Unpacking Novelty: The Anatomy of Vision Advantages

Sinan Aral & Paramveer S. Dhillon

{sinan|dhillon}@mit.edu

Sloan School of Management

Massachusetts Institute of Technology

Abstract

The Strength of Weak Ties and Brokerage Theory rely on the argument that weak bridging ties deliver novel information to brokers. Yet our conceptualization of novelty itself is fundamentally underdeveloped. We therefore develop a theory of how three distinct types of novelty – diversity, total non-redundancy and uniqueness – combine with network structure to create vision advantages. We test our theory using panel data on an evolving corporate email network in a medium-sized digital media firm over twelve months. Three main results emerge from our analysis. First, we confirm the Diversity-Bandwidth Tradeoff at the heart of the vision advantage. As brokers networks become more diverse, their channel bandwidth contracts, creating countervailing effects on access to novel information. Second, we uncover the mechanisms driving the Diversity-Bandwidth Tradeoff and highlight differences in vision advantages across ties. Strong cohesive ties deliver greater information diversity and more total non-redundant information, while weak bridging ties contribute the greatest uniqueness – information which is most different from what other contacts deliver. Third, we find network diversity (or inversely, network constraint) is the dominant factor in the relationship between network structure and longitudinal entropy. This result suggests that weak bridging ties, which provide unique information through low bandwidth, structurally diverse channels, contribute the most to the aggregation of novel information over time.

Key words: information worker productivity, networks, communication, information exchange, information channels, information flow, knowledge transfer, content analysis, email data.

1

1 Introduction

For the last forty years, researchers in disciplines as diverse as economics, sociology, management, marketing and information systems have been developing an important line of social theory. The Strength of Weak Ties and Brokerage Theory have together underpinned tens of thousands of empirical investigations linking network structure to outcomes including wages, job placement, promotion, creativity, innovation, political success, social support, productivity, and performance (Aral et al., 2012, 2007; Baker, 1990; Bulkley and Van Alstyne, 2004; Burt, 2009, 2004; Granovetter, 1973; Hansen, 1999, 2002; Padgett and Ansell, 1993; Podolny, 2001; Reagans and Zuckerman, 2001; Uzzi and Spiro, 2005; Uzzi, 1997). Both of these theories rely on the argument that weak bridging ties deliver novel information to actors in brokerage positions. Burt calls this the broker's "vision advantage." Yet, little empirical evidence has been presented to validate the existence of vision advantages or to document how they work. Perhaps more importantly, our conceptualization of novelty itself is underdeveloped. What is novel information? Although many theories currently rely on a vague notion of 'access to novel information' to explain individual and group outcomes, the precision with which we treat novel information, both theoretically and empirically, is underwhelming. For example, consider the following question: Which of these two messages contains more novel information? 1) A message with two bits of information you have never seen before but that are very closely related to the information you have seen earlier, or alternatively, 2) a message with one bit of information that is very different that any information you have ever seen before? Surely, the volume of new information contained in the messages is important. The first message has two bits of 'novel information' compared to the second message, which has only one. On the other hand, however, the relative distance of the bit of information in the second message to the information you have already seen should also matter to which message contains more 'novelty'. Current social theories of novelty are too imprecise to distinguish or differentiate the novelty in these two messages. We argue that this ambiguity in our conceptualizations of novelty currently impedes our ability to explain outcomes of consequence in network studies.

The key difficulty is in theorizing about, observing and subsequently measuring the novelty of the information content delivered by networked actors. As Burt (Burt, 2008) notes, "Empirical

success in predicting performance with network models has far outstripped our understanding of the way information flow in networks is responsible for network effects. A cluster of network concepts emerged in the 1970s on the idea that advantage results from connections with multiple, otherwise disconnected, groups and individuals. The hubs in a social network were argued to have advantaged access to information and control over its distribution... However, the substance of advantage, information, is almost never observed." According to Burt (Burt, 2005), "The next phase of work is to understand the information-arbitrage mechanisms by which people harvest the value buried in structural holes... More generally, the sociology of information will be central in the work."

Using detailed analysis of the content and structure of an evolving corporate email network over twelve months, we seek to test the vision advantage argument and to investigate the underlying dynamic mechanisms that enable vision advantages by unpacking and theorizing about the basic concept of novelty. We define and develop three new concepts that provide specificity to our theoretical conceptualizations of information novelty: information diversity, information uniqueness and total non-redundant information. We then theorize how network structure should affect access to these conceptually distinct dimensions of novelty and analyze how much novel information each actor in a broker's network delivers to the broker over time using vector space and information theoretic measures of novelty in email content. Temporal information structure plays a critical role, both in our theoretical conceptualization of novelty and in our empirical measures. We therefore also analyze how network dynamics affect the amount of novel information brokers receive, allowing us to extend theory about the types of 'information environments in which brokers receive more or less novel information (Aral and Van Alstyne, 2011).

Three results emerge from our analysis. First, we confirm the Diversity-Bandwidth Tradeoff at the heart of the vision advantage (Aral and Van Alstyne, 2011): As a broker's network becomes more diverse, the bandwidth of their communication channels contracts, creating countervailing effects on access to novel information. Second, our analysis uncovers the mechanics driving the Diversity-Bandwidth Tradeoff and highlights differences between the vision advantages offered by strong cohesive ties and weak bridging ties. As our theory predicts, strong cohesive ties deliver

3

greater information diversity and more total non-redundant information, while weak bridging ties contribute greater uniqueness - information which is the most different from what other contacts are delivering. Finally, our analysis of the networks evolution reveals that network stability (maintenance of the same contacts over time) increases the novelty brokers receive over time, providing some of the first evidence of the role of network dynamics in vision advantages.

The theory we propose and the results of our empirical analysis together represent the first steps toward a dynamic ego- and dyad- level model of the vision advantages that have for forty years been hypothesized to explain The Strength of Weak Ties and Brokerage Theory. Our work therefore makes three key contributions to these important lines of argument. First, we propose the first theoretical explanation for how vision advantages work namely, that weak bridging ties provide brokers with more unique information while strong cohesive ties provide more information diversity and more total non-redundant information. We develop each of these constructs theoretically and operationally, extending the work begun by Aral and Van Alstyne (2011). Second, we provide the first empirical evidence of the types of information weak bridging ties and strong cohesive ties deliver. Third, we find network diversity (or inversely, network constraint) is the dominant factor in the relationship between network structure and longitudinal entropy. This result suggests that weak bridging ties, which provide unique information through low bandwidth, structurally diverse channels, contribute the most to the aggregation of novel information over time. These contributions validate the information based mechanisms theorized to drive The Strength of Weak Ties and Brokerage Theory and serve to advance our understanding of the anatomy and dynamics of vision advantages.

2   Theory

Human social networks tend to cluster due to triadic closure. As Granovetter's (Granovetter, 1973) forbidden triad proposes, two strong tie contacts to a third party are themselves likely to be connected by a strong (or at least a weak) tie because they are more likely to meet, more likely to have similar preferences, and because their discord would inspire cognitive dissonance in the original strong tie friendships. The significant clustering that develops in human social networks as

a result of triadic closure creates small world networks with short global path lengths (Watts and Strogatz, 1998) and heavy tailed degree distributions (Barabási and Albert, 1999). Such structure – densely connected cliques connected by infrequent weak bridging ties – gives rise to opportunity. Brokers with structurally diverse networks, which lack cohesion and structural equivalence but are rich in structural holes, have privileged access to diverse, novel information. Contacts maintained through weak ties are typically unconnected to other contacts and therefore more likely to "move in circles different from our own and thus [to] have access to information different from that which we receive" (Granovetter, 1973). These ties are "the channels through which ideas, influence, or information socially distant from ego may reach him." As Burt (Burt, 2009) argues, "everything else constant, a large, diverse network is the best guarantee of having a contact present where useful information is aired." Since information in local network neighborhoods tends to be redundant, structurally diverse contacts that reach across structural holes should provide channels through which novel information flows (Burt, 2009). Novel information is thought to be valuable because of its local scarcity. Actors with scarce information in a given network neighborhood are better positioned to broker opportunities, make better decisions, and apply information to problems that are intractable given local knowledge (Van Alstyne and Brynjolfsson, 2005; Burt, 2004; Hargadon and Sutton, 1997; Lazer and Friedman, 2007; Reagans and Zuckerman, 2001; Rodan and Galunic, 2004). Access to novel information should increase the breadth of individuals absorptive capacity, strengthen the ability to communicate ideas across a broader range of topics to a broader audience, and improve persuasion and the ability to generate broader support from subject matter experts (Cohen and Levinthal, 1990; Reagans and Zuckerman, 2001; Rodan and Galunic, 2004). For these reasons, networks rich in structural diversity are thought to confer "information benefits" or "vision advantages" that improve performance by providing access to diverse and novel perspectives, ideas, and information (Burt, 2009).

