Completed Projects	Number of Completed Projects	Number of Completed Projects	Number of Completed Projects	Number of Completed Projects
Age	.015 (.066)	.010 (.063)	.006 (.063)	.027 (.068)	.021 (.067)	.054 (.060)	.201 (.065)
Gender	-1.115 (.699)	-1.119* (.632)	-1.176* (.634)	-1.367* (.789)	-1.133 (.712)	-1.141 (.770)	-1.349* (.782)
Education	.066 (.320)	.162 (.289)	.153 (.296)	-.011 (.319)	.068 (.321)	.039 (.303)	-.002 (.318)
Industry Experience	-.029 (.061)	-.012 (.059)	-.009 (.060)	-.016 (.057)	-.026 (.061)	-.032 (.053)	-.021 (.059)
Partner	1.335 (1.627)	1.508 (1.530)	1.596 (1.536)	2.491 (1.912)	1.397 (1.680)	2.456 (1.839)	2.361 (1.816)
Consultant	1.592 (1.079)	1.832* (.952)	1.857* (.962)	2.479 (1.583)	1.660 (1.151)	2.417 (1.545)	2.198 (1.473)
Words Seen	.001*** (.0003)						
Mean Rank		-.225*** (.041)					
Mean Time			-.132*** (.023)				
Rank 10%				.004*** (.001)			
Rank 50%					.002*** (.0003)		
Words Seen In 1 Week						.008*** (.002)	
Words Seen In 1 Month							.003*** (.001)
Constant	-1.597 (5.674)	13.858** (5.179)	17.268*** (5.369)	-.069 (6.109)	-1.349 (5.768)	-2.464 (6.171)	-.446 (5.998)
F-Value (d.f.)	5.13*** (7)	6.73*** (7)	7.07*** (7)	2.94** (7)	4.28*** (7)	3.16** (7)	3.37*** (7)
R^2	.39	.43	.44	.25	.36	.27	.29
Obs.	41	41	41	41	41	41	41

Note: * p < .10; ** p < .05; *** p < .01

The results demonstrate that access to information predicts project output. Each additional ten words seen are associated with an additional 1% of one project completed. We also see that greater mean rank and longer average times to receive words are associated with fewer projects completed holding constant traditional demographic and human capital variables.

Table 6 presents relationships between access to information diffusion and revenues generated, which can be thought of as a quality adjusted measure of output.

Table 6. Information Diffusion & Revenues

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Dependent Variable:	Total Revenues	Total Revenues	Total Revenues	Total Revenues	Total Revenues	Total Revenues	Total Revenues
Age	1127.36 (2821.64)	888.59 (2684.60)	720.11 (2676.23)	1812.38 (3079.27)	1414.45 (2884.71)	2846.80 (2820.33)	1525.45 (2935.73)
Gender	-65152.48* (36796.11)	-65387.82* (34113.54)	-67740.8* (34320.92)	-70968.47* (41507.65)	-65451.9* (37780.24)	-64268.99 (41860.22)	-71504.52* (41663.96)
Education	-3340.51 (13410.84)	1052.76 (12231.44)	453.83 (12489.27)	-9337.38 (13878.34)	-3653.47 (13658.48)	-6093.10 (14103.16)	-8428.63 (13741.45)
Industry Experience	-2517.68 (2771.90)	-1744.85 (2755.45)	-1599.66 (2766.55)	-2061.68 (2749.58)	-2365.72 (2789.73)	-2648.24 (2630.66)	-2222.38 (2771.76)
Partner	121600.4 (77138.46)	129394.5* (70803.72)	133607.9* (71045.74)	171580.1* (96160.11)	125243.7 (81137.58)	171220.7* (91832.3)	167003.2* (90702.31)
Consultant	61777.68 (61463.41)	72674.13 (55064.94)	73515.88 (55743.18)	91727.37 (87988.3)	64306.19 (66203.77)	93837.54 (84155.48)	82969.67 (82146.35)
Words Seen	70.52*** (15.61)						
Mean Rank		-10202.88*** (1992.77)					
Mean Time			-5931.05*** (1130.32)				
Rank < 10%				152.07** (58.76)			
Rank < 50%					64.93*** (16.16)		
Words Seen In 1 Week						321.50*** (114.98)	
Words Seen In 1 Month							114.96*** (38.76)
Constant	64973.45 (247744.40)	765031.8*** (22344.2)	915736.5*** (231192.6)	195308.1 (276691.3)	85886.32 (255321.6)	68776.88 (290924.2)	166804.6 (272595.9)
F-Value (d.f.)	4.46*** (7)	5.39*** (7)	5.54*** (7)	2.64** (7)	3.77*** (7)	3.56*** (7)	2.83** (7)
R^2	.39	.42	.42	.24	.36	.27	.27
Obs.	41	41	41	41	41	41	41

