Productivity Effects of Information Diffusion in Networks

Sinan Aral
NYU Stern School of Business & MIT Sloan School of Management, 44 West 4th St. 8-81, New York, NY, 10012, sinan@stern.nyu.edu.

Erik Brynjolfsson
MIT Sloan School of Management, E53-313, Cambridge, MA 02142, ebrynjo@mit.edu.

Marshall Van Alstyne
Boston University School of Management & MIT Sloan School of Management, 595 Commonwealth Avenue, Boston, MA 02215, marshall@mit.edu.

We examine what drives the diffusion of different types of information through email networks and the effects of these diffusion patterns on the productivity and performance of information workers. 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 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. 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. We find 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 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. We also find access to information strongly predicts project completion and revenue generation. The effects are economically significant, with each additional "word seen" correlated with about $70 of additional revenue generated. Our findings provide some of the first evidence of the economic significance of information diffusion in email networks.

Key words: Networks, Information Diffusion, Productivity, Email.
Acknowledgements: The authors are grateful for the support of Cisco Systems for this research.

Preliminary Draft Version - May 18, 2007

1. Introduction

The process of information diffusion lies at the heart of numerous phenomena in 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.[1] Furthermore, a foundational assumption of social network theory is that strong and weak ties as well as structural holes affect performance because of their influence on information flows. While theories based on information diffusion and social network theory proliferate, empirical evidence on how information spreads within firms and the ultimate economic effects remain scarce.

In large part, this shortfall reflects the difficulty of directly measuring information diffusion in a fine-grained way.[2] Prior diffusion studies typically observed adoption or purchase decisions rather than the movement of information, and studies that do focus on information per se are typically theoretical or simulation based. Existing theory focuses mainly on which global social structures maximize diffusion, and although we know that certain types of information transfer more easily than others (Von Hippel 1998), diffusion studies typically treat information as homogenous, overlooking variation in diffusion based on differences in type. This gap in current research 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 encounter an idea as it spreads? Do different types of information diffuse differently? Can we explicitly link having novel information to better performance?

In this paper, we gather and analyze data that are uniquely well-suited to studying the movement of different types of information through an organization. Using ten months of email data, and five years

[1] Timely access to strategic information, innovative ideas, or current news can also highlight hidden opportunities, provide negotiating leverage (Burt 1992), promote innovation (Burt 2004), and ultimately drive economic performance (Reagans & Zuckerman 2001, Hansen 2002).

[2] Indeed, developing the data for this paper required several thousand hours by a team of half a dozen researchers across multiple institutions.

of revenue data and project co-work relationships, we characterize the social networks of individuals in a mid-sized executive recruiting firm. We identify two types of information diffusing through this network – 'event news' and 'discussion topics' – by their usage characteristics, and observe several thousand diffusion processes of each type. 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 sooner.

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 organizational hierarchy, across relationships with a prior working history, and across stronger ties, while news will diffuse both laterally and vertically, and without regard to the strength or function of relationships. These findings highlight the importance of simultaneously studying structure and content to understand information diffusion. In particular, we argue that the type of information and the types of social relations jointly predict the diffusion path of novel information. While some types of information diffuse more vertically through organizational hierarchy and across functional relationships, other types diffuse laterally and without regard to function or hierarchy.

Strikingly, we find that access to information strongly predicts employees' productivity. Timely access to novel information predicts the number of projects completed by each individual and the amount of revenue each person generates holding other factors constant. Access to novel information is economically significant. Beyond the average, each additional novel word seen predicts roughly $70 of additional revenue; and productivity falls as news is delayed. Our findings provide some of the first evidence of the economic significance of information diffusion in email networks. Indeed, they explicitly validate the long-hypothesized mechanism by which social network structure might influence performance, namely via information diffusion.

2. Theory & Literature

2.1. The Central Role of Information in Diffusion Studies

Theories of the diffusion of innovations (e.g. Rogers 1995) 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 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 traders create abnormal profits for the informed and explain seemingly irrational trading equilibria, such as "herding" or outcomes based on "follow the leader" strategies. Yet, in these models temporal asymmetries in information acquisition are taken as given, and how and why these systematic asymmetries arise remains unknown.

