Antecedents and Consequences of Mutual Knowledge in Teams

Article *in* SSRN Electronic Journal · November 2008

DOI: 10.2139/ssrn.1299260

Erik Brynjolfsson
Massachusetts Institute of Technology

Marshall Van Alstyne
Boston University

Antecedents and Consequences of Mutual Knowledge in Teams

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

Erik Brynjolfsson
MIT Sloan School of Management, Room: E53-313, 50 Memorial Drive, Cambridge, MA 02142.
erikb@mit.edu

Marshall Van Alstyne
Boston University & MIT, 595 Commonwealth Avenue, Boston, MA 02215.
mva@bu.edu

A tension exists between two well-established streams of literature on the performance of teams. One stream contends that teams with diverse backgrounds, social structures, knowledge, and experience function more effectively because they bring novel information to bear on problems that cannot be solved by groups of homogeneous individuals. In contrast, the literature on mutual knowledge contends that shared information and experience is essential to effective communication, trust, understanding and coordination among team members. Furthermore, several distinct antecedents of mutual information and knowledge have been hypothesized, making it difficult to manage information overlap in teams. In this paper, we use a unique data set of observed email content from 1382 executive recruiting teams and detailed accounting data on their productivity to examine both the antecedents and performance effects of shared versus diverse information. We find clear evidence of an inverted-U shaped relationship between mutual information and team productivity. A significant amount of information overlap among team members is associated with higher performance while extremes of too little or too much mutual information hamper performance. We also find that geographic dispersion and social network distance are strong predictors of mutual knowledge failures, while demographic dissimilarity and organizational distance do not predict the degree of mutual information in our data. Our work helps bring together the divergent streams of literature on mutual knowledge, information diversity, and the management of team performance.

Key words: Mutual Knowledge, Diversity, Social Networks, Demography, Geographic Dispersion, Information Distance, Teams, Performance.

Introduction

A tension exists between two well established arguments about team performance. One stream of literature on structural and demographic diversity contends that teams with diverse backgrounds, social structures, local knowledge and prior experience should function more effectively by bringing the distinct knowledge and information of members to bear on problems that cannot be solved by the shared common knowledge of the team. According to this argument, team members combine their diverse experience, information and social network connections to bring unique benefits to the team. On the other hand, the literature on mutual knowledge emphasizes that shared information and experience enables effective communication, mutual understanding and coordination among team members. Diverse teams, that share little common information and knowledge have difficultly communicating and tend to develop ineffective communication processes that hamper the development of trust and mutual understanding. At the same time, several potential antecedents of mutual information and knowledge have been proposed including demographic similarity, structural cohesion, and geographic co-location. While many arguments about the antecedents and consequences of mutual information and knowledge exist in these various domains, little large scale empirical evidence examines mutual knowledge directly.

In this paper, we use a unique data set of the content of email communication among the members of 1382 executive recruiting teams and detailed accounting data on their productivity to examine both the antecedents and performance effects of shared mutual information and knowledge. We hypothesize that demographic dissimilarity, social network distance, geographic dispersion, organizational distance and differences in project expertise predict the degree to which information in individuals' email inboxes and outboxes is similar or dissimilar. We then construct precise measures of team productivity using detailed accounting data on teams' revenue output and labor inputs. We hypothesize that mutual information and knowledge enables more effective communication, coordination and trust building, but that teams also benefit from bringing unique and diverse knowledge and information to bear on project activities. We argue that the costs and benefits of mutual information combine to create a non-linear relationship between information overlap and team performance.

We find that there is an inverted-U shaped relationship between mutual information and team productivity. A healthy amount of information overlap among team members contributes to performance while too little or too much overlapping information hampers performance. This result helps resolve the apparent tension between arguments that detail the costs and benefits of mutual information and knowledge in teams. In addition, we provide some of the first large-scale quantitative evidence on the antecedents of mutual information in teams. We utilize a vector space model of information content in email communication to analyze the mutual information among team members, and estimate the degree to which different characteristics of teams predict the information overlap observed in email. Our results demonstrate that geographic dispersion and social network distance are the strongest predictors of mutual knowledge failures. In contrast, demographic dissimilarity and organizational distance do not predict the degree of mutual information among team members. While previous project co-work is weakly related to greater mutual information, when social network and geographic dispersion variables are entered into the analysis prior project co-work does not predict information overlap. Geographic dispersion and social networks are the two most salient characteristics of teams that predict mutual information in our setting. As such, managers can calibrate optimal information overlap among team members by analyzing geographic dispersion and social networks during team assignment processes. Our work contributes to the literature on diversity and mutual knowledge and helps managers manage optimal information overlap in teams.

The Mutual Knowledge Problem

Mutual Knowledge and Coordination

Mutual knowledge, the knowledge that communicating parties share in common and know they share (Krauss & Fussell 1990), is essential for mutual understanding, trust and effective communication and coordination in teams (Cramton 2001). The development of mutual knowledge among team members establishes a "common ground" that helps teams avoid misattribution, increases the likelihood that communication is understood, and ensures that perspectives on problem solving and execution are mutually recognized and effectively integrated across varying perspectives. Shared information is essential to developing mutual knowledge and understanding as it enables team members to 'get on the same page' and to understand the context and perspectives of their counterparts. The information exchanged in communication is particularly important because it enables team members to learn what others know (Cramton 2001: 347), an essential element of mutual knowledge and understanding (Clark 1996, Krauss & Fussell 1990). When collaboration occurs without mutual information and knowledge, comprehension is based on the recipient's own unique information, creating opportunities for misinterpretation and misunderstanding.

Mutual knowledge failures disrupt relationships among team members and reduce decision quality and productivity. Mutual knowledge and information are therefore considered "a precondition for effective communication and the performance of cooperative work" (Cramton 2001: 349). Establishing informational common ground enables team performance for several reasons. First, shared information prevents misattribution and misunderstanding which disrupt communication processes (Stasser & Stuart 1992) and relationships among team members (Cramton 2001). Shared information enables the development of communicative and behavioral norms that build trust, guide relationship development and prevent affective, process and task conflict in teams (Jackson 1965, Eisenhardt et al. 1997, Hinds & Bailey 2003). Development of trust and avoidance of conflict are important for information sharing and positive, efficient group interaction essential for effective collaboration (Hinds & Bailey 2003). Second, shared information enables team members to know 'who knows what' increasing the likelihood that the most relevant expertise is brought to bear on tasks and problems encountered during teamwork (Stasser et al. 1995, Contractor 2000). Mutual knowledge increases the likelihood of information sharing and the effectiveness of knowledge transfers (Reagans & McEvily 2003), and helps teams convey information necessary for joint decision making and problem solving (Denis 1996, Stasser & Titus 1985). Third, shared information enables group identification which in turn enables effective collective action (Portes & Sensenbrenner 1993, Coleman 1988, Reagans and Zuckerman 2001). Common goals and shared understanding also facilitate coordination (Van Alstyne 1997). For these reasons, shared mutual information promotes the development of common ground, enables effective communication, improves joint decision making and task execution, and creates a will toward collective action that should improve teams' productivity.

Information Diversity and Team Productivity

While teams who share significant amounts of mutual information and knowledge may be more harmonious and better able to communicate and coordinate effectively, knowledge homogeneity can reduce learning (Pfeffer 1983), creative problem solving (Reagans & Zuckerman 2001) and ultimately performance (Ancona & Caldwell 1992, Reagans & Zuckerman 2001, Cummings 2004). Scientific breakthroughs have occurred by combining expertise (Van Alstyne & Brynjolfsson 1996, 2005), and teams whose members combine different skills, information, and perspectives are more likely to recognize opportunities and bring useful information to bear on their tasks (Hong & Page 2001). This has the ironic consequence that teams with less expertise but more diversity can outperform teams with more expertise but less diversity. One reason is that the best problem solvers tend to have similar skills, therefore collections of the best problem solvers can perform little better than individuals. In contrast, more diverse teams can apply broader more novel expertise, making it possible for "diversity to trump ability" (Page 2007). Varied perspectives can also affect performance since what is good for the individual is not always good for the group, and among homogeneous teams "group think" can retard the acceptance of new ideas (March 1991).

Creativity and productivity depend critically on novel information (e.g. Burt 1992, Reagans & Zuckerman 2001). In Burt's memorable words, "creativity is an import-export game,… not a creation game." (Burt, 2004b). For instance, Hargdon and Sutton show how engineers broker information flows among industrial sectors and note that actors with access to diverse pools on information "benefit from disparities in the level and value of particular knowledge held by different groups…" (Hargadon & Sutton

1997: 717). Similarly, Cummings (2004) finds that more diverse teams can draw on unique local information to improve their responsiveness to various stakeholders.

New information becomes useful when it is linked to information that a person already knows. In turn, as a person's information base grows, this can increase the effectiveness of additional knowledge transfer (Cohen & Levinthal 1990, Simon 1991). Furthermore, others may also be more likely to share information with a person who has a greater knowledge base and absorptive potential (Reagans & McEvily 2003, Rodan & Galunic 2004).[1] The behavior of the executive recruiters in our study is consistent with these theories. They report being more effective when they have more diverse information about candidates' industries and experience. The greater the diversity of information the more likely they can find a match with a specific position.

