The Anatomy & Dynamics of Vision Advantages

Completed Research Paper

Sinan Aral
New York University, Stern School of Business, IOMS
New York City, New York
sinan@stern.nyu.edu

Vincent J. David
New York University, Stern School of Business, IOMS
New York City, New York
vdavid@stern.nyu.edu

Abstract

The Strength of Weak Ties and Brokerage Theory rely on the argument that weak bridging ties deliver novel information to brokers. Yet, very little empirical evidence exists to validate this claim. Analyzing an evolving corporate email network, we investigate the dynamic mechanisms that enable the "vision advantage". Three results emerge. First, we confirm the Diversity-Bandwidth Tradeoff at the heart of the vision advantage: As brokers' networks becomes more diverse, their channel bandwidth contracts, creating countervailing effects on access to novel information. Second, we uncover the mechanics 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, network stability increases the novelty brokers receive over time, providing the first evidence of the role of network dynamics in vision advantages.

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

Introduction

For the last forty years, researchers in disciplines as diverse as economics, sociology, management 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 such as wages, job placement, promotion, creativity, innovation, political success, social support, productivity, and performance (Aral 2012; Aral 2007; Baker 1990; Bulkley et al. 2004; Burt 1992; Burt 2004; Granovetter 1973; Hansen 1999; Hansen 2002; Padgett et al. 1993; Podolny 2001; Reagans et al. 2001; Uzzi 1996; 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, very little empirical evidence has been presented to validate the existence of vision advantages or to document how they work.

The key difficulty is in observing and measuring the novelty of the information content delivered by brokers' contacts and network structure simultaneously. 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 validate the existence of vision advantages and to investigate the underlying dynamic mechanisms that enable them. We 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. We then analyze how network dynamics affect the amount of novel information brokers receive.

Three results emerge from our analysis. First, we confirm the Diversity-Bandwidth Tradeoff at the heart of the vision advantage (Aral et al. 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 also uncovers the mechanics driving the Diversity-Bandwidth Tradeoff and highlights differences in the vision advantages offered by strong cohesive ties and weak bridging 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 are delivering. Finally, we find 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 dyadic- 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 provide the first empirical evidence of the role of network stability in vision advantages. 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.

Theory

Human social networks tend to cluster due to triadic closure. As Granovetter's (Granovetter 1973) 'forbidden triad' demonstrates, 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 et al. 1998) and heavy tailed degree distributions (Barabási et al. 1999). Such structure — densely connected cliques tied together by infrequent weak bridging ties — gives rise to opportunity. Brokers with structurally diverse networks, which are low in cohesion and structural equivalence and 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 1992) 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 1992).

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 (Alstyne et al. 2005; Burt 2004; Hargadon et al. 1997; Lazer et al. 2007; Reagans et al. 2001; Rodan et al. 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 et al. 1990; Reagans et al. 2001; Rodan et al. 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 1992).

The first paper to explore the information benefits to structural diversity uncovered a tradeoff between network diversity and channel bandwidth (Aral & 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'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 1992), 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, raising questions about whether vision advantages operated the way

Granovetter and Burt had theorized. However, Aral and Van Alstyne 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. 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. 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 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 ego's 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 the opposite – that cohesion promotes innovation (Obstfeld 2005; Uzzi et al. 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 novelty 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 one's 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 there for 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.

Methods & Data

Research 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 firm's 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.

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 five 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 what we observed. 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 firm's evolving corporate email network as described below. Figure 1 displays the largest connected component of this graph using aggregated data over the twelve months colored to distinguish communities identified by the Blondel (2008) community detection algorithm. This figure gives a sense of the distinct clusters of communication that exist in the firm's email network.