Much of current literature on information diffusion and contagion 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." Two core models have emerged to explain the diffusion of influence in and contagion. Threshold models posit that individuals adopt innovations after reaching and surpassing their own private "threshold" of influence (e.g. Granovetter 1978, Schelling 1978). Cascade models 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). While both models assume an information transmission between adopters and non-adopters, they 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*." [emphasis added].

Finally, there is a small body of literature on knowledge transfers and performance (e.g. Reagans & Zuckerman 2001). 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), although this work focuses on dyadic transfers of information rather than on the diffusion paths of information through a population.

2.1. Information Diffusion in Organizations

Although some information diffusion studies exist, they typically rely on computer simulations of a handful of agents (e.g. Buskens & Yamaguchi 1999, Newman et. al. 2002, Reagans & Zuckerman 2006), treat information as uniform and homogeneous (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). A current focus on global network properties that maximize information diffusion (e.g. Watts & Stogatz 1998) deemphasizes predictors of access to information cascades and their economic consequences. In addition, assumptions of information homogeneity are problematic in light of 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 (Nelson 1990), complexity (Uzzi 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 argue that several other important factors influence information diffusion 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 affect the movement of information, and have different effects across different types of information.

We therefore propose four extensions to current work. First, in addition to global network structures, there exist hierarchal, demographic and task based drivers of information diffusion. For example, 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. Third, we argue that content and structure jointly predict the diffusion path information - that different social and structural factors will govern the diffusion of different types of information. Finally, and perhaps most importantly, we explicitly link performance differences among individuals to the information diffusion, and by implication, the determinants of information diffusion including social network structure.

2.2. Social Drivers of Information Diffusion

We hypothesize four categories of factors that may impact information dynamics in organizations:

1. *Demograpics*. Individuals' personal characteristics and dissimilarity are likely to affect social choices about information seeking and transmission. Similar individuals tend to flock together

(McPherson, Smith-Loving, & Cook 2001), which creates parity in perspectives and information among similar individuals in organizations (Burt 1992, Reagans & Zuckerman 2001). Demographic diversity can also create social divisions (Pfeffer 1983), reducing the likelihood that individuals will seek each other for advice or information sharing. We therefore measure the demographic characteristics of individuals and the demographic dissimilarity of pairs of individuals focusing on age, gender, and education, three of the most important variables in organizational demography.[3]

2. *Organizational Hierarchy.* 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. As project teams in our organization are organized hierarchically, task related information is likely to flow vertically rather than laterally across an organizational level. We therefore measure each individual's position in the organizational hierarchy (e.g. partner, consultant, and researcher).

3. *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 performance (e.g. Burt 1992, 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). 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

[3] We do not have access to race or organizational tenure variables (although we do measure industry tenure).

[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.

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] (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:

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

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

Finally, a great deal of evidence links physical proximity to communication between actors (e.g. Allen 1977). In the case of email, geographic distance may be associated with more email communication between actors who find it costly to communicate face to face. We therefore measure *physical proximity* by whether two people work in the same office.

4. *Functional Task Characteristics*. Working relationships are conduits of information flow. They necessitate exchanges of task related information and create relatively stable ties that individuals rely on for advice on future projects. However, relationships can decay over time (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. We expect information to diffuse more easily between employees with the same industry tenure, who have been through similar work related milestones and may already be familiar with one another through industry relationships (Pfeffer 1983). Status and authority differences also may prevent less experienced workers from soliciting or sharing

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

information across industry tenure gradients while more experienced workers, less constrained by status and authority differences, may rely on other experienced workers for information.

2.3. Dimensions of Information Content

Characteristics of information content are also likely to affect diffusion patterns. Certain types of information are "stickier" and have higher transfer costs (Von Hippel 1998). We describe two contrasting information types, 'event news' and 'discussion topics,' which serve as archetypes for comparison.[6]

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 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, specialized groups. At this firm, work groups discuss particular projects, and most frequently have back and forth discussion about particular candidates or clients. 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.

