SANDIA REPORT
SAND2009-5975
Unlimited Release
Printed October 2009

Quantifiable and Objective Approach to Organizational Performance Enhancement

Examining Social Structure and Linguistic Content during Collaborative Group Work from a Network Perspective

Andrew J. Scholand, Yla R. Tausczik, and James W. Pennebaker

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Sandia National Laboratories

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: `reports@adonis.osti.gov`
Online ordering: `http://www.osti.gov/bridge`

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: `orders@ntis.fedworld.gov`
Online order: `http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online`

SAND2009-5975
Unlimited Release
Printed October 2009

Quantifiable and Objective Approach to Organizational Performance Enhancement

Examining Social Structure and Linguistic Content during Collaborative Group Work from a Network Perspective

Andrew J. Scholand, Ph.D.

Decision Support Systems
Sandia National Laboratories
Albuquerque, NM 87185-1243

Yla R. Tausczik, M.A.
Professor and Chair James W. Pennebaker

Department of Psychology
University of Texas at Austin
Austin, TX 78712

Abstract

This report describes a new methodology, social language network analysis (SLNA), that combines tools from social language processing and network analysis to identify socially situated relationships between individuals which, though subtle, are highly influential. Specifically, SLNA aims to identify and characterize the nature of working relationships by processing artifacts generated with computer-mediated communication systems, such as instant message texts or emails. Because social language processing is able to identify psychological, social, and emotional processes that individuals are not able to fully mask, social language network analysis can clarify and highlight complex interdependencies between group members, even when these relationships are latent or unrecognized. This report outlines the philosophical antecedents of SLNA, the mechanics of preprocessing, processing, and post-processing stages, and some example results obtained by applying this approach to a 15-month corporate discussion archive.

Acknowledgements

Special thanks to the Seniors' Council for funding this investigation. The work described in this report performed by The University of Texas (UT) at Austin was funded under Contract Number 26-3963-70.

Thanks also to John Ganter for numerous engaging conversations and helpful suggestions for improvement of this work.

The University of Texas Researchers thank the Army Research Institute (W91WAW-07-C-0029) for additional support.

Contents

List of Tables

List of Figures

1 Introduction

Sandia, like many large institutions, faces a tough future. Projected nuclear weapons funding is monotonically decreasing, and other areas of national security are not the exclusive purview of Sandia. Increasing efficiency and effectiveness is a critical priority if Sandia is to continue providing exceptional service in the national interest.

This research addresses technical means to make socially informed efficiency improvements, based on the fundamental assumption that organizational and interpersonal improvements are likely to provide the highest gradient of returns at a technically skilled organization like Sandia. Furthermore, much of Sandia's knowledge and expertise in successfully executing work quickly is not written down because this information is developed, shared, and acted upon in an operational context through informal conversations among and between groups of individuals. The fundamental research proposition is that digital records arising from interactions in the 'as-is' organization can be analyzed to create an approximate but meaningful representation of the work-centered social dynamics within the organization. 'Meaningful' in this context implies facets of information relevant to interpersonal dynamics, aspects of distributed cognition and group work, and the development of organizational power and control. By constructing an explicit representation of working collectives of Sandians, it will be possible to better understand how work is actually accomplished, which in turn enables effective systemic improvements.

This work combines linguistic analysis with social network processing to both predict underlying structural relations and retrospectively describe patterns of group interaction. Survey-based evaluations of the predictions indicated a high degree of accuracy in assessing two well-established components of group membership, friendship and consultation networks. Descriptive insights of the group under study match previous ethnographic findings about work at Sandia (40; 41). We believe the discovered quantitative descriptors and predictors will be reproducible in future studies since function words are not context specific and the University of Texas' linguistic categories have been shown to have consistent relationships over multiple studies.

This report is organized as follows. Section 2 provides an overview of the approach taken to quantify how work is accomplished and describes the work group under study. Section 3 briefly covers literature in domains relevant to this study. Section 4 describes aspects of the data source that are evident from a traditional social language analysis of the data. Section 5 describes the new methodology developed in this work extending social language analyses beyond attributional descriptions. Section 6 provides the results of applying this methodology, and Section 7 steps through the survey-based validation of these results. Section 8 discusses some areas for future work, and Section 9 concludes.