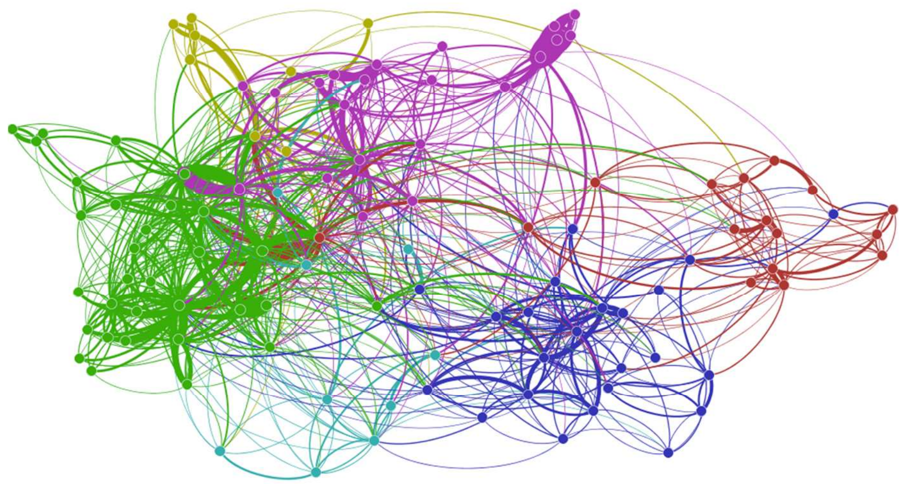


Figure 1.

Data

The quantitative data collection focused on two areas: (i) human resource information such as employees' age, gender, and date of hire, and (ii) internal email communications captured on the corporate email server for all employees. Overall we collected two million emails sent and received from 214 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 then stemmed, e.g. 'multitasking' and 'multi-task' become 'multitask'. We then calculated the keyword frequency for each email and keywords were replaced by unique hash codes ensuring anonymity of the actual content (see Aral & Van Alstyne 2011 for further details on email data collection and hashing). The full dataset was then processed using Latent Dirichlet Allocation (LDA) (Blei et al. 2003). This probabilistic generative model enabled us to classify keywords according to their frequency and co-occurrence across emails into a set of topics. Each message was then represented as a probability distribution over topics. We then applied our statistical methods to the lower-dimensional space of topics rather than the higher-dimensional and very sparse space of keywords.

Latent Dirichlet Allocation

Using LDA we identify a topic distribution vector $\mathbf{\Gamma_k}$ for each message sent (indexed k). The probability space is n-dimensional, such that:

$$\mathbf{\Gamma_k} = (\gamma_{k1}, \gamma_{k2}, \ldots, \gamma_{kn}).$$

This representation of each email represents the likelihood of the email being about a specific topic based on its content. While the content itself is anonymized and the topics are not further determined in terms of their actual meaning, LDA quantifies the co-occurrence of terms in exactly the same way it does with clear text. Based on the topical distribution of each message we then derive measures of information diversity, non-redundancy and uniqueness for the streams of emails within the company.

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 either with a given contact (dyad-level) or across all contacts (ego-level), (ii) the total volume of non-redundant information received 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 all of ego's 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.

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:

(i) The information diversity ID of all N messages Ego[1] (indexed e) received from all his peers. Let the set of these messages be $\mathcal{M}_e$. We summed the cosine similarity of the topic distribution vectors of all the email messages received by ego in a given time period $\mathbf{\Gamma}_k \in \mathcal{M}_e$ to their mean value $\bar{\mathbf{\Gamma}}_k$:

$$ID = \frac{1}{N}\sum_{k=1}^{N}[1 - \cos(\mathbf{\Gamma}_k, \bar{\mathbf{\Gamma}}_k)]^2, \qquad \mathbf{\Gamma}_k \in \mathcal{M}_e.$$

By this definition information diversity measures the variance or spread of topic distribution vectors across all received messages relative to their mean. A richer, more diverse set of communication will result in a higher ID, while very specific communications focusing on a small set of topics will result in lower ID. This is same procedure used by Aral and Van Alstyne (2011).