The first paper to explore the information benefits to structural diversity uncovered a tradeoff between network diversity and channel bandwidth (Aral and Van Alstyne, 2011). In the executive recruiting firm they studied, Aral and Van Alstyne (2011) found that as an individual's ego network diversity increased, the bandwidth of their communication channels contracted, creating

countervailing effects on access to novel information. The theoretical arguments underpinning this 'Diversity-Bandwidth Tradeoff' highlighted unexplored aspects of the Strength of Weak Ties and Brokerage Theory. In particular, if bridging ties are by nature weak and infrequent, they are also likely to deliver less novel information per unit time on a smaller number of topical dimensions (for more detail on this theory see (Aral and Van Alstyne, 2011). These discoveries raised questions about exactly how novelty flows through network structure. However, it remains to be seen whether Aral and Van Alstyne (2011)'s findings generalize and whether their results can be replicated in other settings. We therefore begin by testing the hypotheses that underlie the Diversity-Bandwidth Tradeoff in order to evaluate their generality:

- Hypothesis 1a: Greater network diversity is associated with lower channel bandwidth.

- Hypothesis 1b: Network diversity and channel bandwidth are both associated with receiving more diverse information and more total non-redundant information.

Though the Diversity-Bandwidth Tradeoff begins to explain how vision advantages operate and how network structure and information flow are related, the mechanisms underlying the tradeoff are not well understood. The first wave of theory linking network diversity to novel information focused almost exclusively on the relative diversity of the information received across different alters in a network (Granovetter, 1973; Burt, 2009), overlooking the diversity and volume of novel information flowing within each tie or channel over time. Aral and Van Alstyne (2011) argued that although dense, cohesive networks tend to deliver information that is redundant across channels (with each alter providing the same or similar information), relationships in such networks are also typically stronger, implying greater frequency of interaction, richer information flows and thus access to more diversity and total novelty within each channel over time. This evidence raised new questions about whether vision advantages operated the way Granovetter and Burt had theorized. However, Aral and Van Alstyne (2011) only measured ego network level proxies for the information delivered over ties, using averages across all the ties in a network to make statements about the general tendencies of different types of ties to deliver different types of information.

We seek to reconcile the apparent tension in these theories with a simple unifying claim: strong embedded ties deliver greater information diversity and greater total non-redundant information,

but they do so at the expense of information uniqueness. We argue the information benefits to bridging ties are not in delivering greater information diversity or greater total non-redundant information, but rather in delivering information that is unique —information ego is unlikely to get from anyone else in their contact network. The only way to test this unifying claim is to examine tie level data that distinguishes the types of information delivered by strong embedded ties as compared to weak bridging ties. In this way our empirical analysis extends the literature on vision advantages and brokerage theory by examining information diversity, total non-redundant information and information uniqueness at the dyadic level.

Information diversity is a measure of the topical variance of a set of information. Total non-redundant information is a measure of the volume of novel (or unrepeated) bits in a set of information. In contrast, information uniqueness is a measure of the distance between one set of information and another (we develop these concepts in detail later in the paper). Strong cohesive ties are likely to provide a broker with greater information diversity and greater total non-redundant information because interaction through rich high-bandwidth channels tends to be more detailed, cover more topics, and address more complex, interdependent concepts over time (see (Aral and Van Alstyne, 2011) for a full development of this argument). On the other hand, though weak bridging ties are likely to provide less information diversity and less total novelty, they are more likely to provide unique information that ego does not receive from other contacts, because they are communicating in social circles that are the most distant from egos other contacts. We therefore hypothesize:

- Hypothesis 2a: Strong cohesive ties deliver more information diversity and more total non-redundant information than weak bridging ties.

- Hypothesis 2b: Weak bridging ties deliver more information uniqueness than strong cohesive ties.

Unpacking differences between diversity, non-redundancy and uniqueness adds a theoretical subtlety to the information advantage argument which could help reconcile conflicting evidence that has accumulated both for and against Brokerage Theory over the years. Some research has found that diverse networks are associated with innovation (Burt, 2005), while other work has found

7

the opposite - that cohesion promotes innovation (Obstfeld, 2005; Uzzi and Spiro, 2005). We believe that one possible explanation for these contradictory results is that in situations where uniqueness matters, structural diversity is more valuable and in situations where diversity or total non-redundancy matters, cohesion is more valuable.

The Strength of Weak Ties and Brokerage Theory have to date been static theories, foregrounding equilibrium states of networks and access to information rather than the dynamical processes that lead to those states. The majority of the work in this area has neglected the role of network and information dynamics in the creation of vision advantages and how changes in network or information structure over time affect the diversity, non-redundancy and uniqueness of the information brokers receive. Aral and Van Alstyne (2011) highlighted the importance of information dynamics when they considered the moderating effect of information turbulence on vision advantages. They found that as the refresh rate of alters information increased, brokers received more novel information and that channel bandwidth had an even stronger effect on the volume of novel information they received.

We embrace and extend this line of inquiry into dynamics by considering the role of network dynamics rather than the role of information dynamics. In particular, we focus on a theoretical concept that has recently attracted the attention of network and management scholars alike network stability.

Network stability describes the degree to which the composition of ones network is changing over time —the more one's contacts change from period to period, the more unstable the network. Past research has examined the role of network stability in the localization of network externalities (Tucker 2011) and mobile content generation and consumption (Ghose & Iyengar 2012). However, none has examined the impact of network stability on access to novel information, which is a key element of any dynamic model of vision advantages.

It is not immediately clear how network stability will affect access to novel information. On one hand, changes in communication partners over time may expose people to new ideas, perspectives and information. If new communication partners come from disparate areas of the network, such changes may enable access to new information. Even when controlling for network diversity and

channel bandwidth, new communication partners may create access to new information if the overlap in information amongst those in cohesive networks is not comprehensive. Simply put, talking to new people exposes us to new ideas and information. We therefore hypothesize:

- Hypothesis 3a: Network stability is associated with receiving less information diversity and less total non-redundant information.

On the other hand, network instability may reduce the trust and depth of the relationships with communication partners. If the incidence of communication over time is low, even if the volume is high, contacts may be less willing to share new or sensitive information. Prior research has shown that trust impacts the willingness to share information (Coleman, 1988) and that infrequent ties typically share information of lower complexity and detail (Hansen, 1999; Uzzi, 1997). We therefore also propose the following competing hypothesis:

- Hypothesis 3b: Network stability is associated with receiving more information diversity and more total non-redundant information.

3 Unpacking "Novelty": Defining Variables

Current literature remains vague in precisely defining the dimensions of novelty or novel information that should matter for vision advantages. However, following recent work, we believe three distinct aspects of novelty are important —(i) the diversity of the information received, which can be thought of as the variance of the topics being discussed by ego either with a given contact (dyad-level) or across all contacts (ego-level), (ii) the total volume of non-redundant information received by ego either from a given contact (dyad-level) or across all contacts (ego-level), and (iii) the uniqueness of the information received, which can be thought of as the distance between the topics discussed with one contact and the topics discussed with all of egos other contacts. The distinctions among these three measures of novelty have clear implications for our theory and we develop three distinct empirical measures that correspond to these concepts below.

3.1 Information Diversity

We measured the degree to which a specific stream of information was focused or diverse by measuring the dissimilarity of their topic distributions. Sets of messages were compared to each other, and the degree to which they were about a set of focused topics, or rather about a wider set of diverse topics was characterized. To remain consistent with current literature, we used the most common measure of document similarity, cosine similarity, to construct our measures of information diversity, which we measured at both the ego network level and the dyad level as follows:

In the following sections we index the employees (ego) by i and represent the total number of messages received by ego i in each time period by n_{it} (such that $\sum_i \sum_t n_{it} = N$).

1. The information diversity (ID) of all the n_{it} messages that ego i receives from all his peers in a given time period t is the variance of the topic distribution vectors across all messages received by that ego in that time period. Let Γ_{ijt} be the k dimensional topic distribution vector for the jth message received by ego i in time period t and let $\overline{\Gamma}_{it}$ be the average topic distribution vector ($= \frac{1}{n_t} \sum_{j=1}^{n_t} \Gamma_{ijt}$), then, ID is defined as:

$$ID_{it} = \frac{1}{n_{it}} \sum_{j=1}^{n_{it}} \left[1 - cos(\Gamma_{ijt}, \overline{\Gamma}_{it}) \right]^2 \tag{1}$$

By this definition of information diversity, a richer, more diverse set of communications will result in a higher ID, while very specific communications, focusing on a small set of topics, will result in lower ID.