2 Technical Approach

2.1 Overview of Approach

It has long been recognized that organizations can significantly benefit from social network analysis, which measures and represents the regularities in the patterns of relations among entities (24). Three decades ago, Tichy et. al. (43) pointed to the stable patterns of interaction within the social groupings of an organization as especially suitable for analysis of the causes and consequences of these relationships. Work by Sparrowe et. al. (39) confirmed that measurements of social networks (both positive and negative) correlates to job performance (as reported by supervisors) in modern industry settings. Baldwin and colleagues (2) found similar associations in master of business administration (MBA) student teams. Hossain et. al. (17) showed a statistically significant relationship between network centrality in Enron email and project coordination. The strength of a knowledge transmission network between divisions in a company predicts time to complete a project (15). Finally centrality in an advice network, not job rank, predicts obtaining high status privileges such as acceptance, the ability to take risk, and information access (19).

Social network analysis, however, focuses primarily on the structure prescribed by the existence of links between entities. In many cases links are treated as being simply binary, namely being either present or absent. A deeper analysis of the state of relations between two entities based on the language used between them has traditionally been the domain of text processing. These text processing algorithms can be classified into two disparate categories, 'top down' and 'bottom up' approaches. Bottom up algorithms analyze the statistical co-occurrences of words and cluster words used together into similarity groups. An example of this kind of algorithm is Latent Semantic Analysis. In contrast, top down approaches attempt to categorize a document based on the use of words predefined to be in certain categories. Section 3 below discusses in further detail an example of this type of algorithm, the social language processing approach, which uses psychological categories to assess over 80 different relationship dimensions.

The key technical approach of this work is to combine the high fidelity assessment of relationships between entities made possible by textual analysis with the contextual framework of social network analysis. By selectively extracting, combining, and processing different psychological, social, and emotional linguistic markers it is possible to map the rich relationships within and across organizations, making difficult tasks such as managing organizational change, organizational design, and interorganizational relationships easier. This report documents our initial findings in these areas, and outlines areas for future research.

Early organizational studies relied on letters, memos, organizational charts, meeting minutes, survey data, interviews, and direct observation to provide data on the social networks of interest. Modern computer mediated communication technologies, however, enable knowledge-intensive collaborative work (22) while providing a rich record of how that work was accomplished. The next section describes one such corpus used in this study.

2.2 Data Source

The National Infrastructure Simulation and Analysis Center (NISAC) developed a programmable collaboration library to facilitate secure collaborative interaction by geographically distributed decision-makers. The collaboration framework offers the usual collaborative services (chat and file transfer) as well as the ability to publish multiple images for collaborative text and graphical annotation. These capabilities focus primarily on *synchronous* capabilities that allow the integration of multiple perspectives and quick convergence on a shared view of a problem to facilitate high-pressure, time-constrained analyses. Figure 1 illustrates the incorporation of these services into the NISAC Agent Based Laboratory for Economics (N-ABLE™) for use in various computational economic analyses.

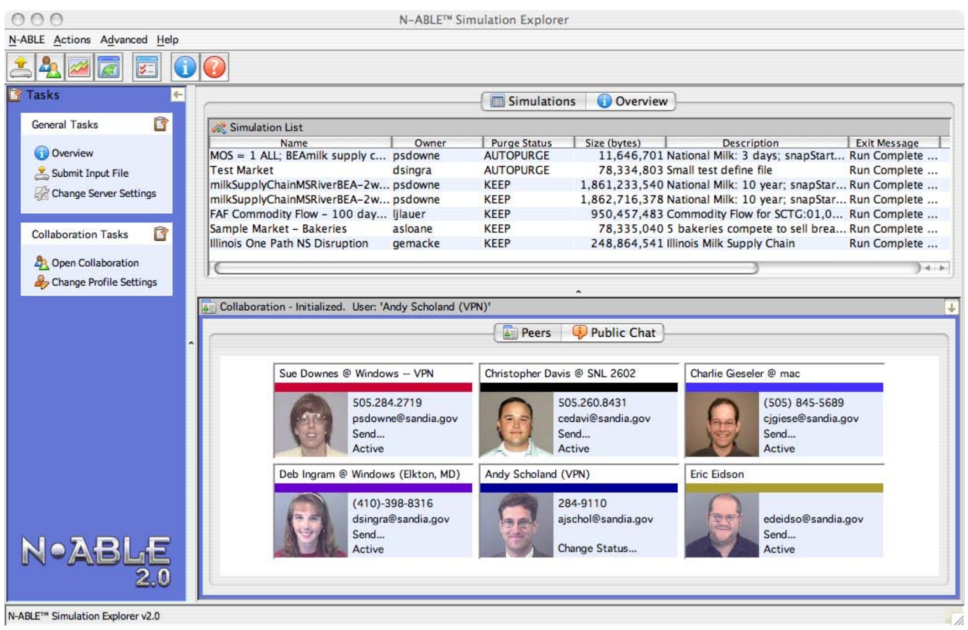

Figure 1: Integration of Collaboration Services in a NISAC Analysis Tool

The tab marked 'Public Chat' allowed instant messages typed by participants to be visible to all users currently logged on the system. Conversation frequently[1] centered on shared images, as shown in Figure 2, as well as other background information, such as geospatially referenced transportation data.