Reconciling the Costs and Benefits of Mutual Information in Teams

Prior theory highlights both costs and benefits to mutual information in teams. On one hand, mutual information promotes effective communication and coordination; on the other hand, diverse information provides novel expertise, and improves decision making, learning, and creative problem solving. Both factors can improve productivity. Empirical research on the performance consequences of team diversity and mutual information and knowledge are inconclusive. No consistent effects of diversity or information overlap have been reported and no consensus exists on the role of mutual information in team performance (Williams & O'Reilly 1998, Jehn et. al. 1999). Some studies find that informational diversity improves performance while others find that diversity creates conflict, hampers coordination and reduces performance, while still others find no conclusive results. Several empirical studies have tried to qualify these relationships by exploring the conditions under which diversity and mutual knowledge may positively or negatively impact performance. For example, some work has examined the moderating effects of personal conflict, arguing that informational diversity improves performance if teams can manage

[1] There may be additional benefits, as documented in the research literature, such as, the potential to increase autonomy (Simmel 1923, Burgelman 1991, Burt 1992), political maneuverability (Padgett & Ansell 1993), or access to resources (Rodan & Galunic 2004).

conflict effectively (Pelled 1996). Other work has demonstrated the importance of task characteristics. For example, Jehn et. al. (1999) find that informational diversity increases team performance when tasks are complex rather than routine, and Van de Van et. al. (1976) show that task interdependence moderates the relationship between diversity and performance. These studies and others like them further our understanding of the relationship between informational diversity and team performance by taking a contingency perspective on performance relationships. We complement and extend this thinking by exploring a related yet distinct perspective which views the relationship between mutual knowledge and performance as non-linear. We hypothesize that while there may be linear contingent effects of knowledge diversity on team performance, there may also be a generalizable non-linearity in the relationship between information overlap and performance.

While prior empirical investigations of mutual knowledge in teams could not confirm an effect on team productivity (Cramton 2001), there is evidence that the performance benefits of access to diverse information are non-linear. Aral & Van Alstyne (2007) show that in the context of executive recruiting there are diminishing marginal productivity returns to novel information. They argue that limits to human cognitive capacity, bounded rationality, and information overload create positive but diminishing performance returns to diverse, novel information.

These conflicting views can be represented in a simple model that accounts for the benefits of both similarity and diversity. We follow common practice and represent mutual knowledge as overlapping expertise in an array of possible topic areas (Blau 1977; March 1991; Hong & Page 2001; Reagans & McEvily 2003; Aral & Van Alstyne 2007), and then use a natural proximity metric – cosine similarity – to measure the degree to which information in teams is mutual. This measures the degree to which expertise profiles (i.e. vectors of expertise levels across multiple topics) point in the same direction – the smaller the angle, the greater the similarity and the greater the mutual knowledge. To measure the benefits of diverse knowledge, an obvious extension is a sine index of dissimilarity – the greater the angle, the greater the dissimilarity, and the greater the value of novel information. The composition of these two performance benefits, one for information similarity $Cos[\angle\ ]$ and one for information dissimilarity

$Sin[\angle\ ]$, defines a natural inverted-U shape of performance as shown in Figure 1. Relative weights can favor either more diverse or more mutual knowledge. Equation 1 represents these tradeoffs in a parsimonious expression that captures the benefits of both information diversity and mutual knowledge for team performance, which vary with the relative importance of either dimension across different work contexts:

$$Performance = \kappa * Cos(\omega\angle) * Sin(\omega\angle) \qquad [1]$$

The benefit of mutual knowledge is captured by the cosine term and scaled by a parameter ω which indicates the degree to which mutual knowledge or information diversity is more important to team performance in a given work environment. The benefit of diversity is captured by sine term and scaled by the same parameter ω. As ω increases, the salience of information diversity for determining team performance increases. On the other hand, as ω decreases, the salience of mutual knowledge for team performance increases.[2] The parameter κ simply scales the magnitude of the impact of either information diversity or mutual knowledge on team performance to accommodate for the possibility that in some contexts the distribution of information among team members may be more important than in others.

We speculate that the tradeoff described in this simple framework makes it more difficult to detect clear positive or negative performance effects of information overlap in teams and that the costs and benefits of mutual information combine to create an inverted-U shaped relationship between mutual information and team productivity. We hypothesize that with too little information overlap teams find communication and collaboration difficult, but that too much mutual information makes the contribution of team members redundant, reducing problem solving efficiency and productivity.

> *Hypothesis 1: There is an inverted U-shaped relationship between mutual information and team productivity.*

Figure 1 describes how the costs and benefits of mutual information in teams combine to create an inverted-U shaped relationship with team performance. Such a relationship implies an optimal infor-

[2] In order to capture the tradeoff between diversity and mutual knowledge and to restrict the values of the parameter which adjusts for their relative importance across work environments, we constrain ω such that $0 < \omega < \frac{\pi\angle}{2}$.

mation overlap in teams (MK*) whereby top performing teams have neither too little nor too much mutual information among team members. In different work contexts the optimal level of information overlap in teams may change. As the importance of information diversity for team performance increases (ω increasing), the optimal information overlap (MK*) moves to the left implying that greater diversity is beneficial. As the importance of mutual information for team performance increases (ω decreasing), the optimal information overlap (MK*) moves to the right implying that greater mutual knowledge is beneficial. These varying contexts capture the relationship between mutual knowledge and team performance in common team environments such as innovation teams or routine operations and production teams. These contexts also map directly to James March's conceptions of exploitation and exploration (March 1991). For teams whose performance is tied to creativity and innovation, information diversity may be more important, while for teams whose performance is tied to efficient exploitation of known organizational processes or tasks, mutual knowledge may be more important.

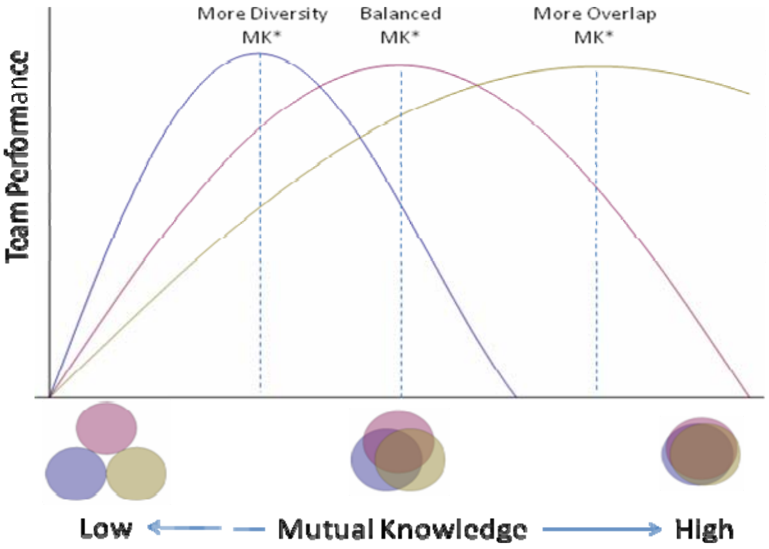


Figure 1. An inverted U-shaped relationship shows the benefits of mutual knowledge for team performance. This plots the composite benefits of (information overlap)*(information diversity) modeled as $Cos[\angle]*Sin[\angle]$ for the angle between knowledge profiles.

This parsimonious and extensible framework for evaluating the relationship between knowledge overlap and team performance accommodates the varying importance of information diversity and mutual knowledge while maintaining the non-linearity of the relationship between mutual knowledge and performance. We expect to see such a curvilinear relationship in our data.[3]

Antecedents of Mutual Information and Knowledge

Managing optimal knowledge and information overlap in teams requires an understanding of both the non-linearity of performance effects and the antecedents of greater overlap or divergence. In this section we review literature addressing potential drivers of information overlap and build hypotheses about team characteristics that may predict the degree of information overlap in teams.

Demography and Expertise

Demographic homogeneity is linked to shared knowledge and information in a variety of literatures. The literature on organizational demography contends that homogeneity in organizational and industry tenure breeds shared experience, shared identity and shared knowledge of circumstance and activity (Blau 1977; Pfeffer 1983). Employees who enter an organization at the same time develop cohort affinity and redundancy in perspectives and information (e.g. Ancona & Caldwell 1992, Reagans & Zuckerman 2001). Cohort affinity increases information sharing among organizational groups with similar industry experience and organizational tenure. Differences in educational background, and industry and organizational tenure create informational diversity in workgroups, and create diversity in perspectives and opinions as well (Stasser 1992, Jehn et. al. 1999). In fact, the sharing of information and the development of shared perspectives complement one another to create harmonious interaction and cooperation which is essential to the argument that shared information improves coordination, cooperation and ultimately team performance. At the same time, social category diversity and differences in social category membership, such as age, gender and ethnicity create social identity effects (Tajfel & Turner 1986) that

[3] It should be noted that our framework remains agnostic about whether innovation requires more diverse information while routine contexts require more mutual information. The framework simply predicts performance as a non-linear function of mutual knowledge that can vary with the varying relative importance of diversity or similarity.

are associated with shared prior experience, shared information, and a greater likelihood of interaction and communication on common topics and issues (Jackson 1992, McGrath et. al. 1996, Pelled 1996, Jehn et. al. 1999). We therefore expect that demographic and expertise distances are negatively associated with information overlap in teams.

> *Hypothesis 2a: Demographic distance - differences in age and gender - is negatively associated with information overlap in teams.*

> *Hypothesis 2b: Expertise distance - differences in education, industry experience, and organizational tenure - is negatively associated with information overlap in teams.*

Geographic Dispersion

Geographic dispersion prevents the development of shared experience and context, making it difficult to share perspectives and common information about work environments, task experiences, prior physical and contextual experience, and to develop multidimensional relationships that are essential to information exchange across topics of shared interest (Schober 1998, Mortenson & Hinds 2001, Hinds & Bailey 2003). Interaction among people in the same organization falls dramatically as a function of distance (Allen 1977). Geographic distance creates different perspectives on task related activities in work groups and disparities in the information team members have about project work (Tyre & von Hippel 1997). Without shared context, team members are less able to develop mutual understanding and common knowledge (Fussell & Krauss 1992). Information is a critical part of developing mutual understanding. As collocation breeds familiarity and friendship through informal interaction, unplanned encounters and multipurpose activities, and the availability of visual cues (Grinter et al. 1999, Kraut et. al. 2002, Hinds & Bailey 2003), and as familiarity increases the frequency of information exchanges across a greater number of topics, we expect dispersed teams to communicate on fewer mutual topics, and to have less in common to talk about. We therefore expect dispersed teams to share less common information.