2. We also measured information diversity within a specific dyad. Dyadic information diversity is defined as the information diversity (ID) of all the messages between a specific sending alter (say) r and ego i in time-period t. The definition is equivalent to that of ID, with the exception that we only sum over the n_{itr} messages exchanged between i and r in time-period t.

$$ID_{itr} = \frac{1}{n_{itr}} \sum_{j=1}^{n_{itr}} \left[1 - cos(\Gamma_{ijtr}, \overline{\Gamma}_{itr}) \right]^2 \tag{2}$$

10

3.2 Information Uniqueness

Information uniqueness (IU) is a dyadic level variable and measures the distance of topic distributions between the ties in a given time-period. IU quantifies how similar the information conveyed to an ego i by one contact r is to the information conveyed to ego by all their other contacts q $\forall$ $q \neq r$ and from whom the ego received atleast one message in time-period t. Let $\overline{\Gamma}_{itr}$ be the average topic distribution of the messages received by ego i from alter r in time period and further let S_{it} be the number of contacts from whom the ego i received atleast one message in time-period t, then, the IU is defined as:

$$IU_{itr} = \frac{1}{S_{it} - 1} \sum_{q=1}^{S_{it}} \left[1 - cos(\overline{\Gamma}_{itr}, \overline{\Gamma}_{itq}) \right] \tag{3}$$

A greater distance between the information content a particular contact provides and what all other contacts provide indicates that the information conveyed over that specific dyad is unique compared to the information the broker receives from everyone else.

3.3 Non-Redundant Information

While the total amount of non-redundant information is clearly a volumetric measure (i.e. measured in number of bits) we recognize that simply measuring the total raw amount of information is unsatisfactory, since the information received might be highly redundant. We therefore needed to develop a measure which quantifies the potential amount of information conveyed by a message given its topic distribution vector. An ideal measure for this purpose is the information entropy $\mathbb{H}(\Gamma) = \mathbb{E}[-ln\Gamma]$. Information entropy measures how much information there is in an event. If ego learns something they already know, the novel information they receive is very small. A message containing mostly information already known to ego will have very low entropy. If we want to determine the amount of non-redundant information conveyed through messages along a tie $\langle i, r \rangle$, we want to account for all other information i receives through all other ties j (with $i \neq j$) and control for redundancy in the information provided by those other ties.

Conditional entropy (CE), define per tie/dyad measures the average amount of non-redundant

information go receives from a specific alter p, given the topic distribution vectors of all other alters communicating with the ego. Conditional entropy therefore measures the marginal amount of novel information provided to ego by a specific alter, relative to the set of information ego is receiving from all of his other contacts. Any overlap in information (i.e. information redundancy) between $\langle i, r \rangle$ and any other $\langle i, q \rangle$, with $q \neq r$ is thus discounted and CE measures only the fraction of truly non-redundant information provided by alter r. Let $\overline{\Gamma}_{itr}$ and $\overline{\Gamma}_{itq}$ be the average topic distribution vectors for the ties $\langle i, r \rangle$ and $\langle i, q \rangle$ respectively. We then define the conditional entropy CE of tie $\langle i, r \rangle$ to be:

$$CE_{itr} = \mathbb{H}(\overline{\Gamma}_{itr} | \overline{\Gamma}_{it1}, \overline{\Gamma}_{it2}, \ldots, \overline{\Gamma}_{it(S_{it}-1)}) \tag{4}$$

where $S_{it} - 1$ is the total number of ties (excluding r) from which i received atleast one message in time-period t.

If we further consider the total amount of non-redundant information i receives from all his sources, we can examine the joint entropy (JE) of information received from all of i's contacts. JE measures the total amount of non-redundant information which can be encoded given the set of topic distribution vectors. We therefore define joint entropy as follows:

$$JE_{it} = \mathbb{H}(\overline{\Gamma}_{it1}, \overline{\Gamma}_{it2}, \ldots, \overline{\Gamma}_{itS_{it}}) \tag{5}$$

Conditional entropy and joint entropy measure the amount of non-redundant information provided to ego by a given contact and by all of egos contacts respectively. We denote both measures of non-redundant information for the dyadic and ego level cases as NRI in what follows. To illustrate the differences between these three measures of information novelty, consider the following examples: First, in considering information diversity we should find a set of received messages covering many different topics, such as accounting, projects, IT, social gatherings, and news to have a large information diversity, whereas if most messages are about one or two topics the corresponding information diversity of that set will be smaller. Second, considering information uniqueness, we propose that a set of messages from one contact and covering a topic no one else talks about (i.e.

sports) has a high information uniqueness, if, for example, no other sources mention the topic in the corpus. Finally, non-redundant information quantifies the amount of additional information a given source contributes (as measured through conditional entropy). Hence, if only a single source talks about a given topic (i.e. sports) the amount of information novelty is identical to the total volume of novel information from that source. However, if at least one other source mentions that topic, the amount of non-redundant information provided by the first source is reduced by a measure proportional to the amount everyone else talks about that topic. After crosschecking all sources for references to given topics, the joint entropy is identical to the total raw amount of novel information received if and only if there are no redundancies, otherwise it is smaller.

3.4   Longitudinal Entropy

The value of information we receive depends, among other factors, on prior knowledge. Up to this point we have discussed the connection between network diversity and the diversity of information received in a static sense, considering information sent and received in a single (or pooled) period of time. A more complete characterization of information novelty, however, should consider how these factors depend on prior knowledge, or what one knows or has learned in the past. To incorporate prior knowledge, we use the same information entropy framework, but in a longitudinal setting. In particular we ask: "What is the amount of non-redundant information received during time period t given prior knowledge received in t-n?"

We take a relative view of information received when studying information novelty, by comparing the novelty of information across different dyads or individuals in the email network. For instance when analyzing the information diversity between dyads we measure the amount of information received across one dyad relative to all other dyads. When analyzing information novelty in a temporally static setting, this conceptualization is symmetric in that the dyad of reference is interchangeable. In the case of knowledge aggregation, however, the point of reference (the prior information of the receiving node) is not interchangeable and the amount of aggregated information grows over time.

Additionally we would like to account for memory loss or the decay of the value of information

over time. To comprehensively characterize information decay or memory loss, we consider two extreme cases of the degree to which information decays, comparing model results for the two diametric cases: 1) information aggregated in panels $\{1, \ldots, t-1\}$, relative to panel in time-period t, which represents information aggregation without any decay. We call this the memory (mem) model as illustrated in Figure 1; and 2) information aggregated only in panel $t-1$, relative to panel t. We call this the memoryless (ml) model.

Both cases describe different aspects of information aggregation and capture different dimensions of novelty relevant to the theory we develop and to our modeling approach. In the first case, which retains a long term memory of information received over time, the amount of prior information each actor is aware of grows over time and the new novel information obtained per unit of time decreases systematically as a consequence because new information is compared to a larger body of potentially redundant information already known to the actor as illustrated in Figure 1. Assuming knowledge in a topic area is finite and does not increase over time, as one gains knowledge of that topic area over time, if they retain that knowledge with no memory loss, then new information obtained on that topic in each period is likely to be less and less novel to them and the total amount of novel information they receive about that topic will decrease over time as they learn all there is to know about that topic.

In the second case, we only consider information aggregated in the prior period as our point of reference. In essence this model considers a complete decay of information from one time frame to the next. The corresponding hypothetical scenario is one of a memoryless Markov process, i.e. the amount of additional non-redundant information is only a function of the new information received and information known in the prior period, but no prior information obtained in time periods 1 to $t-2$.