[1]See (27) for a exposition of the role of shared images in this environment as coordinating group sensemaking and consensus building.

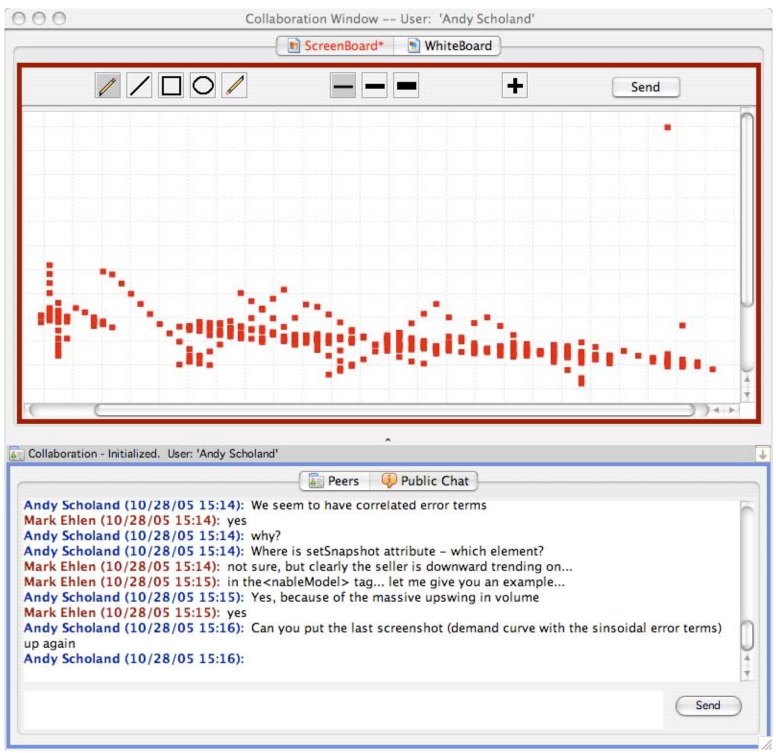

Figure 2: Example Chat Excerpt Related to Shared Image

This framework has been used since 2003 by the geographically distributed Computational Economics Group to plan, stage, execute, debug, and interpret high performance computing simulations of the national economy subject to regional disruptions. The group also used the tool to evaluate simulation initialization specifications derived from data fused across multiple government and commercial data sources. These work-related instant message conversations between 18 team members were collected for this analysis from September 2006 to November 2007.[2] These participants included 7 females and 11 males, varying in age from 22 to 64 years old. Four other chat participants were excluded due to contributing less than 250 words in public chat during the period of the study.

It is extremely important in social network analysis to appropriately determine the boundaries of the network under study, as errors can distort the overall configuration of actors in a system (19; 36). One of the strengths of this data source is that it is largely a self-contained system. In social network analysis terms, the system boundary is established by a realist strategy because the boundary is explicitly recognized by the participants[3] rather than being a perspective imposed solely for analysis.

Another strength of this data source is the unobtrusive way in which it was collected, an attribute shared by most electronic communication systems. Due to extensive monitoring opportunities, computer-mediated communication data is considered highly resistant to measurement errors (24) such as inaccurate recall, bias, and elicitation priming. The lack of an explicit audience (such as the author of a questionnaire) in such automatically recorded data is also important from a linguistics perspective, which holds that dialog changes to accommodate the addressed audience.

[2]The use of these data has been reviewed and approved by Sandia's Human Studies Board in Research Protocol SNL0806.

[3]In an anonymous survey, 71% of respondents agreed with the statement, "Participating in collaboration provides me with a sense of belonging and group identity within the CEG team." In a separate, non-anonymous survey 87% agreed to some degree with the statement, "Group chat increased the sense of community within the group." See Tables 18 and 21 in Appendix D for these and other survey responses.