Theories of information transfer support our distinctions between event news and discussion topics. Complex knowledge is more difficult and costly to transfer requiring strong dyadic ties for

[6] Archetypes are not mutually exclusive but rather serve to evoke underlying properties that correlate with diffusion patterns and usage behaviors of particular words in email. Our contention is that information of the types described is likely to diffuse according to specific patterns, and that patterns and word character proxy one another. The precise mapping is not critical. Our goal is to demonstrate that different characteristics of people, job relationships, and social structure affect access to information.

effective transfers (Hansen 1999). A theoretical distinction is also 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. 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).[7] 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 due to their shared common interests and their ability to more effectively communicate ideas based on their "common ground" (Cramton 1991). We therefore hypothesize diffusion of event news will be driven by demographic and network factors that 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 betweenness centrality, constraint and path length.*

On the other hand, information passed back and forth among small groups is likely to be task specific and relevant to those socially proximate to the originator. At our research site, work groups are organized vertically, with teams typically composed of members from different organizational levels, implying task related information passes vertically up and down the organizational hierarchy. We hypothesize that diffusion of discussion topics, is driven not only by demographic and network factors, but also by project co-work and organizational hierarchy.

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

[7] While distinctions exist between knowledge and information (e.g. Orlikowski 2002), we assume characteristics that make knowledge complex and costly to transfer influence the types of information employees in this firm send and receive.

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 match information about job candidates with information about positions available. Better information should improve the timeliness and quality of these matches increasing project completion rates and revenue generation.

H3: *Project completion and revenue generation by individuals is 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 on 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) surveys of demographic characteristics, education, and industry tenure. Email data cover 10 months of complete email history 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.[8] 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 in Aral, Brynjolfsson & Van Alstyne (2006). Since cryptographic techniques

[8] 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).

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. 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%. A data summary appears in Table 1.

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
Betweenness 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

3.2. Identifying Heterogeneous Information Types

Our initial dataset[9] consists of approximately 1.5 million words whose frequencies are distributed according to the standard Zipf's Law distribution (see Figure 1). We eliminated words with low information content by culling the most infrequent words (term frequency < 11), words that are too commonly used (term frequency > weekly), and words with low term frequency-cumulative inverse

[9] We thank <name deleted> for tireless coding efforts that extracted and manipulated the email data described in this section.

document frequency (tf-cidf), a common metric used to identify spikes in usage (Gruhl et. al. 2004).[10] These three methods reduced the sample to 120,000 words.

In selecting event news, we sought words with a spike in activity and a rapid, pervasive diffusion to members of the organization, followed by a decline in use. We chose words seen by more than 30 people with a coefficient of variation one standard deviation above the mean.[11] To select words likely to display rapid propagation, of words that reached 30 people, we selected words with a coefficient of variation of activity one standard deviation above the mean - 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).[12] 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.[13] The result is a sample of 3,275 words at first rarely used, then suddenly are used much more frequently and by more than 30 people, followed by a decline in use. We then selected a sample of discussion topic words where users both received and sent the word in email. This simple criterion selected approximately 4,100 words from the original candidate set. Examples of the usage characteristics of event news and discussion topics are shown in Figures 3 & 4. Words in the discussion topic sample display a lack of use, followed by a shallow diffusion to a limited number of people in back and forth discussion, which in the case of the word shown in Figure 4 lasts close to 3 months. These words are shared in back and forth conversation as shown in Figure 5. 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 &

[10] 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. The cutoff of 11 produced similar results as cutoffs in the neighborhood of 11.

[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 (see Figure 2). To be conservative, we selected any word seen by more than 30 people as a potential observation of event news.

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

[13] Event driven spikes in use *not* part of diffusion processes will downward bias our estimates, making them more conservative.

Kleinberg (2006) have noted, information cascades are typically shallow, but are 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 saw 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. T-tests demonstrate that they differ significantly across all dimensions of interest related to their use and diffusion (see Table 3).