These results demonstrate economically significant relationships between access to information diffusing the network and output. An additional ‘word seen’ is associated with about $70 of additional revenue generated. Strikingly, access to information diffusing in the network is a much stronger predictor of productivity than traditional human capital variables such as education or industry experience.

5. Discussion & Conclusion

We develop theory on how different types of information diffuse through social groups in organizations. In order to build our understanding of organizational information dynamics, we examine factors that affect the diffusion of two different types of information – event news and discussion topics – and their effects on individual performance. We empirically test our hypotheses by observing several thousand diffusion processes of each type of information in real email communication data collected over ten months in a mid-sized executive recruiting firm combined with detailed accounting data at the level of individual information workers. Our results reveal that the confluence of social structure and information content together determine the movement of information in our firm. In turn, better access to information strongly predicts the performance of these workers.

We demonstrate that demographic distance plays a significant role in reducing the diffusion of information across all information types, and that traditional social network characteristics such as the strength of ties and the path length between individuals increase and decrease the likelihood of receiving information respectively. We find that prior project co-work increases the likelihood of receiving information, demonstrating the influence of prior work history on the diffusion patterns of information in organizations. We also find gender to be a significant predictor of the flow of information in our firm, with men being 55% more likely than women to receive a given piece of information during its diffusion process.

We also find that different types of information diffuse differently. Upon examining the diffusion of event news and discussion topics we find striking differences in the factors associated with their movement. While demographic distance reduces the likelihood of seeing both types of information, task characteristics such as project co-work and industry tenure differences are more likely to reduce the likelihood of receiving discussion information than event news. Discussion topics are more likely to diffuse vertically up and down the organizational hierarchy, across relationships with a prior working history, and across stronger ties, while news is more likely to diffuse laterally as well as vertically, and without regard to the strength or function of relationships.

Furthermore, these differences in diffusion patterns strongly correlate with performance. Information workers who receive more novel information, or who receive it earlier than their colleagues complete projects at a statistically faster rate than their colleagues and generate significantly more revenue for the firm.

These findings highlight the importance of considering both structure and content in information diffusion. The confluence of structural factors related to demographic, organizational and social relationships, and the content of the information diffusing through a social group determines its diffusion path and the likelihood of its diffusion to given individuals. The theory and results developed shed light on who is likely to become aware of a given piece of strategic information in an organization and who is likely to become aware of it sooner. They also provide some of the first evidence quantifying the role of information diffusion on the productivity of information workers. The theory and evidence developed here addresses some of the fundamental assumptions underlying a variety of research streams from the diffusion of innovations, to dynamic trading behavior, to word of mouth viral marketing.

References

Adamic, L., Rajan, L., Puniyani, A. & Huberman, B. 2001. "Search in power law networks." *Physical Review E.* 64, 046135.

Ahuja, G. 2000. "Collaboration networks, structural holes and innovation: A longitudinal study." *Administrative Science Quarterly*, 45: 425-455.

Allen, T. J. 1977. Managing the flow of technology. Cambridge, MA, MIT Press.

Ancona, D.G. & Caldwell, D.F. 1992. Demography & Design: Predictors of new Product Team Performance. *Organization Science*, 3(3): 321-341.