3 Background

3.1 LIWC Background

This work leverages social language analysis performed by the Linguistic Inquiry and Word Count (LIWC) text analysis software program developed by University of Texas researchers James W. Pennebaker, Roger J. Booth, and Martha E. Francis. LIWC is a program for quantitative text analysis that uses a word count strategy for both the analysis of content (what is being said) and style (how it is being said). Word count strategies are based on the assumption that the words people use convey psychological information over and above their literal meaning and independent of their semantic context. In this sense, they are "top down" in that they explore text within the context of previously defined psychological content dimensions or word categories. (In contrast, word pattern strategies such as latent semantic analysis mathematically detect "bottom-up" how words co-vary across large samples of text, typically to determine the degree to which two texts are similar in terms of their content.) LIWC searches for over 2300 words or word stems previously categorized by independent judges into over 80 linguistic dimensions. These dimensions include standard language categories (e.g., articles, prepositions, pronouns– including first person singular, first person plural, etc.), psychological processes (e.g., positive and negative emotion categories, cognitive processes such as use of causation words, self-discrepancies), relativity-related words (e.g., time, verb tense, motion, space), and traditional Freudian content dimensions.

The use of LIWC allows for the indirect measurement of various attributes of interest, based on the robust premise that word use reflects basic social, personality, cognitive and biological processes. Certain LIWC categories are strong markers for specific psychological behaviors. The relative[4] use of first person singular pronouns is "a particularly robust marker of the status of two people in an interaction" (5). The relationship between use of first person singular pronouns ("I-words") and status is an inverse relation; in a conversation between two individuals, the person with the lower use of I-words tends[5] to be higher in relative status. Higher status individuals also tend to use more first person plural pronouns (37). Cognitive mechanism words (e.g. cause, know, ought) are often used to make causal statements or reappraisals. These words can show increased cognitive complexity (42).

Function words also indicate important emotional dimensions of social relationships. Groups that used more positive emotion words and more frequently used assenting words in reply to dissenting responses had less negative interpersonal behavior and better team performance (13). Similarly, successful coalitions of negotiating business students used more assent words than pairs or triads who did not form an alliance (18). However, frequent use of assent words by an individual can

[4]The relativity of use is important, as baseline rates have been found to correlate to individual attributes such as age, gender, culture, and psychological health.

[5]LIWC embodies an inherently probabilistic approach to social language analysis in the interests of computational efficiency.

alternately indicate passivity and acquiescence. Leshed and colleagues (26) found that individuals in a small group engaged in a collective task that used more assent words were rated by group members as being less involved and not as task focused.

3.2 Related Work on Communication

Previous research has investigated how instant messaging has been used in a work place. Isaacs et. al. (20) recorded 21,000 instant message conversations between 437 dyads discussing work and non-work related topics. The authors set out to describe the functional uses that instant messaging plays in work place. They coded whether a subset of 500 conversations included statements for the following functional categories: simple questions and information, work related, scheduling and coordination, personal, saying "hi", and no response. While the majority of conversations pertained to work, none of the other categories made up more than one third of the conversations.

One of the advantages of instant messaging in the workplace may be better group communication. Scholl et. al. (35) survey participant attitudes toward communication through both chat and audio channels. Their participants preferred communicating though chat because it is both asynchronous and synchronous, it creates a permanent record, there is more time to think between turns, and communication with a large group is easier. Communicating using instant messaging in a large group is easier than an audio channel because of fewer problems with turn taking and collisions between multiple conversations at the same time.

Observing communication and problem solving in informal group communication informs the study of real world problem solving. Shin et. al. (38) argue that different skills are needed to solve open-ended problems, including more emphasis on the regulation of cognition, specifically the meta-cognitive phases in planning how to problem solve.

Hirokawa (16) noted common phases that groups go through in problem solving: orientation, problem solving, conflict, and decision emergence. He compared small group decision processes in an open ended traffic control problem. He found that groups classified as successful made more procedural statements at the beginning and end of discussion and produced task-oriented statements later. Unsuccessful groups, in contrast, were task-oriented early on and made procedural statements in the middle of discussion. Hirokawa concluded that more successful groups analyze the problem before generating and evaluating solutions. Similarly, Artzt and Armour-Thomas's (1) work with middle school children working in groups to solve a math problem showed the importance of meta-cognitive group processes interwoven with cognitive behaviors. In their study, fewer metacognitive statements presaged failure to reach a solution. The social dimension of small group problem solving was also highlighted in this work, with the attitudes (positive or negative) of high-ability students affecting the problem-solving behaviors of the group.

3.3 Related Sociolinguistic Work

Work in the sociolinguistic field has also combined network analysis and linguistic style to understand linguistic variation with respect to social position (29), although with the social position provided from external sources. Eckert (11) describes the tradition in sociolinguistics of studying style variation in relation to “social categories of socioeconomic class, sex class, and age,” at work in “ethnographic studies of more locally-defined populations” and “as a resource for the construction of social meaning.” Sociolinguistic work has mostly examined variation in pronunciation in spoken language, as opposed to sensory-depleted use of written language in both synchronous and asynchronous communication. Rubini and Semin (34), however, have examined language use in relation to membership in groups. Their case study examined members of both the Communist party and Catholic church describing similar activities, such as reading a group specific newspaper or attending group specific meetings. They found individuals used positive and generalizable terms when describing group congruent behaviors to enhance in-group identity. In contrast, incongruent behaviors were described in concrete terms to particularize them. Maass et. al. (28) similarly found that people communicate desirable in-group and undesirable out-group behaviors more abstractly than the converse behaviors. The behaviors under question were not social norms specific to the groups (local sports team affiliations), but were behaviors generally deemed socially desirable.