(ii) Information diversity *within* a specific dyad *IDW*. In this case we measured the information diversity of all messages between a specific sending alter and another receiving Ego. The definition is equivalent to that of ID, with the exception that we summed only over the N_i messages sent from a specific alter (indexed i). Let this set of messages be $\mathcal{M}_{ie}$:

$$IDW = \frac{1}{N_i}\sum_{k=1}^{N_i}[1 - \cos(\mathbf{\Gamma}_k, \bar{\mathbf{\Gamma}}_k)]^2, \qquad \mathbf{\Gamma}_k \in \mathcal{M}_{ie}$$

Information Uniqueness

Information uniqueness is quantified by measuring the distance of topic distributions *between* ties. Consider the cosine similarity between the aggregate topic distribution $\mathbf{\Gamma}_{ie} = \sum \mathbf{\Gamma}_k, \mathbf{\Gamma}_k \in \mathcal{M}_{ie}$ of tie (i, e) and all other $\mathbf{\Gamma}_{je}, i \neq j$ representing ties (j, e) and defined accordingly. Holding Ego e and alter i constant information uniqueness *IU* is defined as:

$$IU = \frac{1}{N_t - 1}\sum_{j=1}^{N_t}\left[1 - \cos(\mathbf{\Gamma}_{ie}, \mathbf{\Gamma}_{je})\right], \quad i \neq j$$

This variable quantifies how similar the information conveyed to ego by one contact is to the information conveyed to ego by all their other contacts. 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.

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

[1] We consider all communications from the point of view of the recipient, which we call Ego. All sending individuals from Ego's point of view are consequently called alter(s).

which quantifies the potential amount of information conveyed by a message given its topic distribution vector $\mathbf{\Gamma}$. An ideal measure for this purpose is information entropy $H(\mathbf{\Gamma}) = E[-\ln \mathbf{\Gamma}]$.

If we want to determine the amount of non-redundant information conveyed through messages along a tie (i, e), we want to account for all other information Ego (indexed e) receives through all other ties (j, e), with $i \neq j$ and control for redundancy in the information provided by those other ties. Let $\mathbf{\Gamma}_{ie} = \sum \mathbf{\Gamma}_k, \mathbf{\Gamma}_k \in \mathcal{M}_{ie}$ be the topic distribution vector for tie (i, e) and $\mathbf{\Gamma}_{je}$ for (j, e). We then define the conditional entropy CE of tie (i, e) to be:

$$CE = H\left(\mathbf{\Gamma}_{ie} | \mathbf{\Gamma}_{j_1e}, \mathbf{\Gamma}_{j_2e}, \ldots, \mathbf{\Gamma}_{j_ne} \right),$$

With the set of indices $J_{ie} = \{j_1, j_2, \ldots, j_n\}$ indexing all ties leading to Ego with the exception of $i \neq j$. Conditional entropy per tie measures the average amount of non-redundant information Ego receives from a specific alter i, given the topic distribution vectors of all other alters communicating with Ego. Any overlap in information (i.e. information redundancy) between (i, e) and any other (j, e) is thus discounted and CE measures only the fraction of truly non-redundant information provided by alter i.

If we further consider the total amount of non-redundant information Ego receives from all his sources, we find the joint entropy JE to be a suitable measure. Accounting for the overlap of information Ego receives from all ties(i, e), joint entropy measures the total average amount of non-redundant information which can be encoded given the set of topic distribution vectors $\left\{\mathbf{\Gamma}_{i_1e}, \mathbf{\Gamma}_{i_2e}, \ldots, \mathbf{\Gamma}_{i_ne}\right\}$. We therefore define joint entropy to be:

$$JE = H\left(\mathbf{\Gamma}_{i_1e}, \mathbf{\Gamma}_{i_2e}, \ldots, \mathbf{\Gamma}_{i_ne}\right).$$

We find both conditional and joint entropy measure the amount of non-redundant information provided to ego by a given contact and by all of ego's 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, consider the following examples: First, in considering information diversity we would 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 would be smaller. Second, considering information uniqueness, we propose that a set of messages originating from one source 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 raw 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 information received if and only if there are no redundancies, otherwise it is smaller.

Network Variables

Besides the information content conveyed through the communications network we are also interested in the structure of the network itself and its effect on the flow of information.