3.3. Data Structure

We observe the diffusion of several thousand words of each information type 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, and the time between the first use of the word and the receipt of the word by each employee. An observation is a word-recipient pair (one for each possible recipient in the firm). For each word, our data record dyadic characteristics of each first user-recipient pair, such as the difference in their ages or industry tenures, for all potential recipients and individual characteristics of recipients (e.g. gender, network position).

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

Table 2: Correlation Matrix

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Gender (M = 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. Common Friends	.04	-.02	.03	.02	.50	.38	-.40	.08	1.00									
10. Betweenness 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. 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	1.00					
14. Same Area Specialty	-.01	-.16	-.00	.01	.12	.40	-.12	.27	.19	.05	-.04	.33	-.12	1.00				
15. Managerial Level Difference	.05	.52	.05	.11	.03	-.10	.01	-.03	.01	.02	-.07	.03	.50	-.21	1.00			
16. Partner	.21	.06	.06	.02	-.06	-.05	-.14	-.06	.07	-.03	-.31	.09	.26	-.05	.23	1.00		
17. Consultant	-.12	-.07	-.03	-.04	-.31	-.09	.29	-.26	-.19	-.22	.19	-.01	-.13	-.07	-.21	-.56	1.00	
18. Researcher	-.09	.01	-.03	.03	.40	.15	-.17	.35	.13	.27	.12	-.08	-.13	.14	-.01	-.44	-.51	1.00