> *Hypothesis 3: Geographic dispersion is negatively associated with information overlap in teams.*

Social Networks

In the research literature, it is often assumed that network distance corresponds to information distance. Individuals whose network includes more common contacts and many strong ties are presumed to swim in the same pools of information. For instance, network cohesion may increase the likelihood of information sharing and the effectiveness of knowledge transfer between individuals (Reagans & McEvily 2003). Social cohesion motivates individuals to devote time and effort to communicating with and assisting others due to the cooperative nature of ties surrounded by other third party ties (Granovetter 1985, Coleman 1988).

As a result, Granovetter (1973) argues that weak ties will deliver more novel information about socially distant opportunities. Specifically, contacts maintained through weak ties typically "move in circles different from our own and thus have access to information different from that which we receive… [and are therefore]… the channels through which ideas, influence, or information socially distant from ego may reach him" (Granovetter 1973: 1371). Similarly, Burt (1992) argues that information in local network neighborhoods is typically redundant so novel information will disproportionately come from structurally diverse contacts.

A closely related point is that longer path lengths increase the likelihood and severity of distortion as information is passed from individual to individual in a network via misunderstanding, vagueness, filtering or even deliberate withholding and falsification (March & Simon 1958, Huber 1982, Hansen 2002, Huber & Daft 1987). A common example is the "telephone game" in which messages are distorted as they are passed along a chain of contacts (e.g. Aral et. al. 2006, Van Alstyne & Brynolfsson, 2005). Furthermore, intermediaries must be willing to pass information even if it has no direct value to them. Using email, a modern recreation of Milgram's famous letter passing experiment showed that 98% of chains between geographically separated and unfamiliar participants failed to complete (Dodds et. al. 2003). We therefore expect longer path lengths, weak ties and fewer contacts in common to be associated with lower information overlap in teams.

Hypothesis 4: Network distance - long path lengths, weak ties, and a lack of network cohesion - is negatively associated with information overlap in teams.

The Setting – Executive Recruiting[4]

We studied a medium-sized executive recruiting firm with fourteen offices across the United States. Our interviews revealed that the core of executive recruiters' work involves matching job candidates to clients' requirements. This matching process is information-intensive and requires activities geared toward assembling, analyzing, and making decisions based on information gathered from team members, other firm employees, and contacts outside the firm. The process for executing a contract is relatively standardized: A partner secures a contract with a client and assembles a project team (team size mean = 1.9, min = 1, max = 5). The team then establishes a universe of potential candidates including those in similar positions at other firms and those drawn from the firm's internal database of resumes and other leads.[5] These candidates are vetted on the basis of perceived quality, their match with the job description and other factors. After conducting initial due diligence, the team chooses a subset of candidates for internal interviews, approximately six of which are forwarded to the client along with detailed background information, notes and a formal report of the team's due diligence. The team then facilitates the client's interviews with each candidate, and the client, if satisfied with the pool, makes offers to one or more candidates. A contract is considered complete when a candidate accepts an offer.

Qualitative studies have described executive recruiting teams as "brokers" between clients and candidates and found that they rely heavily on information flows to complete their work effectively (Finlay & Coverdill 2000). In our context, more precise or accurate information about the candidate pool reduces time wasted interviewing unsuitable candidates and increases the quality of placement decisions (Aral et. al. 2006). In addition, the sharing of procedural information can improve efficiency and effectiveness (Szulanski 1996) and executive recruiters report learning to deal with difficult situations through communication with peers. Effective recruiters rely on being "in the know" and delivering candidates that

[4] The description of the setting, data and methods for this study draw heavily on our prior work as document in Aral et al. (2006), Aral and Van Alstyne (2007) and Aral et al. (2007). Additional details can be found therein.

[5] We have also studied executive recruiters' use of information contained in the firm's internal database or 'Executive Search System.' For more detailed analyses on how use of the Executive Search System impacts productivity and performance see Aral et. al. (2006).

display professional and personal attributes that fit client needs. To accomplish this, recruiters must be aware of several different information channels to match different candidates with different client requirements.

Methods

Data

We are able to precisely measure the overlap of information recruiters send and receive by analyzing message content in email communication. Although recruiters exchange information through several channels, including face to face communication and phone conversations, email provides a context in which we can analyze written transcripts of electronic communication. In contrast, instant messaging is not widespread in the firm that we studied, so our analysis is relatively comprehensive in its coverage of codified information exchanges. Since our content measures consider the similarity of topics across the entire network, poor coverage of the firm could bias our estimates of the relative novelty or diversity of topics discussed via email. We therefore take several steps to ensure a high level of participation (described below). 87% of eligible recruiters agreed to participate. Our inability to observe the remaining 13% is limited to messages between two employees who both opted out of the study. As a result, we have nearly full coverage of the firm's email network and individual content data.[6]

Our data come from three sources: (i) detailed accounting records of project assignments and team productivity, (ii) email data from the corporate server, and (iii) survey data on demographic characteristics, human capital and information seeking behaviors. Internal accounting data describe: revenues generated per project, contract start and stop dates, project team composition and share weighted labor inputs into each project by each recruiter. These data provide excellent productivity 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

[6] F-tests comparing performance levels of those who opted out with those who remained did not show statistically significant differences. F (Sig): Rev02 2.295 (.136), Comp02 .837 (.365), Multitasking .386 (.538).

and from October 1, 2003 to March 1, 2004. Participants received $100 in exchange for permitting use of their data, resulting in 87% coverage of eligible recruiters and more than 125,000 email messages captured. Details of email data collection are described by Aral et. al. (2006) and Van Alstyne & Zhang (2003). The third data set contains survey responses on demographic and human capital variables such as age, education, industry experience, and information-seeking behaviors. 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%.[7]

A Vector Space Model of Mutual Information in Email Communication

Vector Space Models are widely used in information retrieval and search query optimization algorithms to identify documents that are similar to each other or pertain to topics identified by search terms. They represent textual content as vectors of topics in multidimensional space based on the relative prevalence of topic keywords. We therefore measure the mutual information in recruiting teams' email communication using a Vector Space Model of the topics present in email content (e.g. Salton et. al. 1975).[8] In our model, each email is represented as a multidimensional 'topic vector' whose elements are the frequencies of keywords in the email. The prevalence of certain keywords indicates that a topic that corresponds to those keywords is being discussed. For example, an email about a job in the health care sector might include frequent mentions of the words "hospital," "nursing," and "medical;" while an email

[7] We wrote and developed email capture software specific to this project and took multiple steps to maximize data integrity. 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 and content was masked using cryptographic techniques to preserve privacy (see Van Alstyne & Zhang 2003). Spam messages were excluded by eliminating external contacts who did not receive at least one message from someone inside the firm.

[8] While email is not the only source of employees' communication, it is one of the most pervasive media that preserves content. It is also a good proxy for other social sources of information in organizations where email is widely used. In our data, the average number of contacts by phone (ρ= .30, $p < .01$) and instant messenger ($\rho = .15$, $p < .01$) are positively and significantly correlated with email contacts. Our interviews indicate that in our firm, email is a primary communication media.

about an information technology job might mention the words “computing,” “programming,” and “technology.” The relative topic similarity of two emails can then be assessed by topic vector convergence or divergence – the degree to which the vectors point in the same or orthogonal directions.[9] To measure mutual information, we characterize all emails as topic vectors and measure the similarity or dissimilarity of topic vectors in individuals’ inboxes and outboxes. Emails about similar topics contain similar language on average, and vectors used to represent them are therefore closer in multidimensional space. The mutual information of teams is then measured as the relatively similarity or dissimilarity of the topic vectors that collectively represent team members’ emails.

Construction of Topic Vectors & Keyword Selection. Vector Space Models characterize documents D_i by keywords k_j weighted according to their frequency of use. Each document is represented as an n-dimensional vector of keywords in topic space,

$$\overrightarrow{D_i} = (k_{i1}, k_{i2}, ..., k_{in}),$$

where k_{ij} represents the weight of the *j*th keyword.

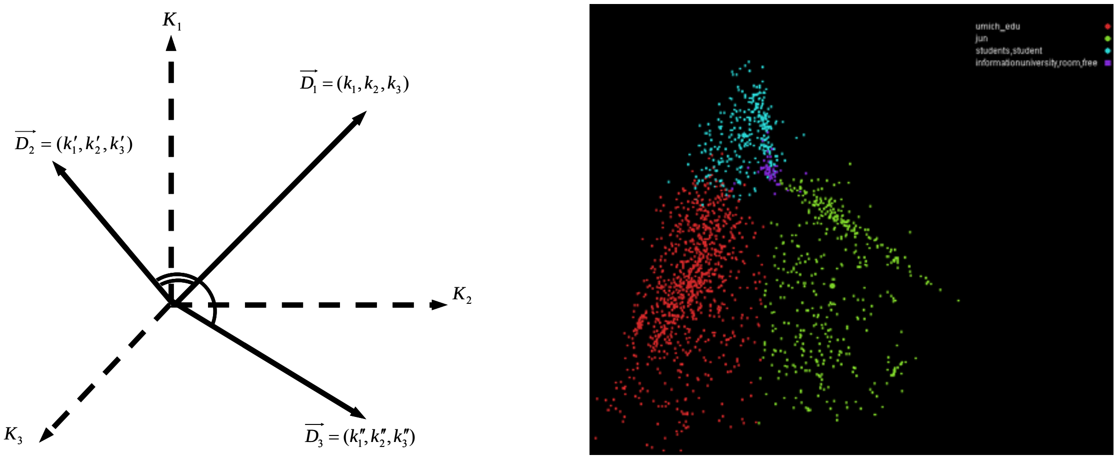

K_1
$\overrightarrow{D_1} = (k_1, k_2, k_3)$
$\overrightarrow{D_2} = (k_1', k_2', k_3')$
K_2
K_3
$\overrightarrow{D_3} = (k_1'', k_2'', k_3'')$
umich_edu
jun
students,student
informationuniversity,room,free

Figure 2. A three dimensional Vector Space Model of three documents is shown on the left. A Vector Space Model containing a test inbox with emails clustered along three dimensions is shown on the right.