We compute these measures, which we call *longitudinal entropy*, by determining the difference of the JE received during any given (current) panel $2, \ldots, t$ and the joint entropy of prior information. In the case of the memory model, we compare new information received in the current time period to prior information aggregated across the set of all prior time periods $1, \ldots, t-1$, whereas, in the memoryless model, we only compare new information received in the current time period to prior
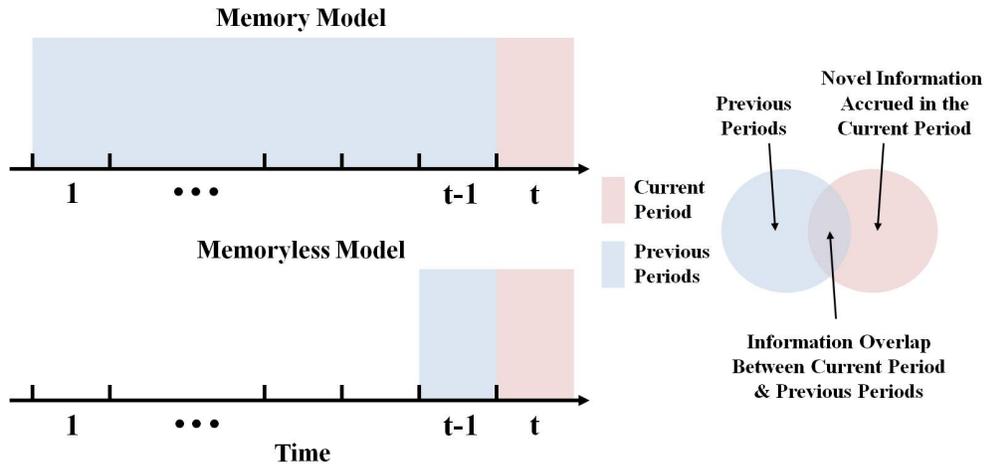
14

Memory Model

1   • • •   t-1   t

Current Period

Previous Periods

Memoryless Model

1   • • •   t-1   t

Time

Previous Periods

Novel Information Accrued in the Current Period

Information Overlap Between Current Period & Previous Periods

Illustration of Marginal Novelty in the Current Period

Memory Model

Previous Periods

Novel Information Accrued in the Current Period

Information Overlap Between Current Period & Previous Periods

Memoryless Model

Previous Period

Novel Information Accrued in the Current Period

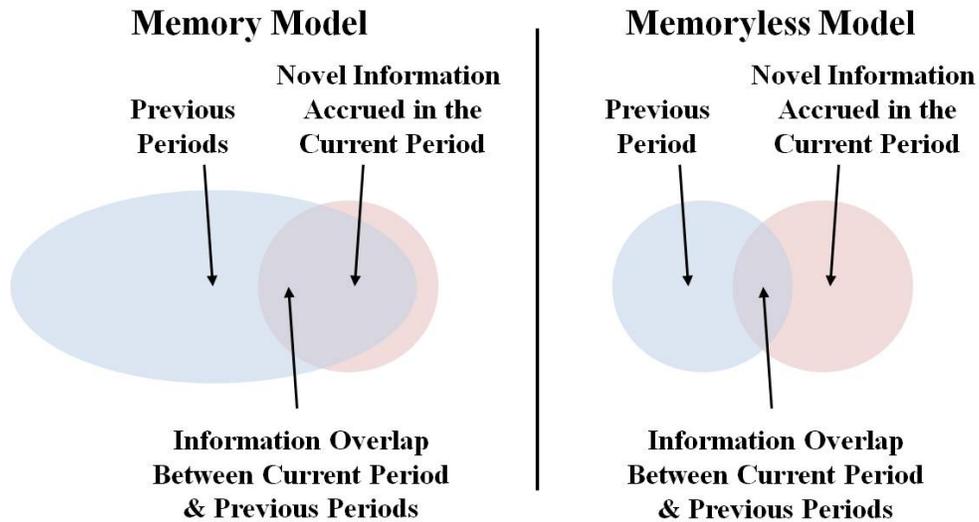Information Overlap Between Current Period & Previous Periods

Figure 1: Illustration of the Memory & Memoryless Models of Longitudinal Entropy. The amount of novel (non-redundant) information accrued in the current panel is the set of information in the "Current Period" in the Venn diagram that does not overlap with the information in "Previous Periods." The memory model considers information accumulated in all previous periods, while the memoryless model only considers information accumulated in the last period (t-1) before the current period (t).

information received in the time period $t-1$. The resulting definitions of longitudinal entropy LE in both cases are:

$$LE_{it}^{mem} = \sum_{s=1}^{t} JE_{is} - \sum_{s=1}^{t-1} JE_{is} \tag{6}$$

$$LE_{it}^{ml} = \sum_{s=t-1}^{t} JE_{is} - \sum_{s=t-1}^{t-1} JE_{is} \tag{7}$$

3.5 Network Size and Bandwidth

Network size S_{it} is defined as the number of contacts from which an ego i received at least one message during a given time period t.

Channel bandwidth per tie $\langle i, r \rangle$ is simply the number of messages n_{itr} received by ego i from alter r in the time-period t. An ego i's average channel bandwidth is defined as the total number of messages he receives over all the incoming ties during that time-period. In other words:

$$B_{it} = \frac{n_{it}}{S_{it}} \tag{8}$$

3.6 Network Constraint

We use Burt's network constraint metric to measure brokerage. Specifically we break this measure down into its individual components in order to apply it not only to the constraint of each brokers ego network but also to measure egos investment in specific ties. Derived by using the bidirectional traffic of emails between any two brokers, we denote the proportion of time and effort invested by ego i in a specific alter r as p_{ir}. We denote this as his direct investment DI_{ir}. Further we consider secondary or redundant investments via mutual relationships (indexed j) in the communications network. This measure of redundant investment is defined as:

$$RI_{ir} = \sum_{j \neq i \neq r} p_{ij} p_{jr} \tag{9}$$

Both factors allow ego to invest in his relationship to alter i and we expect both to have an

influence on the diversity and amount of non-redundant information ego receives from a specific contact in his network. To quantify the amount of network constraint ego experiences in his personal network, we sum over all contributions (direct and redundant) for all of his peers:

$$NC_{it} = \sum_{r=1}^{S_{it}} (DI_{irt} + RI_{irt})^2 = \sum_{r=1}^{S_{it}} (p_{ir} + \sum_{j \neq i \neq r} p_{ij} p_{jr})^2 \tag{10}$$

The use of the individual contributions of direct and redundant investments allows us to evaluate the effects of investment in individual ties. By using the aggregated term of investments over all ties, we can further evaluate the network constraint the information broker experiences. Understanding the relationship between network structure and novel information on both levels of individual dyadic ties and aggregated over the full set of ties that comprise the ego network allows us to truly understand the mechanics of the vision advantage.

4 Empirical Setting

We explored the anatomy and dynamics of vision advantages by analyzing the content and structure of an evolving corporate email network over twelve months. The firm that we studied was a medium sized, global digital media firm delivering language and localization services such as translation, dubbing, and sub-titling for film, digital gaming and web content producing clients around the world. The services provided by employees of this firm required constant information seeking and communication to solve problems that were highly localized. For example, translating a movie from English into thirty other languages required translators to seek information about current local language use and modern day idioms from country and regional experts in the firm. In interviews and during participant observation, employees frequently reported and were observed seeking information from people in disparate parts of the firms communication network in order to solve these highly idiosyncratic problems. In this way, novel information drove the speed and quality of the work product. Our interviews revealed that timely access to such novel information from disparate parts of the network were important drivers of project completion rates and error rates.

17

The lead author first collected data from 10 weeks of participant observation in the firm over a 6 month period prior to the start of quantitative data collection. During this period, we collected data from interviews of the entire senior executive team and key informants from the three main operational teams —sales, technology and operations. We also conducted interviews with employees in each of the major language teams that produce the localization work. These areas represent a comprehensive set of all of the types of employees in the firm. In addition to these interviews we observed the employees of each of these divisions performing their work, taking detailed notes of our observations. This initial data collection helped us understand the setting, the work that was being done, the role of novel information in the work and the nature of the social network dynamics at play in the communications of the firm.

Following this qualitative data collection, we collected complete and comprehensive data on the content and structure of the firms evolving corporate email network as described below. Figure 2 displays the largest connected component of this graph using aggregated data over the twelve months colored to distinguish communities identified by the Blondel et al. (2008) community detection algorithm. This figure gives a sense of the distinct clusters of communication that exist in the firms email network. The distinct communities that have developed within the firms communication structure fulfill different roles in the companys work flow. They accumulate distinct pools of knowledge and information and create a setting in which employees must reach outside of their local networks, through weak bridging ties, to gain access to novel information they need to complete their work. The variegated nature of the community communications in the firm provides a perfect opportunity to study the role of structural diversity in providing privileged access to novel information and therefore to investigate the dynamics of vision advantages.