Network Size & Bandwidth

To capture the key elements of the communications network structure we use a set of basic measures, such as bandwidth and network size. Bandwidth per tie BWM is simply defined as the number of messages N_t sent across that tie during a given amount of time (i.e. monthly panels). Consequently Ego's average channel bandwidth is $\overline{BWM} = E[BWM]$, considering all incoming ties, over which Ego receives messages. Network size NWS is simply the number of contacts, from which Ego received messages (at least one, incoming) during a given time period.

Network Constraint

We use Burt's network constraint 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 broker's ego network but also to measure ego's 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 e in a specific alter i as p_{ei}. We denote this as his direct investment $DI = p_{ei}$. Further we consider secondary or redundant investments via mutual relationships (indexed j) in the communications network. This measure of redundant investment we denote as:

$$RI = \sum_{e \neq j \neq i} p_{ej} p_{ji} \,.$$

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:

$$NC = \sum_i (DI + RI)^2 = \sum_i \left(p_{ei} + \sum_{e \neq j \neq i} p_{ej} p_{ji} \right)^2 .$$

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.

Network Stability

As we have a panel data set of email communications over twelve months, we can investigate temporal effects of the structure and stability of the communications network on the flow of novel information. We identify two principle approaches to assess network stability.

We quantify the number of contacts that each employee retains from one time interval (panel) to the next. Let the set of active ties ego e retains in panel t be $\mathbb{L}_{t,e}$ and for the previous panel be $\mathbb{L}_{t-1,e}$. Furthermore we determine the intersection of these two sets to be:

$$\mathbb{I}_{t,e} = \mathbb{L}_{t-1,e} \cap \mathbb{L}_{t,e}$$

Consequently we define network stability NS as fraction of contacts, which are retained from one panel to the next:

$$NS = \frac{\left|\mathbb{I}_{t,e}\right|}{\left|\mathbb{L}_{t-1,e}\right|} .$$

Network stability quantifies the fraction of retained contacts from one time interval to the next.

Control Variables

We control for the effects of demographic characteristics of each information broker such as age, 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 account for two primary demographic factors: difference in dates of hire and difference in gender. We compute both demographic control variables on the level of individual ties and aggregate them (i.e. determine average values) for the ego networks of information brokers over all incoming ties of their respective ego network.

Model Specification

We use monthly panel data 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. We first estimated a set of random effects models at the ego level to examine the relationships between network structure and information diversity (*ID*) and total non-redundant information (*NRI*) as follows:

$$(B_{et}, NC_{et}) = \gamma_e + \beta_1 NC_{et} + \beta_2 B_{et} + \sum_k \beta_k C_{ket} + \varepsilon_{et}$$

$$(ID^I_{et}, NRI^I_{et}) = \gamma_i + \beta_1 NS_{et} + \beta_2 NC_{et} + \beta_3 S_{et} + \beta_4 B_{et} + \sum_k \beta_k C_{ket} + \varepsilon_{et}$$

The first set of equations models the diversity-bandwidth tradeoff for communication received by ego e at time t, while the second set examines the relationships between network structure and access to novel information. The variable ID^I_{et} represents the diversity of the information, NRI^I_{et} represents the total amount of novel information, NS_{et} represents network stability, NC_{et} represents network constraint, S_{et} represents network size, B_{et} represents total bandwidth, and $\sum_k \beta_k C_{ket}$ represents controls for demographic variables (Gender, Hire Date).

We then estimated three sets of random effects models at the dyadic level to examine the relationships between tie characteristics (e.g. embeddedness, strength of tie, bandwidth) and information diversity (*ID*), non-redundant information (*NRI*) and information uniqueness (*IU*) as follows:

$$(ID^I_{jet}, NRI^I_{jet}, IU^I_{jet}) = \gamma_{je} + \beta_1 DI_{ejt} + \beta_2 RI_{ejt} + \beta_3 B_{jet} + \sum_k \beta_k C_{kjet} + \varepsilon_{jet}$$

Where ID^I_{jet} represents the diversity of information sent from alter j to ego e at time t, NRI^I_{jet} represents the total amount of novel information, DI_{ejt} represents direct investment of ego e into alter j, RI_{ejt} represents redundant investment, B_{jet} represents channel bandwidth, and $\sum_k \beta_k C_{kjet}$ represents controls for differences in demographic variables (Gender, Hire Date) between alter and ego.