[9] Each email may pertain to multiple topics based on keyword prevalence, and topic vectors representing emails can emphasize one topic more than another based on the relative frequencies of keywords associated with different topics. In this way, our framework captures nuances of emails that may pertain to several topics of differing emphasis.

Weights define the degree to which a particular keyword impacts the vector characterization of a document. Words that discriminate topics are weighted more heavily than words less useful in distinguishing topics. As terms that appear frequently in a document are typically thematic and relate to the document's subject matter, we use the 'term frequency' of keywords in email as weights to construct topic vectors and refine our keyword selection with criteria designed to select words that *distinguish* and *represent* topics.[10]

In order to minimize their impact on the clustering process, we initialized our data by excluding common "stop words," such as "a, "an," "the," "and," and other common words with high frequency across all emails that are likely to create noise in content measures (these so-called "stop words" produce 0 weights in the vectors above). We then implemented an iterative, k-means clustering algorithm to group emails into clusters that use the same words, similar words or words that frequently appeared together.[11] The result of iterative k-means clustering is a series of assignments of emails to clusters based on their language similarity. Rather than imposing exogenous keywords on the topic space, we extract topic keywords likely to characterize topics by using a series of algorithms guided by three basic principles.

First, in order to identify distinct topics in our corpus, keywords should *distinguish* topics from one another. We therefore chose keywords that maximize the variance of their mean frequencies across k-means clusters. This refinement favors words with widely differing mean frequencies across clusters, retaining words with an ability to distinguish between topics. In our data, we find the coefficient of variation of the mean frequencies across topics to be a good indicator of this dispersion.[12]

[10] Another common weighting scheme is the 'term-frequency/inverse-document frequency.' However, we use a more sophisticated keyword selection refinement method specific to this dataset described in detail in the remainder this section.

[11] K-means clustering generates clusters by locally optimizing the mean squared distance of all documents in a corpus. The algorithm first creates an initial set of clusters based on language similarities, computes the 'centriod' of each cluster, and then reassigns documents to clusters whose centriod is the closest to that document in topic space. The algorithm stops iterating when no reassignment is performed or when the objective function falls below a pre-specified threshold.

[12] The coefficient of variation is particularly useful due to its scale invariance, enabling comparisons of datasets, like ours, with heterogeneous mean values (Ancona & Caldwell 1992). To ease computation we use the square of the coefficient of variation, which produces a monotonic transformation of the coefficient without affecting our keyword selection.

$$C_v = \frac{\sqrt{\frac{1}{n}\sum_i \left(m_i - \overline{M}\right)^2}}{\overline{M}}$$

Second, keywords should *represent* the topics they are intended to identify. In other words, keywords identifying a given topic should frequently appear in emails about that topic. To achieve this goal we chose keywords that minimize the mean frequency variance within clusters, favoring words that are consistently used across emails discussing a particular topic:[13]

$$ITF_i = \frac{\sqrt{\sum_c \sum_i \left(f_i - \overline{M_c}\right)^2}}{\overline{M_c}}$$

Third, keywords should not occur too infrequently. Infrequent keywords will not represent or distinguish topics and will create sparse topic vectors that are difficult to compare. We therefore select high frequency words (not eliminated by the "stop word" list of common words) that maximize the inter-topic coefficient of variation and minimize intra-topic mean frequency variation. This process generated topical keywords from usage characteristics of the email communication of employees at our research site.[14]

Measures of Mutual Information. Using the keywords generated by our usage analysis, we populated topic vectors representing the subject matter of the emails in our data. We then measured the similarity and dissimilarity of information in teams' email communication by measuring the distance of email topic vectors that characterized team members' inboxes and outboxes. We created five separate measures of information distance based on techniques from the information retrieval, document similarity and information theory literatures (see Appendix A for detailed descriptions of each measure). The approach of all five measures is to compare the distance of topics in team individuals' emails, and to characterize the degree to which emails are similar or dissimilar. We used two common document similarity measures

[13] i indexes emails and c indexes k-means clusters. We squared the variation to ease computation.

[14] We conducted sensitivity analysis of our keyword selection process by choosing different thresholds at which to select words based on our criteria and found results were robust to all specifications and generated keyword sets more precise than those used in traditional term frequency/inverse document frequency weighted vector space models that do not refine keyword selection.

(Cosine similarity and Dice's coefficient) and three measures enhanced by an information theoretic weighting of emails based on their "information content."[15] We performed extensive validation tests of our measures of mutual information and their correlations, including application to an independent dataset from Wikipedia. A detailed description of the validation process and results appears in Appendix B. As all diversity measures are highly correlated (~ corr = .98; see Appendix A), our specifications use the aggregated cosine distance of team members' incoming and outgoing email topic vectors ID_{ij} to measure information distance:

$$ID_{ij} = \sum_{n=1}^{N}\sum_{k=1}^{K}\left(Cos\left(d_{in}, d_{jk}\right)\right)^2 \text{, where:}$$

$$Cos(\angle) = Cos(d_{i1}, d_{j1}) = \frac{d_{i1} \bullet d_{j1}}{|d_{i1}||d_{j1}|} \text{, such that } 0 \leq ID_{ij} \leq 1.$$

This measure aggregates the cosine distances of email vectors in team members' inboxes and outboxes, approximating the similarity or dissimilarity of the information content team members received and sent.

Measuring Team Productivity

We measure the productivity of executive recruiting teams by the amount of revenue they generated per person day of labor input. Recruiters' primary task is to fill vacant positions for client firms. They work in teams of one to five members and earn revenue for the firm equal to one third of the salary of the position they fill. As such, their productivity can be measured by an intuitive assessment of the real economic output they generate (revenues earned for the firm) divided by the labor inputs (in time and effort) they put into a project. As recruiters take on multiple projects simultaneously, they are typically working on a number of projects in parallel (Aral et. al. 2006). We therefore measure labor inputs by the shared weighted number of days team members collectively spend on a given project. We create this

[15] Information Content is used to describe how informative a word or phrase is based on its level of abstraction. Formally, the information content of a concept c is quantified as its negative log likelihood –log p(c).

measure by assigning a labor day to the project for each day a team member worked on the project, divided by the number of other projects the team member was working on during that day. For instance, if a team member is working on three total projects on one of the days they are working on the given focal project, we assign 1/3rd of a day of labor input to that project from that team member. When that team member takes on an additional project, we start assigning 1/4th of a day's labor input to the focal project from that team member until they complete a project or take on another project. We sum these labor inputs across all team members for every day that the focal project is active. Team productivity is therefore defined as real output in revenues divided by total labor input measured as a sum of the person days that all project team members contribute to that project:

$$Performance = \left(\frac{TeamOutput}{TeamInput}\right) = \left(\frac{R_p}{\sum_i \sum_t \left(\frac{1}{MT_{it}}\right)}\right),$$

where R_p is the revenue value of the project and MT_{it} is the number of other projects team member *i* worked on during day *t* of the focal project (the amount of project 'multitasking' person *i* is engaged in on that day) (Aral et. al. 2006). As teams fill vacancies more quickly they generate more revenue with fewer labor day inputs and are therefore more productive in real economic terms.

In our context, measuring team performance based on economic productivity is the most appropriate way to capture the multidimensional performance of executive recruiting teams. Recruiters generate revenue by filling positions for clients. They earn revenue for the firm equal to one third the salary of each placed candidate. Filling a position entails meeting the requirements and the minimal thresholds of quality client firms expect during a search. A recruiting team earns the same revenue for filling a position in three months as they do for filling it in four or five months. However, more labor inputs (and costs in salaries and firm resources) are expended as a search takes longer. Therefore, the speed with which recruiting teams fill positions and the revenues they generate relative to labor inputs together measure the efficiency with which they create economic output and therefore their relative productivity.

Table 1. Descriptive Statistics

Variable	Obs.	Mean	SD	Min.	Max
Dyadic Variables					
Same Gender	69	.55	.50	0	1
Age Difference	55	21.26	10.22	4	39
Education Difference	55	.95	1.02	0	3
Industry Experience Diff.	55	27.11	8.96	7	38
Org. Position Difference	69	.88	.75	0	2
Project Co-Work	69	.41	1.30	0	9
Same Region	69	.15	.35	0	1
Same Office	69	.48	.50	0	1
Network Distance	69	1.44	.50	1	2
Contacts in Common	69	24.38	9.28	1	37
Tie Strength	69	2.57	4.85	0	29
Dyadic Information Overlap	69	.61	.12	.24	.93
Team Variables					
Team Size	1382	1.98	.60	1	5
Age	1372	45.07	7.77	27	63
Education	1372	17.74	1.02	15	20
Industry Experience	1372	14.47	7.94	1	39
Mutual Information	994	0	1	-2.89	3.59
Mutual Info. Received	994	0	1	-2.75	2.98
Mutual Information Sent	994	0	1	-1.55	1.82

Estimation Procedures

We recorded data on the dyadic relationships between executive recruiters in a set of matrices that reflected the values characterizing each relationship. For example, we created matrices that recorded the age differences or network distance between recruiters in the cells of the matrices that corresponded to each dyad. In the case of distance variables, we measured the absolute value of the differences between individuals. In the case of gender differences and geographic dispersion variables (same office and same region), we constructed binary dyadic variables X in which $X_{ij} = 1$ if two recruiters i and j had the same value and $X_{ij} = 0$ otherwise. We used these dyadic variables to test the following model predicting the information distance between recruiters, where information distance is measured as the cosine distance between feature vectors characterizing the information content in recruiters' email messages:

$$\begin{aligned} ID_{i,j} &= \gamma_{i,j} + \beta_1 GD_{i,j} + \beta_2 AD_{i,j} + \beta_3 ED_{i,j} + \beta_3 IED_{i,j} + \beta_4 OTD_{i,j} + \beta_5 PCW_t + \beta_6 SO_i \\ &+ \beta_7 SR_i + \beta_8 ND_{i,j} + \beta_9 CC_{i,j} + \beta_{10} TS_{i,j} + \varepsilon_{i,j} \end{aligned} \quad [2]$$

The model estimates relationships between information distance ($ID_{i,j}$), gender difference ($GD_{i,j}$), age difference ($ID_{i,j}$), education difference ($ED_{i,j}$), industry experience difference ($IED_{i,j}$), organizational tenure difference ($OTD_{i,j}$), prior project co-work ($PCW_{i,j}$), whether recruiters work in the same office ($SO_{i,j}$), the same region ($SR_{i,j}$), network distance ($ND_{i,j}$), the number of contacts recruiters have in common ($CC_{i,j}$), and tie strength ($TS_{i,j}$).