4.1 Data

Our quantitative data collection focused on two areas: (i) human resource information such as employees gender, and date of hire, and (ii) internal email communications captured on the corporate email server for all employees. The goal of our analysis is to test the theoretical mechanisms that establish and enable "vision advantages" from structural holes, "the strength of weak ties,"
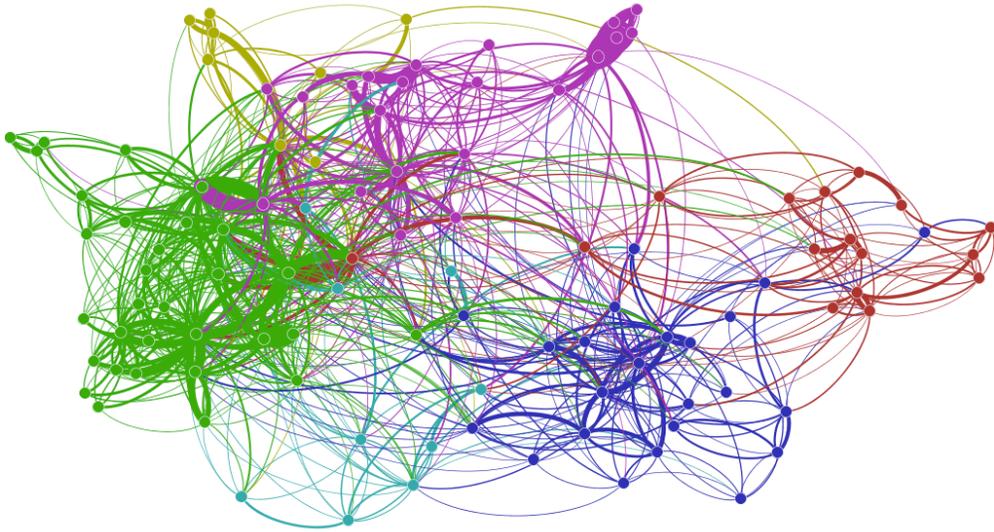
Figure 2: Communities in Network Structure of the Global Media Firm.

and "the diversity-bandwidth tradeoff." To do so, we unpack and operationalize the concept of "novel information" to reflect the theoretical distinctions between information diversity, total non-redundant information and information uniqueness. We then measure information diversity, total non-redundant information and information uniqueness in the content of the emails exchanged between employees and statistically relate variance in these measures to the dynamic structural characteristics of the evolving corporate email network over time (specifically Burt's constraint measure, Aral and Van Alstyne (2011)'s bandwidth measure, a measure of network stability and relevant control variables). The result is the first empirical evidence of how dynamic network structure is related to the flow of novel information in a firm —specifically the diversity of the information exchanged, the ebb and flow of the volume of total novel information exchanged and the variation in the uniqueness of the information exchanged. The results provide intriguing evidence of how vision advantages, the strength of weak ties and the diversity-bandwidth tradeoff all operate in practice.

Overall we collected two million emails sent and received from 232 employees over twelve months during 2010. All email aliases were associated with a single user. The email content was anonymized using the same hashing algorithm used by Aral and Van Alstyne (2011) during data collection on

the email server before being transferred off-site for further analysis. During this process, the subject and content of each message was processed for 'stop words, such as 'a', 'an, 'the, 'and and other high frequency words, which were removed. Remaining keywords were root stemmed, e.g. 'multitasking' and 'multi-task' become 'multitask'. We then calculated the keyword frequency for each email. Keywords were replaced by unique hash codes ensuring anonymity of the actual content (see (Aral and Van Alstyne, 2011) for further details on email data collection and hashing).

The descriptive statistics of the dataset are shown in Table 1 and Table 2 shows the correlations between different variables.

Variable	Mean (μ)	SD (σ)	Minimum	Maximum
Gender Difference	0.50	0.20	0	1
Hire Date Difference	-0.11	3.94	-15.07	12.48
Total Incoming Emails (n_{it})	243.1	305.1	1	2566
Network Size (S_{it})	23.08	19.18	1	96
Channel Bandwidth (B_{it})	9.58	8.21	1	99
Network Constraint (NC_{it})	0.34	0.27	0.07	1.30
Information Diversity (ID_{it})	0.81	0.21	0	0.98
Non-Redundant Information (JE_{it})	49.1	44.19	0.25	236.4

Table 1: Descriptive Statistics (Panel of 232 employees (N=2300) who received atleast one email over a 12 month period from Jan-Dec. 2010). Hire date and gender differences are the differences (to-from) averaged over all the contacts who sent atleast one email during that panel period.

Variable	1	2	3	4	5	6	7	8
1. Gender Difference	-	-	-	-	-	-	-	-
2. Hire Date Difference	-0.05	-	-	-	-	-	-	-
3. Total Incoming Emails (n_{it})	-0.01	0.05	-	-	-	-	-	-
4. Network Size (S_{it})	-0.03	0.05	0.81	-	-	-	-	-
5. Channel Bandwidth (B_{it})	0.02	0.02	0.50	0.14	-	-	-	-
6. Network Constraint (NC_{it})	0.07	-0.05	-0.39	-0.65	0.09	-	-	-
7. Information Diversity (ID_{it})	-0.06	0.04	0.39	0.55	0.17	-0.77	-	-
8. Non-Redundant Information (JE_{it})	-0.02	0.04	0.84	0.99	0.16	-0.62	0.53	-

Table 2: Pairwise Correlations between different variables of the panel of 232 employees.

The full dataset was then processed using Latent Dirichlet Allocation (LDA) (Blei et al., 2003). Latent Dirichlet Allocation (LDA) (Blei et al., 2003) is a probabilistic model for text corpora (collection of documents). Each document is composed of a sequence of 'n' words $\mathbf{w} = \{w_1, w_2, \ldots, w_n\}$. Further, the corpus (collection of documents) C is composed of a total of 'N' doc-

uments $C = \{\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_N}\}$. LDA represents each document as a random mixture of k latent topics (specified by the user) and further each topic is characterized by a distribution over words and provides estimates of the the topic distribution of each document, i.e. how is the composition of a document split among all the k topics?

In our case, the corpus is the entire collection of emails exchanged between the employees and each message is a different document. Hence, we are able to estimate a topic distribution (a 'k' dimensional vector whose entries sum to 1) for each email. We chose the total number of topics to be 50 (a typical number used in many text modeling studies); our results are robust to trying different number of topics.

Note that in our setting the content is anonymized and the meaning of the topics is not further explicated, but LDA quantifies the co-occurrence of terms in exactly the same way it does with clean text (see (Aral and Van Alstyne, 2011) for descriptions of robustness checks that establish the robustness of these procedures using both hashed and unhashed clean text). We derive measures of information diversity, non-redundancy and uniqueness for the streams of emails within the company based on the topical distribution of each message.

4.2 Model Specification

We use monthly panel data (i.e. the time-period t is a month) to estimate the relationship between network structure and information novelty. To understand the exact underlying principles, we not only investigate the relationship between ego network structure and the information ego receives but also the flow of information along individual dyadic ties.

In all the following analysis, we control for the effects of demographic characteristics of each information broker such as gender and hire date using data from the human resources department of the firm. We are also interested in controlling for the effects of differences in those demographic factors between senders and receivers in the email network. We therefore account for two primary dyadic demographic factors: difference in hire date and difference in gender between senders and receivers. We compute demographic control variables both on the level of individual ties and also aggregate them by computing their average values for the ego networks of information brokers over

all incoming ties in that time-period t of their respective ego networks.

4.2.1 Ego-level Analysis

We first replicate the Diversity-Bandwidth tradeoff results (Aral and Van Alstyne, 2011) on our dataset. Equations 11, 12 give the specification and Table 3 shows the estimation results.

$$B_{it} = \gamma_{it} + \delta_t + \beta_1 NC_{it} + \beta_2 S_{it} + \beta_3 (S_{it})^2 + \sum_k \beta_k X_{ik} + \epsilon_{it} \tag{11}$$

$$NC_{it} = \gamma_{it} + \delta_t + \beta_1 B_{it} + \beta_2 S_{it} + \beta_3 (S_{it})^2 + \sum_k \beta_k X_{ik} + \epsilon_{it} \tag{12}$$

	Model 1	Model 2	Model 3	Model 4
Gender Diff	0.020	0.011	-0.019	-0.023
	(0.013)	(0.013)	(0.017)	(0.018)
Hire Date Diff	0.014	0.054	-0.068	-0.142**
	(0.025)	(0.036)	(0.036)	(0.048)
Network Size (S_{it})	-0.967***	-0.862***	0.272***	0.237***
	(0.027)	(0.033)	(0.046)	(0.051)
Network Size Squared (S_{it})^2	0.277***	0.237***	-0.024	-0.017
	(0.012)	(0.013)	(0.018)	(0.019)
Network Constraint (NC_{it})	-	-	0.146***	0.134***
			(0.028)	(0.029)
Channel Bandwidth (B_{it})	0.087***	0.076***	-	-
	(0.015)	(0.015)		
Constant	-0.276***	N/A	-0.187**	N/A
	(0.045)		(0.69)	
Temporal Controls	Month	Month	Month	Month
R^2	0.38	0.23	0.07	0.06
F-statistic (df)	87.09***	45.63***	10.7***	10.02***
	(16)	(16)	(16)	(16)