3.4 Related Computer Mediated Communication Work

Recent research has demonstrated substantial organizational value to social network informed information artifacts constructed from computer mediated communications. IBM’s Atlas for Lotus Connections, a commercial implementation of Ehrlich’s work (12), is advertised as helping users spot connections and relationships between various groups in their personal and corporate networks. Specifically, Atlas provides a visual indication of the important hubs among topic experts and informal groups that have developed while working on similar projects. Users can then identify communication gaps or bottlenecks between groups and manage skills across the organization. Atlas also illustrates how a user is connected to any given expert in the organization, facilitating approaching an individual to form a connection. Joan DiMicco’s work at Sun Microsystems (10) similarly allows visualization of organizational expertise, again allowing one to consider the best path through one’s contacts to an introduction to a given expert. To date, however, these approaches seek to leverage existing structures (co-authorship of papers, organizational and seating charts, email ‘from’ and ‘to’ headers) rather than build them from content analyzed with a particular theoretical viewpoint.

Jonassen and Kwon (21) studied problem solving in the context of computer-mediated communication and found that there are differences in group processes depending on the type of problem. They contrast ill-defined problems from well-defined problems, suggesting that the majority of problems outside of the classroom are ill-defined problems. An ill-defined problem is characterized by an unclear goal and multiple potential solutions. Jonassen and Kwon tested the effectiveness of face to face and computer mediated communication in solving both types of problems in a group. They found that participants rated themselves as more effective using computer mediated communication to solve ill-defined problems, and there were fewer non-task personal statements and both more agreement and disagreement in computer mediated communication. They cite past research showing face to face to communication is favored because it engenders social processes, and found more non-task personal statements do occur with face to face communication. Jonassen and Kwon conclude computer mediated communication facilitates more critical discussion, communication of ideas, and better decisions. Additionally the authors found that in computer mediated communication there is an iterated process in which the group goes through the sequence of problem inspection to solution evaluation repeatedly, whereas in face to face communication these steps only occur linearly once.

Paolillo (30) studied a community of Asian Indians on Internet Relay Chat, an early manifestation of Computer Mediated Communication. He found that social cliques formed very quickly, and that vernacular usage was highest by peripheral and newly admitted individuals attempting to integrate themselves into the group.

4 Baseline Analysis

4.1 Content Analysis

Addressing the nature of information being communicated and the type of work being conducted in chat requires understanding the content of topics of discussion. To analyze the major themes in the conversation corpus, the University of Texas researchers have developed a method they term the 'Meaning Extraction Method' (6). This approach focuses on the co-occurrence of content words (i.e. adjectives, adverbs, nouns, and regular verbs) as a complement to the function word focus of LIWC.

The text data were grouped into synchronous conversations.[6] Conversations with fewer than 100 words were discarded, leaving 304 substantial conversations for analysis. The text processing program WordSmith was used to generate an exhaustive list of the most frequent non-function words in these conversations. This list was then reduced in two steps. First, uncommon words, defined as words that were not in at least 10% of the selected conversations, were removed. Second, from the remaining common words we removed any symbols, references to people, and condensed repeated word forms (e.g. thought, thinking, thinks). The final list consisted of 105 word stems or lexemes. For each of these word stems we recorded whether each conversation included it, generating a binary matrix of 105 word stem items by 304 conversational observations. The binary variable indicated the presence of each item in each conversation. Principle component analysis of this matrix indicated three factors had an eigenvalue greater than 1. We used verimax rotation on three factors and viewed only those items with loadings greater than 0.30, eliminating 21 word stems. The remaining 84 word stems are listed in Table 1, grouped by factor.

The first factor we call the social coordination of work. It includes social niceties (e.g. lol, hehe), affirmations (e.g. good, yeah, great, cool), actions coordinating people (e.g. call, meeting, chat, send), and descriptions of the communication of ideas (e.g. http, show, read, thinking, interesting, question). In conversations with high scores for this topic, participants planned future times when they would discuss work in detail. Social coordination is important in arranging detailed work discussions for this group, and is one of the main uses of the public chat forum. Past research has found that instant messaging is useful in setting up communication in other mediums (20; 35). The fact that instant messaging is informal and can be used asynchronously as well as synchronously enables individuals to arrange future communication without interrupting important current work. Public chat is also useful in coordinating a group because many individuals can be informed simultaneously again with minimal disruption. In this excerpt individuals share information about the cancellation of a group meeting that was going to be held in-person. Word stems matching this social factor are highlighted in red.