As dyadic network and relational data do not constitute independent observations in the classical statistical sense, we utilize multiple regression quadratic assignment procedure (MRQAP) to estimate the model parameters. MRQAP utilizes a randomized permutation procedure to construct significance tests that account for the non-independence of observations (Krackhardt 1988, Borgatti & Cross 2003). Systematic interdependence arises is dyadic network data because each row value on a given dimension of interest measures the relationship from one individual in the data to all the other individuals in the data. If for instance a given recruiter is a well connected social butterfly, with strong links to many other employees, this behavior characteristic will show up in the strength of tie matrix as above average tie strength with every other member of the organization. Quadratic assignment procedure (QAP), developed by Hubert (1987) and extended to multivariate settings by Krackhardt (1988), proceeds by first regressing the dependent variable matrix on independent variable matrices to recover unbiased estimates of the beta parameters of the model (Judge 1990). Although the parameter estimates are unbiased, traditional estimates of standard errors are sensitive to network autocorrelation creating biased significance tests that tend to underestimate the standard errors and thus overestimate the significance of correlated observations (Hinds et al 2000). To account for this bias, QAP and MRQAP procedures create a reference distribution against which coefficients are compared by randomly permuting the dependent variable matrix multiple times (in our case 9,999 times), regressing the permuted matrix on the independent variable matrices and comparing resulting coefficients against the observed beta parameter estimates. If less than 1% the observed betas are larger than the betas generated under the randomized permutation procedure, then the observed coefficient is said to be significantly different from random at the .01 confidence level. Parameter estimates of MRQAP have been shown to be robust against interdependence of observations problems typically observed in networked data (Krackhardt 1988, 1993, Carley & Krackhardt 1996).

After estimating associations between dyadic relational data, we estimated the productivity of recruiting teams as a function of their mutual information using the following model:

$$P_i = \alpha_i + \beta_1 S_i + \beta_2 A_i + \beta_3 E_i + \beta_4 IE_i + \beta_5 MI_i + \beta_6 MI_i^2 + \sum_j \beta_j Y_{i,j} + \sum_k \beta_k JC_{i,k} + \varepsilon_i, \quad [3]$$

where P_i represents the productivity of team *i* measured by revenue per person day of labor input, S_i represents the size of team *i* – the number of team members, A_i represents the average age of the team, E_i represents the average educational attainment of the team in years, IE_i represents the average number of years of industry experience of the team, MI_i represents the information overlap or mutual information of the team, MI_i^2 represents mutual information squared, $\sum_j \beta_j Y_{i,j}$ represents year controls and $\sum_k \beta_k JC_{i,k}$ represents job class controls for difficulty differences across projects created by the varying difficulty of filling different types of positions. We estimated team productivity using ordinary least squares regression. We could not reject the hypothesis of no heteroskedasticity using a Durbin-Watson test, and therefore report standard errors according to the White correction (White 1980). As project analyses may cluster on groups of project team members, we report robust standard errors clustered by project team in project-level analyses.[16] In the next section we detail the results of these analyses.

Results

We first examined the antecedents of mutual information by analyzing the characteristics of teams that predict greater information distance between recruiters as measured by topics found in email

[16] Clustered robust standard errors treat each project team as a super-observation for part of its contribution to the variance estimate (e.g. $\varepsilon_{ci} = \eta_c + \upsilon_{ci}$, where η_c is a group effect and υ_{ci} the idiosyncratic error). They are robust to correlations within the observations of each group, but are never fully efficient. They represent conservative estimates of standard errors as team members, such that teams with similar composition expend independently varying levels of effort across projects.

communication. Since communicated information is an essential part of common knowledge, text analysis of email content provides a precise view of the degree to which information is shared among team members in topics of conversation. The results of multiple regression quadratic assignment procedure (MRQAP) analyses are shown in Table 2.

Table 2. Predicting Mutual Information in Teams – Antecedents

Dependent Variable	***Cosine Based Information Distance Between Team Members***		
Model:	**1**	**2**	**3**
Specification	**MRQAP**	**MRQAP**	**MRQAP**
Demographic Distance			
Same Gender	-.031	-.034	-.021
Age Difference	.034	.037	.011
Education Difference	-.019	-.029	-.042
Expertise Distance			
Industry Experience Difference		.023	.009
Organizational Distance		-.044	-.052
Project Co-Work		-.143***	-.040
Geographic Dispersion			
Same Region		-.268***	-.183***
Same Office		-.295***	-.272***
Social Network Distance			
Network Distance			.179***
Contacts in Common			-.155*
Tie Strength			-.212***
Adjusted R^2	.002	.246	.388
Observations	3080	3080	3080

Notes: Multiple Regression Quadratic Assignment Procedure Estimation with 10,000 Random Permutations. * p < .10, ** p < .05, *** p < .01.

Our initial analysis finds that basic demography does not predict information distance with any confidence. Model 1 demonstrates that in our firm, age differences, gender differences and education differences do not explain variation in information distance. However, when expertise differences are introduced in Model 2, we find strong evidence that prior project co-work – the degree to which team members of a given project have worked on the same projects in the past – is a strong predictor of lower information distance in teams (β = -.143, p < .01). Having a prior working history entails shared common experience and greater overlap in the tasks and substance of prior projects. Having worked on projects together in the past, team members share learning experiences and retain similar knowledge about the candidate pool, contacts at client organizations, and qualitative details about the personalities and qualifications of candidates as well as the requirements of client firms. This detailed knowledge can then be reused in future projects and team members with shared prior project experiences bring similar information to bear on current and future projects as reflected in the relative similarity of the information they share in email communication.

Model 2 also demonstrates that geographic dispersion is a strong predictor of mutual information in teams. Teams whose members work in the same office (β = -.295, p < .01) or in the same region (β = -.268, p < .01), display much lower information distance than teams whose members are spread across different offices in the same region or in offices that are located in different regions. These results reflect the effects of both dispersion and shared local knowledge on mutual information. Recruiters who work in the same region are more familiar with the clients and the candidate pool available in that region. While recruiters work with clients and candidates outside of their region, candidates move more rarely across cities and regions than they do across firms in the same city or region. Recruiters therefore develop local knowledge and expertise about the labor conditions, clients, candidates and norms of job mobility in a given city or region, and team members from the same city or region share this local knowledge and expertise which is reflected in the similarity of information they share in email. For example, one recruiter told us in an interview that knowledge of local markets was an important aspect of the job, commenting about a potential candidate that "Mary has two inch nails, which doesn't present well in the LA market." At the same time, prior research has shown that collocation is important for the communication processes that help teams develop mutual information and knowledge (e.g. Tyre & von Hippel 1997 Schober 1998, Mortenson & Hinds 2001, Hinds & Bailey 2003), supporting our hypothesis that dispersion should predict lower information overlap in teams.

The results in Model 3 demonstrate that the social networks of team members also predict mutual information. Recruiters with more contacts in common (β = .179, p < .01), shorter path lengths to each other (β = -.155, p < .01), and stronger ties (β = -.212, p < .01) share more mutual information in email. Network distance measures the shortest geodesic path length between two recruiters, measuring the fewest number of contacts that a recruiter must go through to contact another recruiter. Recruiters who are closer together in the firm's communication network share more mutual information. In addition, recruiters with strong direct ties to each other also share more information in common. Controlling for tie strength and network distance, cohesion around a dyad also contributes to greater shared mutual information. Recruiters with more shared contacts in common also share more mutual information and have lower information distance in email. Prior evidence demonstrates that network cohesion around a given relationship contributes to greater knowledge transfer between contacts (Reagans & McEvily 2003), making it more likely that recruiters with shared contacts share common mutual information. These results demonstrate that social networks are a strong predictor of mutual information. They also show that social network characteristics are more salient in predicting information distance than shared prior project experience. While shared project experience predicts mutual information, strong connections and closeness in the flow of information in communication networks is more important in predicting shared information. We suspect that this is because we measure information shared in communication, which is an important aspect of the common ground between team members.

In summary, geographic dispersion and social network distance are strong predictors of mutual knowledge failures, while demographic dissimilarity and organizational distance do not predict the degree of mutual information between employees of this firm.

Performance Effects of Mutual Information

Having investigated the antecedents of mutual information and knowledge, we turned our attention to the productivity and performance effects of mutual information. Prior research could not detect any performance effects of mutual information (Cramton 2001), and we speculated that a reason was the non-linear nature of the relationship. Mutual information in teams has both costs and benefits. On one hand, mutual information establishes common ground and enables effective communication and collaboration. On the other hand, information diversity enables creative problem solving, learning, productivity and performance through the combination of the unique marginal contributions of team members. We therefore hypothesize that an inverted-U shaped relationship exists between mutual information and team productivity.

We measured team productivity by the revenues generated per person day of labor input. This classic measure of productivity is the most appropriate in our context because as teams use their shared and unique information to fill client positions, the speed with which they accomplish their goals and the number of person days they devote to a search are affected by how much relevant information they have at their finger tips and to what degree, in contrast, they must spend time, effort and energy researching candidate options, capturing client requirements or vetting and negotiating with candidates and clients. We present the results of our analysis, which control for variation in output across different types of projects (measured by their job class) and temporal variation in demand or workload in different years in Table 3.