Results

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, we should see the bandwidth of their communication channels contract. Second, we should observe increases in 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.

Table 1.

	(1) Constraint	(2) Bandwidth
Constraint		-0.131*** (0.000)
Bandwidth	-0.126*** (0.000)	
N	2300	2300
chi2	655.0***	83.01***
R2	0.381	0.157
Controls	Hire Date, Gender	

p-values in parentheses
* p<0.05 *** p<0.01 *** p<0.001

We find strong evidence confirming the Diversity-Bandwidth Tradeoff. As employees communicated with contacts which are less well connected to each other the overall bandwidth of their communication channels to those contacts contracted quite rapidly. For instance, we estimate that a one standard deviation increase in structural constraint (NC, i.e. reduced structural diversity) was associated on average with a .13 standard deviation decrease in bandwidth (Model 2, Table 1). As networks become more diverse the thickness of communication channels narrow.

Table 2.

	(1) ID	(2) ID	(3) NRI	(4) NRI	(5) NRI
Network Stability		0.151*** (0.000)			0.210*** (0.000)
Network Constraint	-0.521*** (0.000)	-0.225*** (0.000)	-0.280*** (0.000)		-0.226*** (0.000)
Network Size	0.126*** (0.000)	0.326*** (0.000)		0.978*** (0.000)	
Average Bandwidth	0.0637*** (0.000)	0.136*** (0.000)	0.124*** (0.000)	0.0217*** (0.000)	0.151*** (0.000)
Hire Date Diff.	-0.0146** (0.006)	0.00847 (0.489)	0.00398 (0.518)	0.00121 (0.374)	0.00462 (0.672)
Gender Diff.	5.125*** (0.000)	9.192*** (0.000)	0.413* (0.017)	-0.180*** (0.000)	0.531 (0.223)
Gender Diff. (sq.)	-5.175*** (0.000)	-9.685*** (0.000)	-0.424* (0.011)	0.184*** (0.000)	-0.492 (0.239)
Constant	-1.123*** (0.000)	-2.041*** (0.000)	-0.148* (0.016)	0.0369** (0.004)	-0.270* (0.028)
N	1426	2300	1426	2300	2300
χ^2	1031.2***	4094.4***	853.8***	48817.2***	516.3***
R^2	0.391	0.714	0.430	0.979	0.420

p-values in parentheses
* p<0.05 *** p<0.01 *** p<0.001

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. Specifically, a one standard deviation increase in structural constraint is associated, on average, with a .52 standard deviation decrease in information diversity (Model 1, Table 2), while a one standard deviation increase in bandwidth is associated with a .06 standard deviation increase in information diversity. Network size yields a positive effect as well with a .12 standard deviation increase in information diversity.[2]

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 (Table 2, Model 3, $\beta = 0.28$). In contrast, network size (Table 2, Model 4, $\beta = 0.98$) and bandwidth (Table 2, Model 4, $\beta = 0.22$) are positively associated with NRI.[3]

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 new setting and with a new data set 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 Figure 2. Although ours is a much larger sample and therefore displays much smoother aggregated relationships, the overall patterns in the data are remarkably similar.