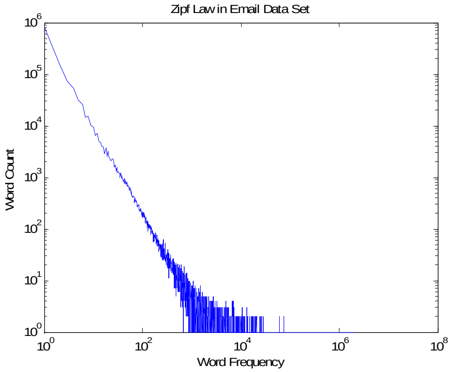

Figure 1. Distribution of Word Frequencies

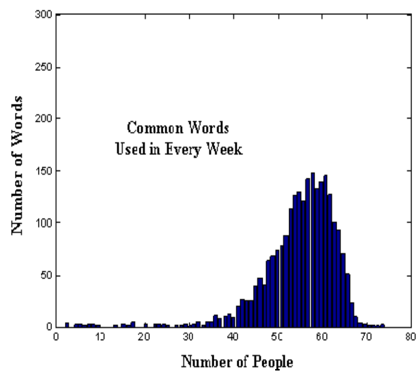

Figure 2. Distribution of Common Words

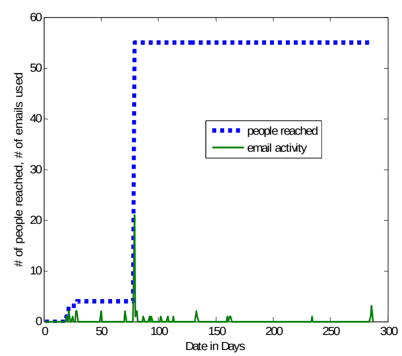

Figure 3. An Example Event News Item

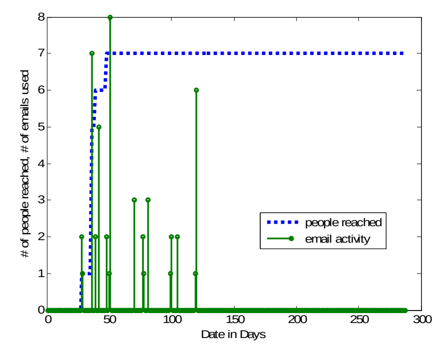

Figure 4. An Example Discussion Topic

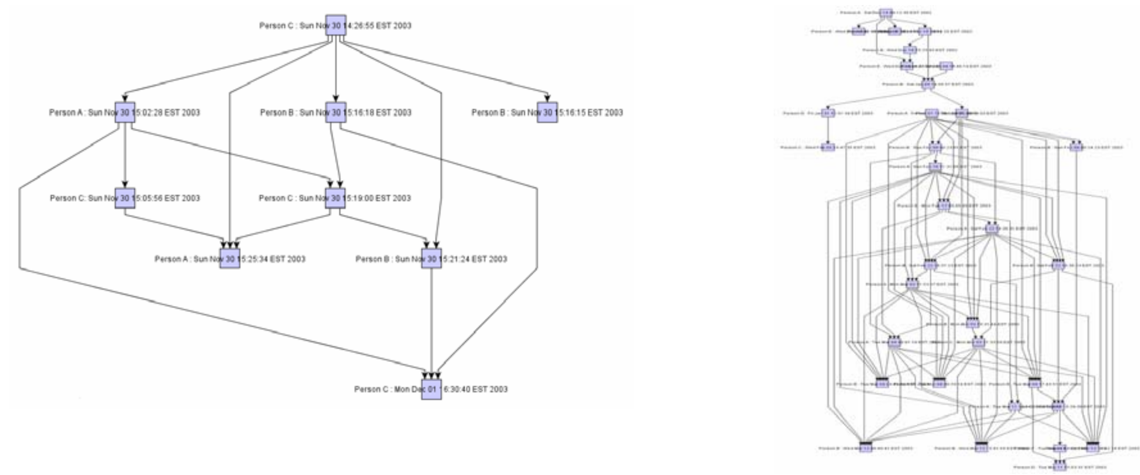

Figure 5. Discussion Paths for Two Discussion Topic Items

3.4. Statistical Specifications

We estimate the impact of hypothesized factors on the likelihood of seeing information and seeing it sooner. Linear estimates of probabilistic outcomes create bias due to non-linearity at upper and lower bounds of the likelihoods of discrete events. They are not well suited to temporal processes in which outcome variables can be conditioned on previous events and they also produce biased estimates of longitudinal data with right censoring (Strang & Tuma 1993). For these reasons we specify logistic and hazard rate models of diffusion. 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.

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

The parameters describe the impact of a given variable on the likelihood of receiving the word during the ten months of email observation. However, pooled cross sectional estimates may wash away temporal variation and allow later events to influence estimates of earlier diffusion (Strang & Tuma 1993).[14] We therefore estimate the rate of receipt of different types of information conditional on having received the

[14] We found no compelling evidence of duration dependence and proceeded with traditional estimations of the Cox model.

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. Coefficients estimate 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).

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

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

where 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.

4. Results

4.1. Estimation of the Diffusion of Information

We first tested the diffusion of all types of information through the firm (see Table 4).

Table 4. 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)*
Demographics[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)
Betweenness 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: Age, Ed, Industry Tenure not significant. * $p < .05$; ** $p < .01$; *** $p < .001$.		

Although employment at the firm is gender balanced and controlling for correlations between gender and organizational position (partner and consultant dummies), men are 55% more likely than women to receive information of all types. Demographic dissimilarity between originator and recipient reduces the likelihood of receiving information by between 1% and 13%, with gender differences recording the largest impact and age differences the smallest. The strength of ties increases the likelihood of receiving information. Ten additional emails sent increases the likelihood of receiving information by 2%. Path length reduces the likelihood of receiving information, with each additional hop reducing the likelihood of diffusion by 29%. Having friends in common seems to reduce the likelihood of receiving an information cascade. However, having friends in common is positively correlated with email volume and the strength of ties. Holding these variables constant, the initially positive effects of friends in common reduce and reverse. Betweenness centrality has a strong positive effect on the likelihood of receiving information, as do stronger project co-work relationships. Hazard rate model estimates of the drivers of the rate of information receipt reveal positive effects for project co-work and betweenness centrality, and a negative relationship between path length and the rate at which information is received. 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

Table 5 presents estimates of the drivers of event news and discussion topic diffusion. Demographic distance reduces the likelihood of receiving both news and discussion topics although with a slightly larger impact for news. One additional year of education difference between two individuals reduces the likelihood that news will diffuse between them by 7.5%, while reducing the likelihood of discussion topics diffusing by nearly 17%. Interestingly, men are over 50% more likely to see news than women although gender has no effect on the likelihood of the diffusion of discussion topics. 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. Ten additional emails exchanged increases the likelihood that discussion topics will diffuse by 7% on average. Path length 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 discussion diffusion by 97%, indicating discussion topics diffuse locally, while news travels across multiple hops. Betweenness centrality increases the likelihood of seeing both news and discussion topics. Perhaps most interestingly, 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%.