The results in Table 3 demonstrate strong evidence of an inverted-U shaped relationship between mutual information and team productivity. A healthy amount of information overlap among team members contributes to performance while too little or too much mutual information hampers performance. This result helps resolve the apparent tension between arguments about the costs and benefits of mutual information in teams. Model 1 demonstrates that analysis of a purely linear relationship between mutual information and performance does not yield useful results – the coefficient on mutual information is not significant. This helps explain why previous research may have been unsuccessful in detecting the performance effects of mutual knowledge. However, when a quadratic relationship is tested, we find strong evidence of a non-linear relationship (see Model 2).

Table 3: Predicting Team Productivity as Function of Mutual Information

Dependent Variable	***Revenues Per Person Day: (Revenue / Labor Days Input)***			
Model:	**1**	**2**	**3**	**4**
Specification	**OLS**	**OLS**	**OLS**	**OLS**
Team Size	-240.05*** (36.96)	-242.26*** (37.19)	-244.80*** (37.16)	-254.12*** (37.57)
Age	-6.83 (4.63)	-7.82* (4.58)	-8.07* (4.37)	-6.77 (4.74)
Education	14.93 (29.34)	9.49 (28.36)	2.24 (29.70)	12.93 (29.35)
Industry Experience	56.89* (30.17)	54.63* (31.38)	59.22* (30.52)	55.02* (33.51)
Mutual Information	-3.63 (29.94)	253.75*** (58.52)		
Mutual Information Squared		-69.42*** (16.38)		
Mutual Information Received			205.08*** (56.98)	
Mutual Information Received Squared			-49.87*** (15.10)	
Mutual Information Sent				61.05 (63.38)
Mutual Information Sent Squared				-23.59 (17.82)
Controls	Job Class, Year	Job Class, Year	Job Class, Year	Job Class, Year
F-Value (d.f.)	11.69*** (16)	10.59*** (17)	11.66*** (17)	10.02*** (17)
R^2	.16	.18	.18	.17
Observations	689	689	689	689

Note: Ordinary Least Squares Estimation with Clustered Robust Standard Errors.
* p < .10, ** p < .05, *** p < .01

The results in Model 2 test the relationship between team productivity and total mutual information combining both incoming and outgoing email communication. The results demonstrate a clear non-linear relationship between mutual information and productivity. The most productive teams are within one standard deviation of the average information overlap observed across all teams. Teams with much greater or less information overlap generate less revenue per person day of labor input than their counterparts. These results suggest that the costs and benefits of mutual information combine to create a non-linear relationship. Too much mutual information can reduce the benefits from the unique contributions of each team member who bring local knowledge and expertise, and diverse perspectives, creativity and problem solving skills to a team. On the other hand, too little mutual information can make communication and collaboration difficult as team members do not share common ground, which establishes and supports mutual understanding and reduces conflict.

These results help resolve the apparent tension between arguments about the costs and benefits of mutual information and knowledge in teams and empirically demonstrate a relationship with productivity and performance. Figure 2 shows the non-linear nature of this relationship. The graph on the left shows the scatter plot of normalized mutual information on the x-axis and revenue per person day of labor input on the y-axis, while the graph on the right shows the fitted values of the quadratic parameter estimates of the relationship between mutual information and productivity. These graphs suggest that teams with mutual information two standard deviations greater or less than average tend to be significantly less productive.

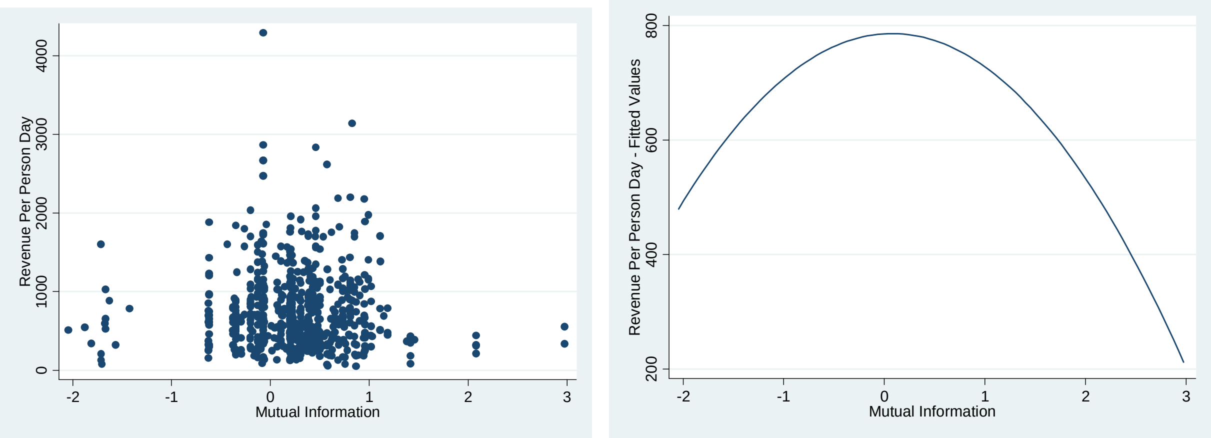


Figure 3: The relationship between mutual information and team productivity.

The results also suggest that greater industry experience is associated with higher productivity on average ($\beta = 54.63$, $p < .10$, Model 2), while older teams, controlling for industry experience and education, appear to be less productive on average ($\beta = - 7.82$, $p < .10$, Model 2), although both of these results are only marginally significant.

In Models 3 and 4 we test whether sent or received information is more salient in explaining the relationship between mutual information and team productivity. Some arguments suggest that shared information received should matter more than shared information sent. An answer-update argument implies that news received from a colleague should matter more than news sent to a colleague. If the project needing an update or the question needing an answer is unique to the recipient, then performance of the recipient should improve more than that of the person taking the time to provide the update or answer. By this thesis, receiving information allows the recipient to perform work. In contrast, a delegation argument implies that shared information benefits senders more than recipients. Individuals commonly describe tasks they want their colleagues or subordinates to complete. By this thesis, sending allows the sender to accomplish work. Alternatively, if knowledge is deeply tacit or complex, a sender may be unable to pass complete knowledge so that a recipient remains at a deficit with respect to the sharing source. By this thesis, sending information signals superior expertise.

Models 3 and 4 show that coefficients have correct signs for arguments favoring senders and receivers, but only mutual information received is statistically significant. We suspect that overlap in the information team members receive is more relevant than that which they send. Novel information shared by a colleague is likely received on the occasion of a specific need. Thus it is the arrival of an opportunity to perform work rather than to off-load it or signal one's expertise that explains why shared information enhances productivity – it meets a specific need. If true, one organizational implication is to manage incentives to improve willingness to share information in response to teammates' requests for information. Furthermore, received information is a proxy for what team members are aware of in their information environments. When team members receive similar information they become jointly aware of project characteristics and progress, and aspects of the environment in which they work. Mutual or joint awareness is the cornerstone of theories of mutual knowledge and common ground (Krauss & Fussell 1990, Clark 1996, Cramton 2001). Common ground and joint awareness can be established if team members are aware of the same information (that which is received) even if they communicate divergent information to others, having reflected on a common set of information that team members know and know they know.

We therefore expect to see a strong relationship between mutual information and productivity when analyzing incoming email, but less of a relationship when analyzing outgoing email. Models 3 and 4 appear to confirm these expectations. Although the results for outgoing mail are in the same direction as for incoming mail, the combined results demonstrate that most of the relationship between total mutual information and productivity can be explained by the overlap in the information received by team members.

Discussion and Conclusion

In this paper, we present some of the first large scale empirical evidence on the antecedents and consequences of mutual knowledge in teams and their implications for productivity. We developed a unique data set of 1382 executive recruiting teams, 125,000 email messages, and 5 years of project performance data. Using these data, we find strong evidence of an inverted-U shaped relationship between mutual information and team performance. This evidence helps resolve the apparent tension in prior literature between research emphasizing the benefits of mutual knowledge and research emphasizing the benefits of diversity.

We also developed a vector space model of mutual information in communication. We constructed topic vectors of email conversation and compared the distance of vectors in recruiters' inboxes and out-boxes to estimate the distance between recruiters' incoming and outgoing information. Using multiple regression quadratic assignment procedure estimation (MRQAP), we found clear evidence for various long-hypothesized effects on mutual knowledge such as geography, social network distance, demographics, and organizational distance.

Our results complement and extend prior work that laid a foundation for organizational research on mutual knowledge through detailed qualitative case studies (e.g. Cramton 2001), analysis of geographic dispersion and collaboration (e.g. Armstrong & Cole 1995, Mortenson & Hinds 2001, Hinds & Bailey 2003), and hypothesis testing of the effects of social network diversity and cohesion on team performance (e.g. Reagans and Zuckerman 2001, Aral et. al. 2006, Aral & Van Alstyne 2007).

Our results demonstrate that geographic and social network distance strongly predict mutual knowledge failures, while demographic and organizational distance do not predict the degree of mutual knowledge among team members. While project co-work weakly predicts greater mutual information, this prediction fails when social network and geographic dispersion variables enter into the analysis. Geographic dispersion and social networks are the two most salient characteristics of teams that predict mutual information in our setting. Mutual information also predicts performance but in our data, we find an single-peaked optimum – extremes in either overlap or diversity correspond with lower productivity.

Our findings suggest that managers may be able to calibrate optimal information overlap among team members by considering geographic dispersion and social networks while placing less emphasis on demographic and organizational distance. These findings contribute to academic and managerial interest in diversity, mutual knowledge, and team performance in organizations.

Acknowledgments

We are grateful to the National Science Foundation (Career Award IIS-9876233 and grant IIS-0085725), Cisco Systems Inc., and the MIT Center for Digital Business for generous funding. We thank Petch Manoharn and Hank Huang for their research assistance.

References

Allen, T. 1984. Managing the Flow of Technology. Cambridge, Mass.: 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.

Aral, S. & M. Van Alstyne. 2007. "Network Structure & Information Advantage" Proceedings of the Academy of Management Conference, Philadelphia, PA.