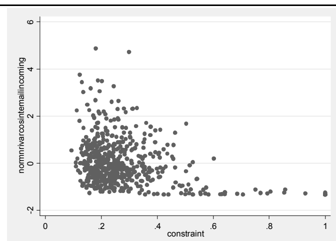
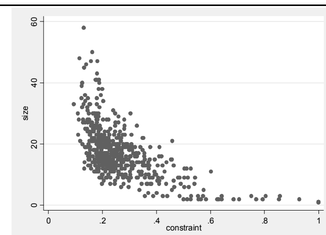
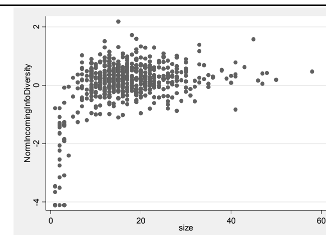
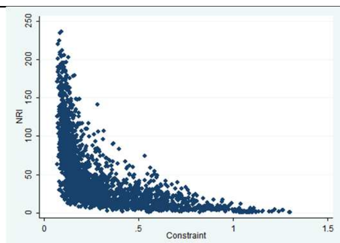
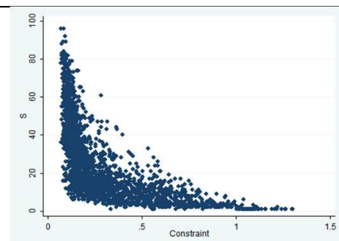
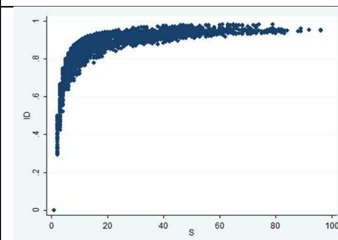
	Constraint and NRI	Network Size and Constraint	Network Size and ID
Aral & Van Alstyne 2011			
This Study			

Figure 2.

[2] Additionally we find an interesting gender difference effect on information diversity. The negative yield of the quadratic gender difference control variable (Table 2, Model 1, $\beta = -5.175$) indicates, that mixed gender communications coincides on average with higher information diversity (i.e. male communicates with females or vice versa). Holding the mixed gender difference effect constant, we further find the linear gender difference control variable having a positive yield on the information diversity between information brokers of the same gender.

[3] Due to significant colinearity between network size and constraint both variables have to be evaluated in separate models. In essence the amount of non-redundant information the broker receives is larger, in the case of a diverse network structure with a decreased number of structural redundancies, while scaling consistently with increased network size.

The 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 the different contributions of the constraint variable NC at the level of individual ties and analyze them separately. We distinguish the two separate terms of Burt's original constraint variable into its dyadic components: direct investment DI and redundant investment RI.

Specifically, we find a highly significant increase ($p < 0.001$) of information diversity *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 .19 standard deviation increase in information diversity within that dyadic channel (IDW) (Model 1, Table 3). This effect is shadowed by a similar effect in redundant investment ($\beta = 0.12,\ p < 0.001$). Together, these results indicate that the diversity of information Ego receives within a particular relationship increases with the amount of time and effort she invests directly and via shared relationships in his peers. Further we find, that higher bandwidth facilitates information diversity within ties ($\beta = 0.33,\ p < 0.001$).

Secondly, we consider the amount of non-redundant information conveyed to Ego per tie, as measured by conditional entropy (NRI). We consistently find positive trends for direct ($\beta = 0.04, p < 0.001$) and redundant investment ($\beta = 0.04, p < 0.001$) as well as tie bandwidth ($\beta = 0.13, p < 0.001$). Bandwidth has the largest impact with a .13 S.D. increase in non-redundant information for a one S.D. change in tie bandwidth. This is consistent and in support of our findings on information diversity within ties as well as the amount of non-redundant information Ego receives across all his peers (measured by its joint entropy *NRI*).

Table 3.

	(1) IDW	(2) NRI	(3) IU
Direct Investment	0.191*** (0.000)	0.0440*** (0.000)	-0.0549*** (0.000)
Redundant Investment	0.118*** (0.000)	0.0422*** (0.000)	-0.0608*** (0.000)
Bandwidth	0.328*** (0.000)	0.130*** (0.000)	-0.0910*** (0.000)
Gender Diff.	0.0109 (0.380)	0.0353* (0.013)	0.0355* (0.032)
Hire Date Diff.	-0.00147 (0.242)	-0.00151 (0.291)	-0.00656*** (0.000)
Constant	-0.216*** (0.000)	-0.112*** (0.000)	0.00401 (0.732)
N	53079	53079	53079
χ^2	8671.0***	4472.2***	5298.4***
R^2	0.193	0.101	0.122