Aral, S., Brynjolfsson, E., & Van Alstyne, M. 2006. "Information, Technology and Information Worker Productivity: Task Level Evidence." *Proceedings of the 27th Annual International Conference on Information Systems*, Milwaukee, Wisconsin.

Beck, N., & J. Katz. 2001. "Throwing out the baby with the bath water: A comment on Green, Kim and Yoon." *International Organization*, 55(2): 487-495.

Bernard, H.R., Killworth, P., & Sailor, L. 1981. "Summary of research on informant accuracy in network data and the reverse small world problem." *Connections*, (4:2): 11-25.

Burt, R. 1987. "Social contagion and innovation: Cohesion versus structural equivalence." *American Journal of Sociology*, 92(6): 1287-1335.

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

Burt, R. 2000. "The network structure of social capital" In B. Staw, & Sutton, R. (Ed.), *Research in organizational behavior* (Vol. 22). New York, NY, JAI Press.

Burt, R. 2002. "Bridge decay." *Social Networks*, 24(4): 333-363.

Burt, R. 2004. "Structural holes & good ideas." *American Journal of Sociology*, (110): 349-99.

Buskens, V. & K. Yamaguchi. 1999. "A new model for information diffusion in heterogeneous social networks." *Sociological Methodology*, 29: 281-325.

Chaimberlin, G. 1980. "Analysis of covariance with qualitative data." *Review of Economic Studies*, 47(1): 225-238.

Cohen, W. & D. Levinthal. 1990. "Absorptive Capacity: A new perspective on learning and innovation." *Administrative Science Quarterly*, 35: 128-152.

Cohen, W. & P. Bacdayan. 1994. "Organizational routines are stored as procedural memory: Evidence from a laboratory study." *Organization Science*, 5(4): 554-568.

Coleman, J.S. 1988. "Social Capital in the Creation of Human Capital" *American Journal of Sociology*, (94): S95-S120.

Coleman, J., Katz, E., & H. Menzel. 1966. Medical innovation: A diffusion study, Bobbs-Merrill, New York.

Cook, K.S., Emerson, R.M., Gilmore, M.R., & Yamagishi, T. 1983. "The distribution of power in exchange networks." *American Journal of Sociology*, 89: 275-305.

Cramton, C.D. 2001. "The mutual knowledge problem and its consequences for dispersed collaboration." *Organization Science*, 12(3): 346-371.

Cummings, J., & Cross, R. 2003. "Structural properties of work groups and their consequences for performance." *Social Networks*, 25(3):197-210.

Cummings, J. 2004. "Work groups, structural diversity, and knowledge sharing in a global organization." *Management Science*, 50(3): 352-364.

De Vany, A. & C. Lee. 2001. "Quality signals in information cascades and the dynamics of the distribution of motion picture box office revenues." *Journal of Economic Dynamics and Control*, 25(3-4): 593-614.

Dellarocas, C. 2003. "The digitization of word of mouth: Promise and challenges of online feedback mechanisms." *Management Science* 49(10): 1407-1424.

Dessein, W. 2002. "Authority and communication in organizations." *Review of Economic Studies*, 69(4): 811-838.

Dewar, R. & J. Dutton. 1986. "The adoption of radical and incremental innovations: An empirical analysis." *Management Science* 32(11): 1422-1433.

Domingos, P., & M. Richardson. 2001. "Mining the network value of customers" *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA: 57-66.

Emerson, R. 1962. "Power-Dependence Relations." *American Sociological Review*, 27: 31-41.

Freeman, L. 1979. "Centrality in social networks: Conceptual clarification." *Social Networks*, 1(3): 215-234.

Gargiulo, M., M. Benassi, 2000. "Trapped in your own net? Network cohesion, structural holes, and the adaptation of social capital." *Organization Science* (11:2): 183-196.

Greve, H. Tuma, D. & N. Strang. 2001. "Estimation of diffusion models from incomplete data." *Sociological Methods & Research*, 29: 435.