Aral, S., Brynjolfsson, E., & Van Alstyne, M. 2007. "Productivity Effects of Information Diffusion in Networks." Proceedings of the 28th Annual International Conference on Information Systems, Montreal, CA.

Armstrong, D., & P. Cole. 1995. "Managing distances and differences in geographically distributed work groups." S. Jackson, M. Ruderman, eds. *Diversity in Work Teams*. American Psychological Association, Washington, DC: 187 – 216.

Blau, P, 1977. Inequality and Heterogeneity. New York: Free Press.

Borgatti, S.P., & R. Cross. 2003. "A relational view of information seeking and learning in social networks." Management Science, 49(4), 432-445.

Burgelman, R.A. 1991. "Intraorganizational Ecology of Strategy Making & Organizational Adaptation: Theory & Field Research." Organization Science, (2:3):239-261.

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

Burt, R. 2004. "Where to get a good idea: Steal it outside your group." As quoted by Michael Erard in The New York Times, May.

Carley, K.M., & D. Krackhardt. 1996. "Cognitive inconsistencies and non-symmetric friendship." Social Networks, 18, 1-27.

Clark, H. 1996. *Using Language*. Cambridge University Press, New York.

Cohen, W.M. & D.A. Levinthal. 1990. "Absorptive Capacity: A New Perspective on Learning & Innovation." Administrative Science Quarterly (35:1): 128-152.

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

Contractor, N. 2000. Social Network Formulations of Knowledge and Distributed Intelligence: Using Computational Models to Extend and Integrate Theories of Transactive Memory and Public Goods. Heterarchies: Distributed Intelligence and the Organization of Diversity Project. Santa Fe, New Mexico: Santa Fe Institute.

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

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

Dennis, A. 1996. "Information exchange and use in small group decision making." Small Group Research, 27(4), 532-549.

Dodds, P. S., R. Muhamad and D. J. Watts."An Experimental Study of Search in Global Social Networks," by Science, Aug 8 2003.

Eisenhardt, K. M., Kahwajy, J. L., Bourgeois, L. J. 1997. "Conflict and strategic choice: How top management teams disagree. California Management Review, 39(2), 42-62.

Finlay, W. & Coverdill, J.E. 2000. "Risk, Opportunism & Structural Holes: How headhunters manage clients and earn fees." Work & Occupations, (27): 377-405.

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

Granovetter, M. 1985. "Economic Action & Social Structure: The Problem of Embeddedness." American Journal of Sociology (91):1420-1443.

Grinter, R. E., J. D., Herbsleb, D. E., Perry. 1999. The geography of coordination: Dealing with distance in R&D work. SIGGROUP Conference, Phoenix, AZ.

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.

Hinds, P. J., & D. E. Bailey. 2003. “Understanding conflict in distributed teams.” Organization Science, 14(6), 615-632.

Hinds, P. J., K. M. Carley, D. Krackhardt & D. Wholey. 2000. “Choosing work group members: Balancing similarity, competence, and familiarity. Organizational Behavior and Human Decision Processes, 81 (2), 226-251.

Hong, L. & S. Page 2001. "Problem Solving by Heterogeneous Agents," Journal of Economic Theory. 97(1), 123-163.

Huber, G. 1982. Organizational information systems: Determinants of their performance and behavior. Management Sci. 28(2) 135–155.

Huber, G., R. Daft. 1987. The information environments of organizations. F. Jablin, L. L. Putnam, K. H. Roberts, L. W. Porter, eds. Handbook of Org. Comm., Sage, Beverly Hills, CA, 130–163.

Hubert, L. J. 1987. *Assignment methods in combinatorial data analysis*. Dekker. New York.

Jackson, J. 1965. “Structural characteristics of norms.” D. Steiner, M. Fishbein, eds. Current Studies in Social Psychology. Holt, Rinehart and Winston, New York.

Jackson, S. 1992. “Team composition in organizations.” S. Worchel, W. Wood, and J. Simpson eds., Group Process and Productivity, 1-12. Sage Publications, London.

Jehn, K. A., G. B. Northcraft, & M. A. Neale. 1999. “Why differences make a difference: A field study of diversity, conflict and performance in workgroups.” Administrative Sciences Quarterly, 44(4), 741-763.

Judge, G. G., Griffitsh, W.E., Carter Hill, R., Lutkepohl, H., & T. Lee. 1990. The theory and practice of econometrics. New York: Wiley.

Krackhardt, D. 1988. Predicting with networks: Nonparametric multiple regression analysis of dyadic data. Social Networks, 10, 359-381.

Krackhardt, D. 1993. “MRQAP: Analytic versus permutation solutions.” Working paper, Carnegie Mellon University.

Krauss, S. & S. Fussell. 1990. Mutual knowledge and communication effectiveness. J. Galegher, R. Kraut, C. Egido, Eds. Intellectual Teamwork: Social and Technological Foundations of Cooperative Work. Lawrence Erlbaum, Hillsdale, NJ, 111-146.

Kraut, R. E., S. R., Fussel, S. E., Brennan & J. Seigel. 2002. “Understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work.” P.J. Hinds, S. Kiesler, eds. *Distributed Work*. MIT Press, Cambridge, MA, 137-162.

March, J. G. 1991 "Exporation and Exploitation in Organizational Learning," Organization Science; 2(1) p71-87.

March, J. G., H. A. Simon. 1958. Organizations. Wiley, New York.

McGrath, J. E., J. L. Berdahl & H. Arrow. 1996. “No one has it but all groups do: Diversity as a collective, complex, and dynamic property of groups.” S. E. Jackson & M. N. Ruderman eds, *Diversity in Work Teams: Research Paradigms for a Changing World*, 42-66. APA Publications, Washington DC.

Mortenson, M., & P.J. Hinds. 2001. “Conflict and shared identity in geographically distributed teams.” International Journal of Conflict Management, 12(3), 212-238.

Padgett, J.F., & C.K. Ansell. 1993. “Robust Action & the Rise of the Medici.” American Journal of Sociology, (98:6): 1259-1319.

Page, S. 2007 The Difference – How the Power of Diversity Creates Better Groups, Firms, Schools & Societies, Princeton University Press.

Pelled, L. H. 1996. Demographic diversity, conflict, and work group outcomes: An intervening process theory. Organization Science, 11, 404-428.

Pfeffer, J. 1983. Organizational demography: Implications for management. California Management Review, 28, 67-81.

Portes, A. & J. Sensenbrenner. 1993. Embeddedness and immigration: Notes on the social determinants of economic action. American Journal of Sociology, 98, 1320-1350.

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.

Rodan, S. & D. Galunic. 2004. “More Than Network Structure: How Knowledge Heterogeneity Influences Managerial Performance & Innovativeness.” Strategic Management Journal (25): 541-562.

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

Schober, M. F. 1998. Different kinds of perspective taking.” S. Fussell, R. Krauss, eds. *Social and Cognitive Approaches to Interpersonal Communication*. Lawrence Erlbaum, Mahwah, NJ.

Simmel, G. (1922) 1955. Conflict & the Web of Group Affiliation. Free Press. New York, NY.

Simon, H. 1991. “Bounded Rationality & Organizational Learning.” Organization Science. (2:1): 125-134.

Stasser, G., & D. Stuart. 1992. “The discovery of hidden profiles by decision-making groups: Solving a problem versus making a judgment.” Journal Personality and Social Psychology, 63, 426-434.

Stasser, G., D. Stewart, & G. Wittenbaum. 1995. “Expert roles and information exchange during discussion: The importance of knowing who knows what.” Journal of Experimental Social Psychology, 31, 244-265.

Stasser, G., & W. Titus. 1985. “Pooling unshared information in group decision making: Biased information sampling during discussion.” Journal Personality and Social Psychology, 48, 1467-1478.

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

Tajfel, H., & J. Turner. 1986. “The social identity of intergroup behavior.” Stephen Worchel, William G. Austin, eds. Psychology and Intergroup Relations. Nelson-Hall, Chicago, IL. 7-24.

Tyre, M. E., & E. von Hippel. 1997. The situated nature of learning in organizations. Organization Science, 8, 71-83.

Van Alstyne, M. “The State of Network Organization: A Survey in Three Frameworks.” Journal of Organizational Computing, 1997, Vol. 7 N.2& 3. pp. 83-152.

Van Alstyne, M. & Brynjolfsson, E. “Could the Internet Balkanize Science?” Science. 274(5292), November 29, 1996, pp. 1479-1480.

Van Alstyne, M. & Brynjolfsson, E “Global Village or CyberBalkans? Measuring and Modeling the Integration of Electronic Communities”. Management Science; 51 (6) (June 2005): pp. 851-868.

Van Alstyne, M. & Zhang, J. "EmailNet: A System for Mining Social Influence & Network Topology in Communication" NAACSOS. North American Association for Computational Social and Organization Sciences. Pittsburgh, PA: Jun 22-25, 2003.

Van de Van, A. H., A., Delbecq, & R. Koenig. 1976. "Determinants of coordination modes within organizations." American Sociological Review, 41, 322-338.

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

Williams, K. Y., & C. A. O'Reilly. 1998. "Demography and diversity in organizations." Barry M. Staw and Robert M. Sutton eds., *Research in Organizational Behavior*, 20, 77-140, JAI Press, Stamford, CT.

Online Appendix A. Descriptions & Correlations of Information Diversity Metrics

1. Cosine Distance Variance

Variance based on cosine distance (cosine similarity):

$$ID_i^I = \frac{\sum_{j=1}^{N}\left(Cos\left(d_{ij}^I, M_i^I\right)\right)^2}{N}, \text{ where } Cos(d_{ij}, M) = \frac{d_i \bullet M_i}{|d_i||M_i|} = \frac{\sum_j w_{ij} \times w_{Mj}}{\sqrt{\sum w_{ij}^2}\sqrt{\sum w_{Mj}^2}}$$

We measure the variance of deviation of email topic vectors from the mean topics vector and average the deviation across emails in a given inbox or outbox. The distance measurement is derived from a well-known document similarity measure – the cosine similarity of two topic vectors.