p-values in parentheses

* p<0.05 *** p<0.01 *** p<0.001

Finally, when analyzing the information uniqueness between ties IU (with the receiving Ego as point of reference), we find a reversed scenario in regard to the effects of direct and redundant investment and bandwidth. Specifically, we find a decrease in the information uniqueness a tie delivers with increased investment (direct $\beta = -0.05, p < 0.001$, and redundant $\beta = -0.06, p < 0.001$, see Model 3, Table 3). This is further supported by the negative relationship of tie bandwidth ($\beta = -0.09, p < 0.001$) on information uniqueness between ties. In other words: weak bridging ties provide information which is on average increasingly distant from or unique when compared to the information provided by other ties.

Putting these results together, we can now describe the underlying anatomy of the vision advantage. A broker receives more diverse information and more total non-redundant information within ties from strong, cohesive ties. However, the information a broker receives from structurally weak bridging ties is on average more unique information, which is remote in topic space (i.e. different) compared to the information he receives from all other ties. On the other, hand the pairwise distance between the information provided by strong, cohesive ties in topic space is on average relatively small. This indicates that structurally weak ties provide potentially unique information, which is significantly different (distant) from the information provided by the core clique surrounding the respective information broker. At the same time, the information provided by a structurally weak tie will also be more specific (i.e. less diverse). Reflecting further on the information diversity and volume of non-redundant information the broker receives (Table 2) overall, we recall that both decrease in the case of highly cohesive or constrained networks. Altogether this indicates that the effects of IDW and IU are countervailing and the overall novel information received by Ego through all his peers is dominated 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 3, Table 3.

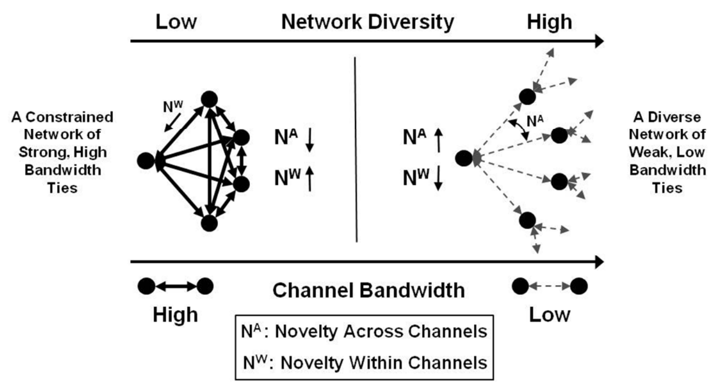


Figure 3.

These results together tell a very compelling story about how vision advantages work, which is graphically described in Figure 3. 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 at the same time novelty 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.

Concluding our analysis we finally investigated the temporal effects of network stability on the flow of novel information. Considering Models 2 and 5 in Table 2 we find, that network stability $p < 0.001$has a positive relationship both with information diversity ($\beta = 0.15, p < 0.001$) and non-redundant information ($\beta = 0.21, p < 0.001$). The effects of network constraint, size and average bandwidth remain qualitatively the same compared to Models 1 and 4 in the same table. This indicates that holding the effects of network structure constant more stable peer relationships over time coincide with higher overall information novelty. In contrast a high churn rate will lead to a reduction of information novelty. This provides supporting evidence for the argument that 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.

Discussion & Conclusion

We analyzed the structure and content of the complete dynamic email network of employees of a medium sized digital global 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 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 also uncovered the mechanics driving the Diversity-Bandwidth Tradeoff and highlighted differences in the 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, we found 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 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. All of these endeavors provide further evidence of the power of nano-level data to uncover social processes driving competitive advantages for information workers.

References

Alstyne, M. V., and Brynjolfsson, E. "Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities," Management Science (51:6) 2005, pp 851-868.

Aral, S., and Alstyne, M. V. "The Diversity-Bandwidth Trade-off," American Journal of Sociology (117:1) 2011, pp 90-171.