Table 3: (Ego-level) The Network Diversity-Bandwidth Tradeoff (N=2300) using Random Effects regression. (Models 1-2) Dependent Variable= Network Constraint (NC_{it}), (1)= Random Effects, (2)=Fixed Effects. (Models 3-4) Dependent Variable= Channel Bandwidth (B_{it}), (3)= Random Effects, (4)= Fixed Effects. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Next, we examine the relationships between network structure and information diversity (ID) and total non-redundant information (NRI). The specification is given in Equation 13, where X_{ik} are ego-specific time-invariant covariates namely averaged gender difference and hire date between

an ego and the alters. Results are shown in Table 4.

$$\{ID_{it}, NRI_{it}\} = \gamma_{it} + \theta_i + \delta_t + \beta_1 NC_{it} + \beta_2 B_{it} + \beta_3 S_{it} + \beta_4 (S_{it})^2 + \sum_k \beta_k X_{ik} + \epsilon_{it} \qquad (13)$$

4.2.2 Dyad-level Analysis

The specification for dyad-level models is given in Equation 14 and the results are shown in Table 5.

$$\{ID_{itr}, NRI_{itr}, IU_{itr}\} = \gamma_{itr} + \delta_t + \beta_1 NC_{itr} + \beta_2 B_{itr} + \beta_3 S_{it} + \beta_4 (S_{it})^2 + \beta_5 DI_{irt} + \beta_6 RI_{irt} + \sum_k \beta_k X_{ikr} + \epsilon_{itr}$$
$$(14)$$

4.2.3 Longitudinal Analysis

Finally, we perform longitudinal analysis where we perform regression on temporal differences of variables. The specification is given in Equation 15.

$$LE_{it} = \beta_1 \Delta(B_{it}) + \beta_2 \Delta(NC_{it}) + \Delta\epsilon_{it} \qquad (15)$$

where $\Delta(B_{it}) = B_{it} - B_{it-1}$ and $\Delta(NC_{it}) = NC_{it} - NC_{it-1}$

5 Results

5.1 The Diversity-Bandwidth Tradeoff

By investigating the relationship between network constraint and bandwidth and their joint effect on the novelty of incoming email, we are able to describe how changes in the communication network structure are associated with changes in the type of information received. If the diversity-bandwidth tradeoff regulates the receipt of novel information, we should observe two phenomena in our data. First, as employees' networks become more diverse (less constrained), we should see the bandwidth of their communication channels contract. Second, we should observe increases in

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Gender Diff	-0.000	-0.001	0.005	0.001	0.000	0.000
	(0.013)	(0.012)	(0.013)	(0.010)	(0.003)	(0.000)
Hire Date Diff	-0.041	-0.045*	-0.133***	-0.002	0.004	0.008
	(0.024)	(0.022)	(0.035)	(0.024)	(0.005)	(0.007)
Network Size (S_{it})	-	0.480***	0.582***	-	0.930***	0.918***
		(0.032)	(0.037)		(0.007)	(0.007)
Network Size Squared (S_{it})^2	-	-0.198***	0.202***	-	0.028***	0.027***
		(0.013)	(0.014)		(0.003)	(0.003)
Network Constraint (NC_{it})	-0.805***	-0.613***	-0.575***	-0.275***	-0.001	-0.003
	(0.017)	(0.020)	(0.021)	(0.014)	(0.004)	(0.004)
Channel Bandwidth (B_{it})	0.149***	0.124***	0.071***	0.094***	0.018***	0.018***
	(0.016)	(0.015)	(0.013)	(0.012)	(0.003)	(0.003)
Constant	-0.071	0.158***	N/A	-0.123*	-0.037***	N/A
	(0.041)	(0.042)		(0.050)	(0.010)	
Temporal Controls	Month	Month	Month	Month	Month	Month
R^2	0.52	0.57	0.44	0.25	0.95	0.84
F-statistic (df)	165.6***	177.93***	116.84***	49.9***	3081.3***	2161.35***
	(15)	(17)	(17)	(15)	(17)	(17)

Table 4: (Ego-level) Predicting Information Diversity and Non-Redundant Information (N=2300) using Random Effects regression. (Models 1-3) Dependent Variable= Information Diversity (ID_{it}), (1, 2)= Random Effects, (3)=Fixed Effects. (Models 4-6) Dependent Variable= Non-Redundant Information (Joint Entropy JE_{it}), (4, 5)= Random Effects, (6)= Fixed Effects. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Gender Diff	0.011	N/A	0.036*	N/A	0.034*	N/A
	(0.013)		(0.014)	(0.016)	(0.015)	
Hire Date Diff	-0.009	N/A	-0.011	N/A	-0.038	N/A
	(0.006)		(0.007)		(0.008)	
Network Size (S_{it})	0.143***	0.222***	0.159***	0.163***	0.128***	0.039***
	(0.006)	(0.011)	(0.006)	(0.013)	(0.007)	(0.011)
Network Size Squared (S_{it})^2	-0.016***	-0.034***	-0.022***	-0.014***	-0.076***	-0.070***
	(0.004)	(0.005)	(0.004)	(0.006)	(0.004)	(0.004)
Direct Investment (DI_{irt})	0.204***	0.304***	0.027***	0.102***	-0.063***	-0.079***
	(0.005)	(0.008)	(0.006)	(0.009)	(0.006)	(0.006)
Redundant Investment (RI_{irt})	0.089***	0.085***	0.011	0.037***	-0.07***	-0.054***
	(0.005)	(0.007)	(0.006)	(0.008)	(0.006)	(0.006)
Channel Bandwidth (B_{it})	0.354***	0.259***	0.186***	0.136***	-0.086***	-0.071***
	(0.005)	(0.006)	(0.006)	(0.007)	(0.006)	(0.005)
Constant	-0.183***	N/A	-0.062***	N/A	0.043***	N/A
	(0.014)		(0.016)		(0.016)	
Temporal Controls	Month	Month	Month	Month	Month	Month
R^2	0.21	0.12	0.05	0.03	0.03	0.02
F-statistic(df)	759.0***	719.0***	158.3***	108.9***	90.8***	59.7***
	(18)	(16)	(18)	(16)	(18)	(16)

Table 5: (Dyad-level) Predicting Information Diversity within a dyad, Non-Redundant Information and Information Uniqueness (N=53079) using Random Effects regression. (Models 1, 2) Dependent Variable= Information Diversity within a dyad (ID_{itr}), (1)= Random Effects, (2)=Fixed Effects. (Models 3,4) Dependent Variable= Non-Redundant Information (Conditional Entropy (CE_{itr})) (3)= Random Effects, (4)=Fixed Effects. (Models 5, 6) Dependent Variable = Information Uniqueness (IU_{itr}). (5)= Random Effects, (6)=Fixed Effects. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

	Model 1	Model 2	Model 3	Model 4
Network Constraint (ΔNC_{it})	-0.690***	-0.723***	-0.673***	-0.739***
	(0.029)	(0.029)	(0.027)	(0.026)
Channel Bandwidth (ΔB_{it})	-	0.0169***	-	0.313***
		(0.020)		(0.019)
Constant	0.004	0.005	0.139***	0.108***
	(0.008)	(0.007)	(0.026)	(0.025)
R^2	0.25	0.28	0.26	0.36
F-statistic (df)	577.4***	337.5***	599.0***	478.0***
	(1)	(2)	(1)	(2)

Table 6: (Ego-level) N=1723 (Models 1-2) Dependent Variable= Longitudinal Entropy (Memoryless) LE_{it}^{ml}. (Models 3-4) Dependent Variable= Longitudinal Entropy (with memory (mem)) LE_{it}^{mem}. Only the people who received at atleast 1 email during all the 12 months of the panel are included. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

the receipt of novel information both as networks become more structurally diverse and as channel bandwidth expands. If these conditions hold, then a tradeoff between network diversity and channel bandwidth is creating countervailing effects on the receipt of novel information.

We find strong evidence confirming the Diversity-Bandwidth Tradeoff. As employees communicated with contacts which are well connected to each other, the overall bandwidth of their communication channels to those contacts widened quite rapidly. For instance, we estimate that a one standard deviation increase in network constraint (NC, i.e. reduced structural diversity) was associated on average with around 0.15 (Model 3, Table 3) standard deviation increase in bandwidth. Similarly, one standard deviation increase in channel bandwidth gives a 0.09 (Model 1, Table 3) standard deviation increase in network constraint. Therefore, as networks become less diverse, the thickness of their communication channels increases.