Granovetter, M. 1973. "The strength of weak ties." *American Journal of Sociology* (78):1360-80.

Granovetter, M. 1978. "Threshold models of collective behavior." *American Journal of Sociology* 83(6):1420-1443.

Green, D., Kim, S.Y., & D. Yoon. 2001. "Dirty pool." *International Organization*, 55(2): 441-468.

Greene, W. 2001. "Fixed and random effects in nonlinear models." *NYU Department of Economics Working Paper.* New York, NY.

Grilliches, Z. 1957. "Hybrid Corn: An exploration of the economics of technical change." *Economtrica* 25(4): 501-522.

Gruhl, et al., 2004. "Information diffusion through blogspace", in *Proceedings of the 13th international conference on World Wide Web*. New York, NY.

Hansen, M. 1999. "The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits." *Administrative Science Quarterly* (44:1):82-111.

Hansen, M. 2002. "Knowledge networks: Explaining effective knowledge sharing in multiunit companies." *Organization Science* (13:3): 232-248.

Hargadon, A. & R, Sutton. 1997. "Technology brokering and innovation in a product development firm." *Administrative Science Quarterly*, (42): 716-49.

Hausman, J., Hall, B., & Z., Grilliches. 1984. "Econometric models for count data with an application to the patents-R&D relationship." *Econometrica*, 52(4): 909-938.

Heckman, J. 1981. "The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process." In Manski, C. & D. McFadden (eds) Structural Estimation of Discrete Data with Econometric Applications, MIT Press, Cambridge, MA, 114-178.

Hirshleifer, D., Subrahmanyam, A., & T., Sheridan. 1994. "Security Analysis and Trading Patterns when Some Investors Receive Information Before Others" *Journal of Finance*, 49(5): 1665-1698.

Hsiao, C. 1996. "Logit & probit models." In Matyas, L. & P. Severstre (eds), The econometrics of panel data: Handbook of theory and applications, Kulwer Academic Publishers, Dordrecht.

Kempe,D., Kleinberg, J., & E. Tardos. 2003. "Maximizing the spread of influence through a social network" *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, D.C.: 137-146.

Kleinberg, J. 1999. "The small-world phenomenon: An algorithmic perspective." Cornell Computer Science Technical Report 99-1776 (October).

Kleinberg, J. 2001. "The small-world phenomenon and the dynamics of information." Advances in Neural Information Processing Systems (NIPS) 14.

Krackhardt, D. & Kilduff, M. 1999. "Whether close or far: Social distance effects on perceived balance in friendship networks." *Journal of personality and social psychology* (76) 770-82.

Kumbasar, E., Romney, A.K., and Batchelder, W.H. 1994. Systematic biases in social perception. *American Journal of Sociology*, (100): 477-505.

Lescovec, J., Singh, A., & J. Kleinberg. 2006. "Patterns of influence in a recommendation network." *Pacific-Asia Conference on Knowledge Discovery & Data Mining (PAKDD).*

Marschak, J. & Radner, R. 1972. Economic Theory of Teams, Yale University Press, CT.

Marsden, P. 1990. "Network Data & Measurement." *Annual Review of Sociology* (16): 435-463.

McPherson, M., L. Smith-Lovin & J. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27: 415-444.

Meyer, J., & B. Rowan. 1977. "Institutionalized organizations: Formal structure as myth and ceremony." *American Journal of Sociology*, 83(2): 340-363.

Mintzberg, H. 1979. The Structuring of Organizations, Prentice-Hall, Englewood Cliffs, NJ.

Nelson, R. "What is public and what is private about technology?" Working Paper 9-90, Center for Research in Management, University of California Berkeley.

Newman, M., Forrest, S. & J. Balthrop. 2002 "Email networks and the spread of a computer virus." *Physical Review E.*, 66, 035101.

Orlikowski, W.J. 2002. "Knowing in practice: Enacting a collective capability in distributed organizing." *Organization Science*, 13(3): 249-273.