2. Dice's Coefficient Variance

Variance based on Dice's Distance and Dice's Coefficient: $VarDice_i^I = \frac{\sum_{j=1}^{N}\left(DistDice\left(d_{ij}^I\right)\right)^2}{N}$, where

$$DistDice(d) = DiceDist(d, M) = 1 - Dice(d, M), \text{ and where}$$

$$Dice(D1, D2) = \frac{2\sum_{i=1}^{T}(t_{D1j} \times t_{D2j})}{\sum_{i=1}^{T} t_{D1j} + \sum_{i=1}^{T} t_{D2j}}$$

Similar to VarCos, variance is used to reflect the deviation of the topic vectors from the mean topic vector. Dice's coefficient is used as an alternative measure of the similarity of two email topic vectors.

3. Average Common Cluster

AvgCommon measures the level to which the documents in the document set reside in different k-means clusters produced by the eClassifier algorithm:

$$AvgCommon_i^I = \frac{\sum_{j=1}^{N}\left(CommonDist\left(d_{1j}^I, d_{2j}^I\right)\right)}{N},$$

where (d_{1j}^I, d_{2j}^I) represents a given pair of documents (1 and 2) in an inbox and j indexes all pairs of documents in an inbox, and where:

$$CommonDist\left(d_{1j}^{I}, d_{2j}^{I}\right) = 1 - CommonSim\left(d_{1j}^{I}, d_{2j}^{I}\right)$$

$$CommonSim\left(d_{1j}^{I}, d_{2j}^{I}\right) = \frac{\sum Iterations\_in\_same\_cluster}{\sum Iterations}$$

AvgCommon is derived from the concept that documents are similar if they are clustered together by k-means clustering and dissimilar if they are not clustered together. The k-means clustering procedure is repeated several times, creating several clustering results with 5, 10, 20, 30, 40 … 200 clusters. This measures counts the number of times during this iterative process two emails were clustered together divided by the number of clustering iterations. Therefore, every two emails in an inbox and outbox that are placed in separate clusters contribute to higher diversity values.

4. Average Common Cluster with Information Content

AvgCommonIC uses a measure of the "information content" of a cluster to weight in which different emails reside. AvgCommonIC extends the AvgCommon concept by compensating for the different amount of information provided in the fact that an email resides in the same bucket for either highly diverse or tightly clustered clusters. For example, the fact that two emails are both in a cluster with low intra-cluster diversity is likely to imply more similarity between the two emails than the fact that two emails reside in a cluster with high intra-cluster diversity.

$$CommonICSim(D_1, D_2) = \frac{1}{\log\left(\frac{1}{\|all\_documents\|}\right)} \cdot \frac{\sum_{D_1, D_2 in\_same\_bucket} \log\left(\frac{\|documents\_in\_the\_bucket\|}{\|all\_documents\|}\right)}{total\_number\_of\_bucket\_levels}$$

$$CommonICDist(D_1, D_2) = 1 - CommonICSim(D_1, D_2)$$

$$AvgCommonIC = \underset{d_1, d_2 \in documents}{average} \left\{CommonICDist(d_1, d_2)\right\}$$

5. Average Cluster Distance

AvgBucDiff measures diversity using the similarity/distance between the clusters that contain the emails:

$$AvgBucDiff = \underset{d_1, d_2 \in documents}{average} \left\{DocBucDist(d_1, d_2)\right\}, \text{ where}$$

$$DocBucketD\ ist(D_1, D_2) = \frac{1}{\|cluster\_iterations\|} \cdot \sum_{i \in cluster\_iterations} \left(BucketDist\ (B_{iteration=i, D_1}, B_{iteration=i, D_2})\right), \text{ and:}$$

$$BucketDist(B_1, B_2) = CosDist(m_{B_1}, m_{B_2}).$$

AvgBucDiff extends the concept of AvgCommon by using the similarity/distance between clusters. While AvgCommon only differentiates whether two emails are in the same cluster, AvgBucDiff also considers the distance between the clusters that contain the emails.

Table A1. Correlations Between the Five Measures of Information Diversity

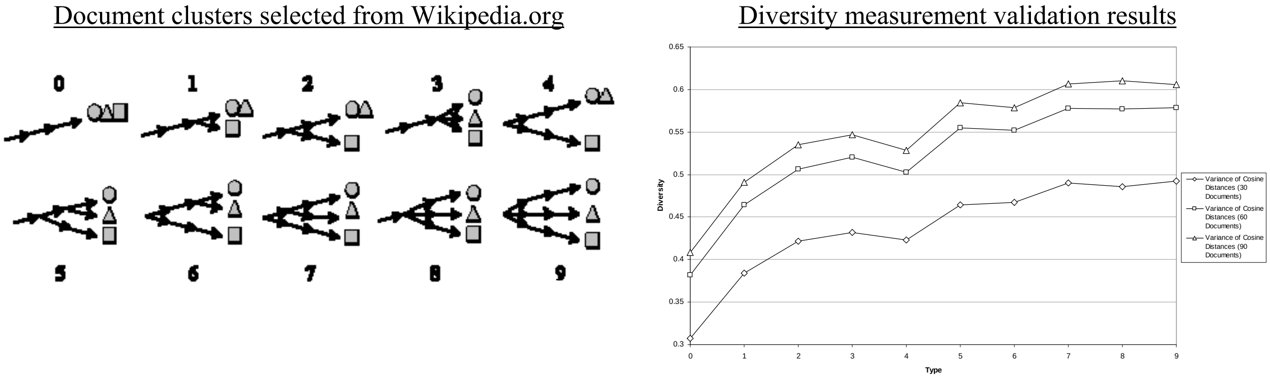
Measure	1	2	3	4	5
1. VarCosSim	1.0000				
2. VarDiceSim	0.9999	1.0000			
3. AvgCommon	0.9855	0.9845	1.0000		
4. AvgCommonIC	0.9943	0.9937	0.9973	1.0000	
5. AvgClusterDist	0.9790	0.9778	0.9993	0.9939	1.0000

Online Appendix B: External Validation of Information Overlap Measures

We validated our information overlap measurement strategy using an independent, publicly available corpus of documents from Wikipedia.org. Wikipedia.org, the user created online encyclopedia, stores entries according to a hierarchy of topics representing successively fine-grained classifications. For example, the page describing "genetic algorithms," is assigned to the "Genetic Algorithms" category, found under "Evolutionary Algorithms," "Machine Learning," "Artificial Intelligence," and subsequently under "Technology and Applied Sciences." This hierarchical structure enables us to construct clusters of entries on diverse and overlapping subjects and to test whether our measurement can successfully characterize diverse and overlapping clusters accurately.

We created a range of high to low diversity clusters of Wikipedia entries by selecting entries from either the same sub-category in the topic hierarchy to create overlapping clusters, or from a diverse set of unrelated subtopics to create diverse clusters. For example, we created a maximum overlap cluster (Type-0) using a fixed number of documents from the same third level sub-category of the topic hierarchy, and a maximum diversity cluster (Type-9) using documents from unrelated third level sub-categories. We then constructed a series of document clusters (Type-0 to Type-9) ranging from low to high topic diversity from 291 individual entries as shown in Figure 3.[17] The topic hierarchy from which documents were selected appears at the end of this section.

If our measurement is robust, our diversity measures should identify Type-0 clusters as the least diverse and Type-9 clusters as the most diverse. We expect diversity will increase relatively monotonically from Type-0 to Type-9 clusters, although there could be debate for example about whether Type-4 clusters are more diverse than Type-3 clusters.[18] After creating this independent dataset, we used the Wikipedia entries to generate keywords and measure diversity using the methods described above. Our methods were very successful in characterizing diversity and ranking clusters from low to high diversity. Figure 3 displays cosine similarity metrics for Type-0 to Type-9 clusters using 30, 60, and 90 documents to populate clusters. All five diversity measures return increasing diversity scores for clusters selected from successively more diverse topics.[19] Overall, these results give us confidence in the ability of our diversity measurement to characterize the subject diversity of groups of text documents of varying sizes.


Document clusters selected from Wikipedia.org

Diversity measurement validation results

Figure B1. Wikipedia.org Document Clusters and Diversity Measurement Validation Results.

[17] We created several sets of clusters for each type and averaged diversity scores for clusters of like type. We repeated the process using 3, 6 and 9 document samples per cluster type to control for the effects of the number of documents on diversity measures.

[18] Whether Type-3 or Type-4 clusters are more diverse depends on whether the similarity of two documents in the same third level sub category is greater or less than the difference of similarities between documents in the same second level sub category as compared to documents in categories from the first hierarchical layer onwards. This is, to some extent, an empirical question.

[19] The measures produce remarkably consistent diversity scores for each cluster type and the diversity scores increase relatively monotonically from Type-0 to Type-9 clusters. The diversity measures are not monotonically increasing for all successive sets, such as Type-4, and it is likely that the information contained in Type-4 clusters are less diverse than Type-3 clusters due simply to the fact that two Type-4 documents are taken from the same third level sub category.

Wikipedia.org Categories:

+ **Computer science >**
+ Artificial intelligence
    + Machine learning
    + Natural language processing
    + Computer vision
+ Cryptography
    + Theory of cryptography
    + Cryptographic algorithms
    + Cryptographic protocols
+ Computer graphics
    + 3D computer graphics
    + Image processing
    + Graphics cards

+ **Geography >**
+ Climate
    + Climate change
    + History of climate
    + Climate forcing
+ Cartography
    + Maps
    + Atlases
    + Navigation
+ Exploration
    + Space exploration
    + Exploration of Australia

+ **Technology >**
+ Robotics
    + Robots
    + Robotics competitions
+ Engineering
    + Electrical engineering
    + Bioengineering
    + Chemical engineering
+ Video and movie technology
    + Display technology
    + Video codecs
    + Digital photography