[6]The methodology for aggregating individual chat statements into conversations is described more fully in Section 5.4.

Table 1: Meaning Extraction Method Topics

Factor	**Word Stems**
Social / Alignment	yeah, lol, hehe, question, good, stuff, hear, people, guys, talk, idea, sounds, kind, nice, true, interesting, point, set, guess, pretty, great, work, time, show, bad, big, hard, thing, call, lot, read, thinking, cool, add, send, meeting, sense, chat, remember, real, http
Work-related Theory	firms, demand, production, BEA, market, supply, define, simulation, number, buy, based, case, means, results, view, answer, analysis, cost, reason, change, day, state, hmm, find, problem, fixed, report
Work-related Implementation	N-ABLE, streamer, file, runs, client, running, code, test, fine, machine, version, email, small, long, data, current

Person C	FYI the department meeting was canceled . . . Person P had a last minute need to cancel . . .
Person B	ahhhhhhhh - it's been cancelled 200 times!!!!
Person B	I just called her about it and she knows. . .
Person B	CSU is on it
Person C	I know . . . I called her too.
Person B	LOL

In another conversation, Person A and Person C coordinate future planning in a different medium through public chat.

Person A	OK cool
Person A	... I'm going to start thinking through an outline. Let me shoot that over in 20min and we can see if that is what you're thinking about too. We can iterate as needed.
Person C	Great!

These conversations are laced with social niceties and affirmations like "lol", "cool", and "great" which ensure messages are interpreted positively. Participants also share information and opinions that form relevant background for their work. In these two adjacent conversation several individuals discuss articles from the theoretical literature. By sharing these resources in public chat, the group's situational awareness is improved as information about what individuals are working on, and with whom, is available. Since public chat is archived, using this channel also ensures that these references are saved for future reference if needed.

Person B	Hey - did you read that paper yet? I am still on page 3 but it is real good so far.
Person D	yeah I had read it before
Person B	ah - that was a sweet find by Person G
Person D	yes I agree
Person G	I am looking for this.........Has anyone seen any data on the percent of the US population with internet access? by income level or geographic area or occupation? do you have suggestions where I might look?
Person I	try: http://www.infoworld.com/articles/hn/xml/01/02/19/010219 hnsurvey.html?p=br&s=5
Person I	i am looking the specific article now
Person G	thank you I am looking also

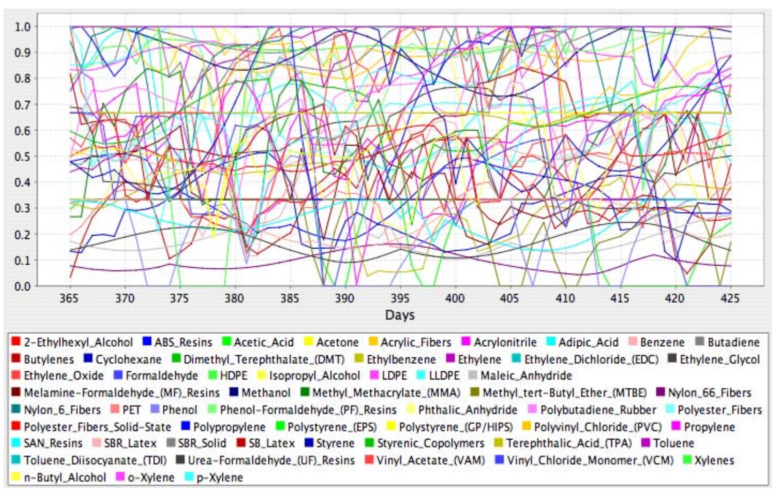

Figure 3: Image Used for Group Evaluation of Chemical Supply Chain Model

The second factor we call work related theory due to its direct relation to the economic (e.g. production, market, supply) and analytic (e.g. results, answer, problem, report) aspects of the work. As an example of this type of combined analytic theory work, Figure 3 is an image shared in collaboration to assist with the group evaluation of both the accuracy of the data processing used to create a supply chain representation and the realism of the resultant firm-level behaviors. Concrete results such as those in Figure 3 are compared to expectations based on economic theory. Conversations that scored high on this topic usually involved fewer participants working on a specific task. These conversations might carry over several days and include long pauses between adjacent conversations. The discussions were highly focused on problem solving, with individuals combining knowledge to resolve problematic issues. In this excerpt Person A and Person C are in the midst of a series of several conversations about the same topic. Person C leaves a series of messages about the problem while Person A is away.