Table 5. 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)*
Demographics[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)**
Betweenness 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: Age, Edu, Industry Tenure not significant. Geographic Proximity: Same Office = 1; * p < .05; ** p < .01; *** p < .001.

Discussion topics are more likely to diffuse up and down the organizational hierarchy rather than laterally. As researcher is the omitted position category, strong positive estimates on partner and consultant variables demonstrate that discussion is more likely to diffuse upward rather than down the hierarchical structure of 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 has a strong positive impact on the hazard rate for discussion topics but not for news, while greater path lengths consistently reduce the hazard rate across both types of information. We see increases in the rate at which employees see discussion topics with greater project co-work (6.6% increase per additional project). Having the same area of expertise increases the rate while industry tenure differences have no effect. The partner and consult dummies show that employees in the top two levels of the organization see information at a higher rate.

4.3. Access to Information & Productivity

Table 6 presents estimates of the impact of access to information on the productivity of individual recruiters as measured by the number of projects completed.

Table 6. Information Diffusion & Project Completions

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Dependent Variable:	Completed Projects	Completed Projects	Completed Projects	Completed Projects	Completed Projects	Completed Projects	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

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. '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). The results show that access to information predicts project output. Each additional ten words seen are associated with an additional 1% of one project completed.

Table 7. 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

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 7 presents relationships between access to information diffusion and revenues generated, which can be thought of as a quality adjusted measure of output. These results show 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. Conclusion

We gather a unique collection of data to examine a set of frequently-made, but seldom-tested assumptions. We find that communication networks matter because they strongly influence information diffusion in firms, and access to novel information, such as new words in email communications, is a highly significant predictor of worker productivity. We demonstrate that demographics, organizational hierarchy, network topology, and task characteristics all influence the diffusion of information and the likelihood of involvement in information cascades. We also find that different types of information diffuse differently. While demographic distance reduces the likelihood of seeing both "news" and "discussion topics", task characteristics such as project co-work and industry tenure differences reduce the likelihood of receiving discussion topics more 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 diffuses laterally as well as vertically, and without regard to the strength of ties or function of relationships. The power of network structure to influence information diffusion validates one of the foundational assumptions of social network theory: information does not diffuse randomly in organizations, but rather reflects the nature and structure of human relationships.

Furthermore, these differences in diffusion patterns strongly predict productivity. Information workers who receive a greater volume of novel information or who receive it sooner complete projects faster and generate significantly more revenue for the firm. In our context, encountering ten novel words beyond the average predicts roughly 1% more of one project completion and $700 in incremental revenues. This effect is arguably the key justification for information systems, and the fundamental basis for the economic value of information.

The effort required to develop and analyze these data necessarily forced us to limit its scope. We focused on a single firm in a particular industry. Thus, we cannot claim that the pattern of relationships we uncover will apply equally to all firms in all industries for all time. Additional research will be needed to generalize our findings and identify factors that differentiate firm, industries and time periods. Furthermore, our approach, like most social science research, cannot directly test causality. We are working on field experiments to address this question in future work.

We are very optimistic about applying this methodology to information research. Direct observation of word-level information flows lets us open up the "black box" of the firm and explicitly test theories about information's effects in organizations. We expect this type of super-micro analysis of information flows and performance to become increasingly common in information systems research, providing a new frontier for theory and discovery.

References

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.

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

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.

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.

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.

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

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.

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

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

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

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.

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.

Kempe,D., Kleinberg, J., & E. Tardos. 2003. “Maximizing the spread of influence through a social network” *Proceedings of the 9th ACM SIGKDD*, Washington, D.C.: 137-146.

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

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

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.

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.

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.

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

Strang, D. & N. Tuma. 1993. "Spatial and temporal heterogeneity in diffusion" *Amer. Jrnl Soc.*, 99(3): 614.

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

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

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

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.