Examining the effects of the Diversity-Bandwidth Tradeoff on information novelty, we find a strong effect on the diversity of information received and the total amount of non-redundant information received. As networks become more structurally diverse, brokers experience an increase in the diversity of the information they receive and the total amount of non-redundant information they receive. Specifically, a one standard deviation increase in structural constraint is associated, on average, with a 0.61 (Model 2, Table 4) standard deviation decrease in information diversity and a one standard deviation increase in bandwidth is associated with a 0.12 (Model 2, Table 4) standard deviation increase in information diversity. Network size yields a positive effect as well with a 0.48 standard deviation increase in information diversity.

Considering the effects of network structure on the total volume of non-redundant information brokers receive, we find a negative relationship between network constraint and non-redundant information. As brokers networks become more constrained, they receive less total non-redundant information, confirming Burt's basic argument. In contrast, network size [1] and channel bandwidth are both positively associated with NRI. In particular, a one standard deviation increase in network constraint decreases non-redundant information by 0.28 standard deviations (Model 4, Table 4) and one standard deviation increase in channel bandwidth is associated with 0.09 (Model 4, Table 4)

---

[1]Since there is severe collinearity between network size and network constraint, so they had to be evaluated in separate regressions.

standard deviation increase in non-redundant information.

These results confirm the Diversity-Bandwidth Tradeoff and validate and replicate the results of Aral and Van Alstyne (2011) in a completely different setting. The robustness of the findings in this new setting and with new data provides strong evidence that the Diversity-Bandwidth Tradeoff is a general phenomenon that holds at the heart of the vision advantage mechanism theorized to explain The Strength of Weak Ties and Brokerage Theory. We were so struck by the consistency of the parameter estimates across these two studies that we requested data from the first study authors and plotted the relationships of key variables across the two samples. The correspondence of these relationships across the two settings is shown in Figure 3. Although ours is a much larger sample and therefore displays much smoother aggregated relationships, the overall patterns in the data are remarkably similar.
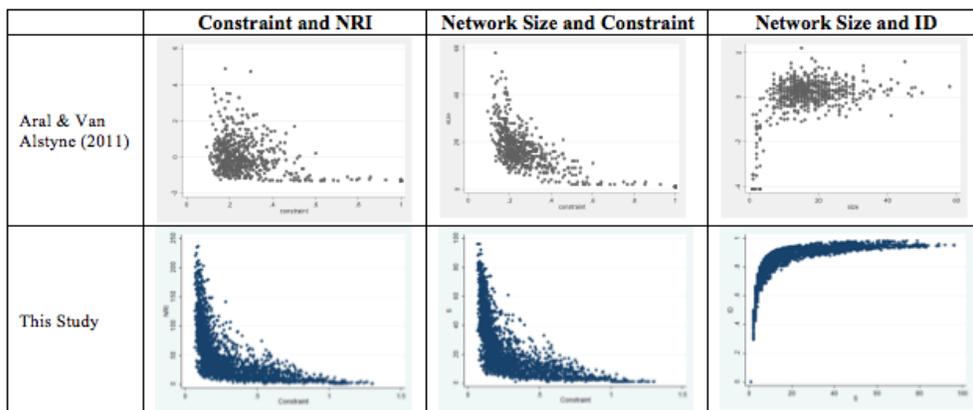

Figure 3: Comparing Network Structure & Novel Information across Studies. Relationships between measures of Burt's Network Constraint, Total Non-Redundant Information (NRI), Network Size and Information Diversity in data taken from Aral and Van Alstyne (2011) and in this study. Measures of all variables are constructed using the methodologies outlined in each paper respectively in order to stay true to the measurement made by the original authors in each paper.

5.2   Anatomy of the Vision Advantage

To further understand the mechanics of vision advantages and the Diversity-Bandwidth Tradeoff we next analyzed the communications network at the level of individual ties. This allowed us to uncover the underlying anatomy of the vision advantage. In particular, we were able to distinguish

several different contributions of network constraint NC at the level of individual ties and analyze them separately. To unpack the relationship between network constraint and access to novelty, we distinguish the two separate terms of Burts original constraint variable into its dyadic components: direct investment DI and redundant investment RI, as described in Section 3.

Specifically, we find a highly significant increase in information diversity received within ties. A one standard deviation increase in ego's direct investment in communicating with a particular alter (the proportion of communication volume they dedicate to that alter) is associated with a 0.2 (Model 1, Table 5) standard deviation increase in information diversity within that dyadic channel (ID_{irt}). This effect is reinforced by a similar positive effect in redundant investment. Together, these results indicate that the diversity of information that an ego receives within a particular relationship increases with the amount of time and effort she invests, directly and via shared relationships, in that peer. Further, we find that higher channel bandwidth facilitates ($\beta = 0.358$) information diversity within ties as well. A greater volume of communication with a particular alter is associated with an increase in the diversity of information received.

Secondly, we consider the total amount of non-redundant information conveyed to ego per tie, as measured by conditional entropy (NRI). We consistently find positive trends for direct and redundant investment as well as tie bandwidth. Bandwidth has the largest impact with a 0.19 standard deviation increase in non-redundant information for a one standard deviation change in bandwidth. This is consistent with and supports our findings on information diversity within ties as well as the amount of non-redundant information ego receives across all his peers (measured by joint entropy NRI).

Finally, when analyzing the information uniqueness between ties IU, with the receiving Ego as the point of reference, we find the opposite effect of direct and redundant investment and bandwidth. Specifically, we find a decrease in the information uniqueness a tie delivers with increased direct and redundant investment in that tie. This is further supported by the negative relationship between channel bandwidth and information uniqueness. In other words: weak bridging ties provide information that is distant from, or unique, compared to the information provided by other ties.

These results, considered together, paint a precise picture of the underlying mechanisms that

enable vision advantages. Brokers receive more diverse information and more total non-redundant information from strong, cohesive, embedded ties. However, the information a broker receives from structurally diverse, weak bridging ties is on average more unique or different (i.e. more remote in topic space) when compared to the information she receives from her other contacts. In contrast to weak bridging ties, the pair wise topical distance between the information provided by strong, cohesive ties is on average relatively small. This indicates that structurally weak ties provide unique information, which is significantly different (distant) from the information provided by the brokers core clique of communication partners. At the same time, the information provided by structurally weak ties is also more specific or topically narrow (i.e. less diverse).

Reflecting further, recall that information diversity and the total volume of non-redundant information that brokers receive both decrease in cohesive or constrained networks (Table 4). Taken together, this implies that the effects of information diversity within a channel (ID_{irt}) and information uniqueness (IU_{irt}) are countervailing —as information uniqueness increases, information diversity decreases, meaning the information provided by weak bridging ties is unique and topically narrow —and that the overall amount of novel information an ego receives through all her contacts is driven more by ties providing unique information . Considering the role of these ties in a network structural sense we find these to be predominantly structural weak ties as illustrated in Model 5.
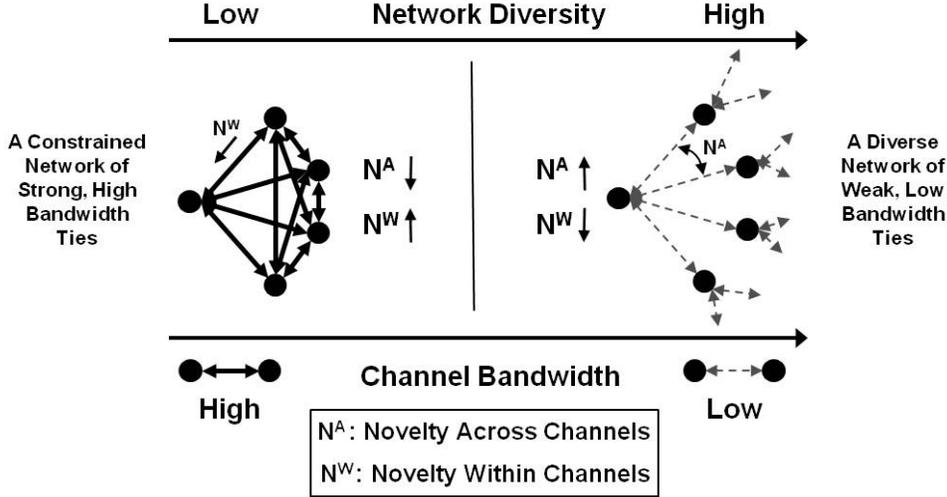

Figure 4: The Mechanics of the Vision Advantage.