Person C	I have spent the last little while checking out unmet demand and consumer surplus on the 365 day run (no disruption). I am using seafood as an example here. Given that the supply-demand ratio is 2.5 (lots of excess supply) we shouldn't have any unmet demand theoretically. However the unmet demand nationally for seafood is 2.2 million. There are two-three DEFINE file issues that could cause this. One is mismatched region names (Person F said he found some wierdness there so maybe tomorrow afternoon we can figure out if this is the issue). The second is too small of a constraint on the maxPreferredSeller list size the buyers can't call enough people. A possible third is that buyers are running out of time during the work day to make enough calls to find what they need. Its the dreaded PrefSellers list rewrite we need! Anyway since none of the firms are snapshot in the current runs I can't look at the results data to figure this out immediately. So I am going to 'hack' the DEFINE file and make the max list size huge and see what happens. New run available in a while.

There may be a few advantages for holding these work discussions over public chat. One, individuals peripherally involved in the problem can listen and participate, and at a minimum know that the conversation took place. For example, at the beginning of one conversation between Person A and Person D, Person A says, "Hi Folks Person D and I are going to be discussing the fixed cost accounting Person D has been working on for a while.", informing everyone of what Person D has been working on and giving others the opportunity to participate or see the final solution. Two, discussions in chat did not require geographic co-location of the individuals involved. Group members had offices in different buildings, and sometimes people worked from home, especially when working during the weekends. Three, it allows both individuals to use their own tools and separately view the elements of the work in their own way. Four, conversations can carry over the entire day, and through multiple days with necessary disruptions. Isaacs (20) found that the most frequent users of instant messaging use it off and on throughout the day. This same pattern is in this chat archive. In the previous conversation Person C was able to carry on in a conversation with Person A while Person A was attending to something else. Often these discussions would be interrupted by meetings and calls from other people at work, but the conversations flowed naturally and the person who was not engaged could continue with work. In this example, problem solving flowed seamlessly through a phone call.

Person D	yes fixed (and variable) costs do get reported with my fixes
Person A	phone - one sec
Person D	ok
Person A	OK I'm back
Person D	interesting. The Accountant's use of storage costs appears to be only in relation to the commodity markets

The in-depth conversations relating to this theory topic illustrate that integrating knowledge and problem solving is a core element of the work being conducted via chat.

The third topic, which we call work-related implementation, covers details such as the software (e.g. N-ABLE), computer equipment (e.g. machine, file, client), and programming (e.g. code, version, data). These conversations about tools were short and often involved solving a technical problem, testing of software and programming, or discussing the relative value of different technologies. For this group, trying to fix a technical problem is conducted very differently from trying to address a theoretical problem. In solving a technical problem multiple team members suggest solutions, work on finding a fix, and/or help test to see what the problem might be. In this conversation a problem with the simulation software arises:

Person C	For those working on Katrina; I have submitted a new run. The DEFINE file is still being verfied so please be patient.
Person C	rats a run-time error!
Person C	anyone know what this error is...?
Person F	which?
Person C	Hold on...
Person F	did we lose the streamer?
Person C	looks like it..
Person C	FirmBuyer.cpp:679: failed assertion 'total != 0.0'
Person F	I'll look it up.
Person C	thanks!

In this conversation, Person B, Person E, and Person H all try to provide suggestions for Person A's problem with the simulation software.

Person A	Has anyone else noticed excessive CPU usage in the client when the streamer dies?
Person B	no but it does seem to be a very Mac thing for the CPU to spin up on a dead/thrashing application
Person E	well if it's thrashing ...
Person A	yeah why is that?
Person B	it's busy looking for the right mouse button? Or perhaps the 'any key'? ;-)
Person E	you could try attaching the ThreadViewer to the NABLE client and see which thread is really busy
Person H	Or use JProfiler...which is much more precise.

Troubleshooting and bug elimination in computer tools are activities well suited for public chat. When a problem arises individuals can pitch in to help solve it, communicating about what they are doing and finding as they test and try out solutions. As each available and interested person claims an aspect of the task without centralized assignment, work is divided up efficiently with little managerial overhead. Also, because several people typically work on this class of problem at once, expertise from multiple individuals is employed at the same time and numerous ideas can be pursued at once. In addition to the effectiveness and rapid response of this 'bottom up' problem solving, the public nature of the discourse serves a cueing function for the group. Due to the centralized nature of the computing resources for this group, technical problems that arise could effect work across the team. Some of these problems are latent, in that their symptoms of failure are non-obvious, and widespread in their impact on simulated outcomes. Problem announcement through public chat helps inform everyone when a potential problem is identified and raises the level of vigilance for anomalous results.