Petersen, T. 1991. "Time aggregation bias in continuous-time hazard-rate models." In Sociological Methodology, edited by Peter Marsden, Basil-Blackwell, Cambridge, MA: 263-290.

Petersen, T. & K. Koput. 1992. "Time aggregation bias in hazard-rate models with covariates." *Sociological Methods & Research*, 21:25-51.

Pfeffer, J. 1983. "Organizational Demography," in Larry L. Cummings and Barry M. Staw (eds.), Research in Organizational Behavior, 5: 299-257. JAI Press, Greenwich, CT.

Podolny, J., Baron, J. 1997. "Resources and relationships: Social networks and mobility in the workplace." *American Sociological Review* (62:5): 673-693.

Reagans, R. & McEvily, B. 2003. "Network Structure & Knowledge Transfer: The Effects of Cohesion & Range." *Administrative Science Quarterly*, (48): 240-67.

Reagans, R. & Zuckerman, E. 2001. "Networks, diversity, and productivity: The social capital of corporate R&D teams." *Organization Science* (12:4): 502-517.

Reagans, R. & Zuckerman, E. 2006. "Why Knowledge Does Not Equal Power: The Network Redundancy Tradeoff" *Working Paper Sloan School of Management* 2006, pp. 1-67.

Rodgers, E. 1995. *The Diffusion of Innovations*. The Free Press, New York.

Rosenberg, N. 1982. *Inside the black box: Technology and economics.* Cambridge University Press, New York.

Ryan, B. & N.C. Gross 1943. "The diffusion of hybrid seed corn in two Iowa communities." *Rural Sociology* 8: 15-24.

Salton, G., Wong, A., & Yang, C. S. 1975. “A Vector Space Model for Automatic Indexing.” *Communications of the ACM*, 18(11): 613-620.

Schelling, T.C. *Micromotives & Macrobehavior*. George J. McLeod Ltd. Toronto.

Sparrowe, R., Liden, R., Wayne, S., & Kraimer, M. 2001. “Social networks and the performance of individuals and groups.” *Academy of Management Journal*, 44(2): 316-325.

Strang, D. & N. Tuma. 1993. “Spatial and temporal heterogeneity in diffusion” *American Journal of Sociology*, 99(3): 614.

Szulanski, G. 1996. "Exploring internal stickiness: Impediments to the transfer of best practice within the firm." *Strategic Management Journal* (17): 27-43.

Tuma, N.B., & Hannan, M.T.1984. Social Dynamics: Models and Methods. Academic Press, New York.

Uzzi, B. 1996. “The sources and consequences of embeddedness for the economic performance of organizations: The network effect” *American Sociological Review*, (61):674-98.

Uzzi, B. 1997. “Social structure and competition in interfirm networks: The paradox of embeddedness.” *Administrative Science Quarterly*, 42: 35-67.

Van Alstyne, M. & Zhang, J. 2003. “EmailNet: A system for automatically mining social networks from organizational email communication,” NAACSOS.

Von Hippel, E. 1998. "Economics of Product Development by Users: The Impact of "Sticky" Local Information" *Management Science* (44:5): 629-644.

Watts, D. 1999. “Networks, dynamics and the small-world phenomenon.” *American Journal of Sociology*, 105(2): 493-527.

Watts, D. & S. Strogatz. 1998. “Collective dynamics of small-world networks.” *Nature*, 393: 440-442.

Wejnert, B. 2002. “Integrating models of diffusion of innovations: A conceptual framework.” *Annual Review of Sociology*, 28: 297-227.

White, H. 1980. “A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity." *Econometrica* (48:4): 817-838.

Wu, F. Huberman, B., Adamic, L., & J. Tyler. 2004. “Information flow in social groups.” *Physica A: Statistical and Theoretical Physics*, 337(1-2): 327-335.

Yamaguchi, K. 1994. “The flow of information through social networks: Diagonal-free measures of inefficiency and structural determinants of inefficiency.” *Social Networks*, 16: 57-86.