These results together tell a very compelling story about how vision advantages work —we depict the mechanisms of the vision advantage graphically in Figure 4. As structural diversity increases, the bandwidth of communication channels contracts (The Diversity-Bandwidth Tradeoff). In diverse networks of weak, low bandwidth, bridging ties, novelty measured across the ties is high (meaning each contact is providing information different from what other contacts are providing), but novelty provided within each channel is decreasing. On the other hand, in constrained networks of strong, high bandwidth embedded ties, novelty across ties is decreasing due to information overlap and redundancy across channels, while at the same time novelty within each channel is increasing due to the rich, frequent, high bandwidth communication in these dyads. The mechanisms of the vision advantage become even clearer and more precise when we consider the effects of Longitudinal Entropy, which captures the accumulation of novel information over time. Panel data models explain how access to different kinds of novelty (diversity, total non-redundant information, uniqueness) changes structural variables such as network constraint and channel bandwidth change. However, examining longitudinal measures of novelty allows us to explore how access to novel information changes as information builds on itself and as actors add new information to what they already know.

As illustrated in Figure 5(a), longitudinal entropy is systematically reduced over time in the memory model due to the effects of extended memory aggregation (as we aggregate more information, novelty in the information we receive is reduced in each subsequent period). In the memoryless model as illustrated in Fig. 5(b) we find no such trend over time. If we quickly forget what we know, new information seems novel even though we may have seen it in the past. We further explore how the relationship between longitudinal entropy and features of network structure, such as channel bandwidth and network constraint, drive access to novelty over time for each user in the email network from the perspective of learning models with strong or weak memory processes in turn. The relationship between network structure and longitudinal entropy proves highly significant and this is most pronounced in the relationship between longitudinal entropy and network constraint as illustrated in Fig 5(c). The figure displays the relationship for the memoryless model, but the same trend is also found in the memory model though it is less precisely estimated.
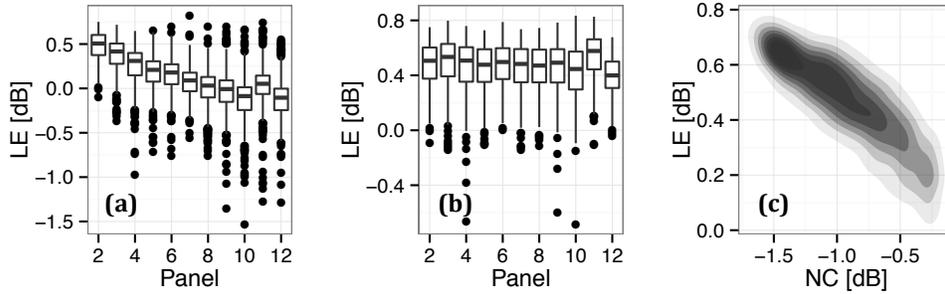
30

Figure 5: Distribution of longitudinal entropy per panel: (a) memory; (b) memoryless. (c) Density plot depicting relationship between longitudinal entropy (LE) and network constraint (NC) for the memoryless model.

In summary, network diversity (or inversely, network constraint) is the dominant factor in the relationship between network structure and longitudinal entropy, by roughly an order of magnitude, when compared to channel bandwidth . This result suggests that weak bridging ties, which provide unique information through low bandwidth, structurally diverse channels, contribute the most to the aggregation of novel information over time compared to high-bandwidth, cohesive ties. We believe this is because unique information —information that is topically distant from what other ties are providing —is more likely to be different than what we learned in the past or what we already knew.

6   Discussion and Conclusion

We analyzed the structure and content of the complete dynamic email network of employees of a medium sized global digital media firm over twelve months in order to empirically validate the vision advantage argument at the heart of the Strength of Weak Ties and Brokerage Theory, and further, to understand the dynamic mechanisms that make vision advantages work. Three results emerged from our analysis.

First, we confirmed the Diversity-Bandwidth Tradeoff at the heart of the vision advantage: As a broker's network becomes more diverse, the bandwidth of their communication channels contracts, creating countervailing effects on access to novel information. These results replicated prior work

on the Diversity-Bandwidth tradeoff with remarkable fidelity.

Second, our analysis uncovered the mechanics driving the Diversity-Bandwidth Tradeoff and highlighted differences in vision advantages offered by strong cohesive ties and weak bridging ties. Strong cohesive ties deliver greater information diversity and more total novelty, while weak bridging ties contribute the greatest uniqueness —information which is most different from what other contacts are delivering. Finally longitudinal entropy, which measures the accumulation of non-redundant information over time, is predominantly driven by network diversity, with bandwidth having a relatively smaller impact. In comparison with our former conclusions this indicates, that structurally weak ties, which provide unique information with limited bandwidth and diversity will contribute most to the aggregation of novel information over time as opposed to high-bandwidth, cohesive ties.

The theory we propose and the results of our empirical analysis together represent the first steps toward a dynamic ego-and-dyadic level model of the vision advantages that have for forty years been hypothesized to explain The Strength of Weak Ties and Brokerage Theory. In addition, the work highlights the power of combining network structure data with network content data to understand how the structure of social relationships is associated with the information content that flows through them (Sundararajan et al., 2013). All of these endeavors provide further evidence of the power of nano-level data to uncover social processes driving competitive advantages for networked actors.

References

Sinan Aral and Marshall Van Alstyne. The diversity-bandwidth trade-off1. *American Journal of Sociology*, 117(1):90–171, 2011.

Sinan Aral, Erik Brynjolfsson, and Marshall W Van Alstyne. Productivity effects of information diffusion in networks. *Available at SSRN 987499*, 2007.

Sinan Aral, Erik Brynjolfsson, and Marshall Van Alstyne. Information, technology, and information worker productivity. *Information Systems Research*, 23(3-part-2):849–867, 2012.

Wayne E Baker. Market networks and corporate behavior. *American journal of sociology*, pages 589–625, 1990.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286 (5439):509–512, 1999.

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008 (10):P10008, 2008.

Nathaniel Bulkley and Marshall W Van Alstyne. Why information should influence productivity. 2004.

Ronald S Burt. Structural holes and good ideas. *American journal of sociology*, 110(2):349–399, 2004.

Ronald S Burt. *Brokerage and closure: An introduction to social capital.* OUP Oxford, 2005.

Ronald S Burt. Information and structural holes: comment on reagans and zuckerman. *Industrial and Corporate Change*, 17(5):953–969, 2008.

Ronald S Burt. *Structural holes: The social structure of competition.* Harvard university press, 2009.

Wesley M Cohen and Daniel A Levinthal. Absorptive capacity: a new perspective on learning and innovation. *Administrative science quarterly*, pages 128–152, 1990.

James S Coleman. Social capital in the creation of human capital. *American journal of sociology*, pages S95–S120, 1988.

Mark S Granovetter. The strength of weak ties. *American journal of sociology*, pages 1360–1380, 1973.

Morten T Hansen. The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative science quarterly*, 44(1):82–111, 1999.

Morten T Hansen. Knowledge networks: Explaining effective knowledge sharing in multiunit companies. *Organization science*, 13(3):232–248, 2002.

Andrew Hargadon and Robert I Sutton. Technology brokering and innovation in a product development firm. *Administrative science quarterly*, pages 716–749, 1997.

David Lazer and Allan Friedman. The network structure of exploration and exploitation. *Administrative Science Quarterly*, 52(4):667–694, 2007.

David Obstfeld. Social networks, the tertius iungens orientation, and involvement in innovation. *Administrative science quarterly*, 50(1):100–130, 2005.

John F Padgett and Christopher K Ansell. Robust action and the rise of the medici, 1400-1434. *American journal of sociology*, pages 1259–1319, 1993.

Joel M Podolny. Networks as the pipes and prisms of the market1. *American journal of sociology*, 107(1):33–60, 2001.

Ray Reagans and Ezra W Zuckerman. Networks, diversity, and productivity: The social capital of corporate r&d teams. *Organization science*, 12(4):502–517, 2001.

Simon Rodan and D Charles Galunic. More than network structure: how knowledge heterogeneity influences managerial performance and innovativeness. *Strategic Management Journal*, 25:541–556, 2004.

Arun Sundararajan, Foster Provost, Gal Oestreicher-Singer, and Sinan Aral. Research commentary-information in digital, economic, and social networks. *Information Systems Research*, 24(4):883–905, 2013.

Brian Uzzi. Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative science quarterly*, pages 35–67, 1997.

Brian Uzzi and Jarrett Spiro. Collaboration and creativity: The small world problem1. *American journal of sociology*, 111(2):447–504, 2005.

Marshall Van Alstyne and Erik Brynjolfsson. Global village or cyber-balkans? modeling and measuring the integration of electronic communities. *Management Science*, 51(6):851–868, 2005.

Duncan J Watts and Steven H Strogatz. Collective dynamics of small-worldnetworks. *nature*, 393 (6684):440–442, 1998.