Aral, S., Brynjolfsson, E. & Van Alstyne, M. 2012. "Information, Technology and Information
Worker Productivity." Information Systems Research, Articles in Advance, March 12: 1-19.Aral, S., Brynjolfsson, E., & Van Alstyne, M. "Productivity Effects of Information Diffusion in Networks," Proceedings of the 28th Annual International Conference on Information Systems, Montreal, CA) 2007.

Baker, W. E. "Market Networks and Corporate Behavior," American Journal of Sociology (96:3) 1990, pp 589-625.

Barabási, A.-L., and Albert, R. "Emergence of Scaling in Random Networks," Science (286:5439), October 15, 1999 1999, pp 509-512.

Blei, D. M., Ng, A. Y., and Jordan, M. I. "Latent dirichlet allocation," J. Mach. Learn. Res. (3) 2003, pp 993-1022.

Bulkley, N., and Van Alstyne, M. W. "Why Information Should Influence Productivity," SSRN eLibrary) 2004.

Burt, R. Brokerage and Closure: An Introduction to Social Capital Oxford University Press. New York, NY., 2005.

Burt, R. S. Structural holes : the social structure of competition Harvard University Press, Cambridge, Mass., 1992, pp. viii, 313 p.

Burt, Ronald S. "Structural Holes and Good Ideas," American Journal of Sociology (110:2) 2004, pp 349-399.

Burt, R. S. "Information and structural holes: comment on Reagans and Zuckerman," Industrial and Corporate Change (17:5), October 1, 2008 2008, pp 953-969.

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

Coleman, J. S. "Social Capital in the Creation of Human Capital," American Journal of Sociology (94:ArticleType: research-article / Issue Title: Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure / Full publication date: 1988 / Copyright © 1988 The University of Chicago Press) 1988, pp S95-S120.

Ghose, A. H., S. Iyengar, R. "Network Stability and Social Contagion on the Mobile Internet," Draft, IOMS, NYU Stern School of Business) 2012.

Granovetter, M "The Strength of Weak Ties," American Journal of Sociology (78:6) 1973, pp 1360-1380.

Hansen, M. T. "The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits," Administrative Science Quarterly (44:1), March 1, 1999 1999, pp 82-111.

Hansen, M. T. "Knowledge Networks: Explaining Effective Knowledge Sharing in Multiunit Companies," Organization Science (13:3) 2002, pp 232-248.
Hargadon, A., and Sutton, R. I. "Technology Brokering and Innovation in a Product Development Firm," Administrative Science Quarterly (42:4) 1997, pp 716-749.
Lazer, D., and Friedman, A. "The Network Structure of Exploration and Exploitation," Administrative Science Quarterly (52:4), December 1, 2007 2007, pp 667-694.
Obstfeld, D. "Social Networks, the Tertius Iungens Orientation, and Involvement in Innovation," Administrative Science Quarterly (50:1), March 1, 2005 2005, pp 100-130.
Padgett, J. F., and Ansell, C. K. "Robust Action and the Rise of the Medici, 1400-1434," American Journal of Sociology (98:6) 1993, pp 1259-1319.
Podolny, Joel M. "Networks as the Pipes and Prisms of the Market," American Journal of Sociology (107:1) 2001, pp 33-60.
Reagans, R., and Zuckerman, E. W. "Networks, Diversity, and Productivity: The Social Capital of Corporate R&D Teams," Organization Science (12:4) 2001, pp 502-517.
Rodan, S., and Galunic, C. "More than network structure: how knowledge heterogeneity influences managerial performance and innovativeness," Strategic Management Journal (25:6) 2004, pp 541-562.
Tucker, C. "Network Stability, Network Externalities and Technology Adoption," SSRN eLibrary) 2011.
Uzzi, B. "Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness," Administrative Science Quarterly (42:1) 1997, pp 35-67.
Uzzi, B., and Spiro, J. "Collaboration and Creativity: The Small World Problem," American Journal of Sociology (111:2) 2005, pp 447-504.
Watts, D. J., and Strogatz, S. H. "Collective dynamics of /`small-world/' networks," Nature (393:6684) 1998, pp 440-442.