4.2 Topical Content and LIWC Categories

Combining the Meaning Extraction Method topics with LIWC's function word categories allows us to address questions of how patterns of style, such as affect words, pronoun use, and text properties such as words per message, change depending on the topic being discussed. Understanding this relationship can tell us what other factors may be necessary in facilitating discussions on these topics. For example, positive affect words may be heavily used in conversations about social organizing to build group coherence and create a positive work environment. Relationships may also be correlational rather than causal, for example more words per message may be necessary for theoretical work conversations because the information being conveyed is both complex and dense.

We examined conversations with at least 500 words because in our text analysis experience these contain enough words to exhibit reliable patterns. Conversations with lower word counts can create misleading artifacts in understanding the language being used. It should be noted that conversations with at least 500 words are necessarily more involved conversations and may not be representative of typical chat communication. Only a small subset, 85 of the 1013 conversations, have at least 500 words.

Table 2: Correlations Between Topics and Language Categories

LIWC Category	**Social / Alignment**	**Work-related Theory**	**Work-related Implementation**
Apostrophes		-0.26	
Articles			0.25
Assent words	0.32	-0.32	
Average response time between messages	Neg		
Average words per message		0.29	
Cognitive mechanisms		0.55	
Exclamation marks	0.22		
First person singular		-0.32	-0.26
Future tense verbs			0.39
Number of speakers			Neg
Past tense verbs		-0.33	-0.31
Positive emotion	0.42	-0.23	
Second person	0.23	-0.35	
Six letter words	-0.34	0.50	
Social	0.50	-0.40	-0.35
Third person plural			-0.28

Table 2 enumerates the statistically significant co-occurrence correlations between the percentage of words from each of the three topics and percentage of words from LIWC's language categories. Although the Meaning Extraction Method extracts only non function words, some of the downselected topic words overlap with the language categories by a few words (e.g. "yeah" is an assent word and a social/alignment topic word). Conversations that scored high in the social/alignment topic typically also had many exclamation marks and positive emotion words, suggesting active positive dialogue. Increased use of second person pronouns in this topic suggests that individuals engaged directly with each other. Assents are an important component of the social/alignment topic, because the purpose of a conversation confirming scheduling or discussing the broad outlines of work is often to reach an agreement. There is a negative correlation between the social/alignment topic and the average response time between messages, suggesting that individuals were quick to respond to each other in social conversations. Individuals may respond quickly to show agreement and enthusiasm. Quick responses, like the increased use of positive emotion and assent words, may represent efforts to create group cohesion. In addition, social statements and questions may be easier to respond to than more work-related queries. The language categories correlated with both social/alignment and work-related theory (assent words, positive emotion words, second person pronouns, six letter words, and social words) are all correlated in opposite directions with both topics because the two topics are negatively related to each other.

Work-related theory conversations were positively correlated with the use of six-letter words and cognitive mechanisms, indicating that these conversations were linguistically complicated and intellectual. In addition there was a positive correlation with the average length of a message, perhaps because these conversations address more complicated ideas that can only be communicated through longer messages. Past tense verbs, first person singular, and apostrophes were all negatively correlated with work theory discussions. The negative correlation with apostrophes suggests that work theory discussions are more formal than other conversations. The reduction in first person singular pronouns may also contribute to a reduction in apostrophes because many contractions use first person singular. The first person singular pronouns are found less often because the typical conversational focus is on the problem being discussed and solved. It also suggests that in discussing work problems individuals do not overuse hedging statements such as "I think," that is they readily make suggestions to each other. This indicates the working relationships between team members engaged in work-related theoretical conversations are typically egalitarian, i.e. non-hierarchical.

The most interesting correlations for work-related implementation discussions of tools, computers, and programming are the positive correlation between this topic and future tense verbs and articles, and the negative correlation between this topic and the number of speakers. Future tense verbs may be used more in combination with discussions of tools because individuals are stating intentionality (what they are planning to do next) or when a problem arises, proposing next steps as a solution. Heavier use of articles suggests that these conversations are more concrete, which we expect because they are often about specific tool use or concrete problems that arise. The negative correlation between number of speakers and this topic could arise in two ways. More people may join a conversation on a computer or programming related problem because they feel they can contribute a valuable experience or perspective and assist in fixing the problem. Alternately, individuals may ask more questions that elicit a tool related discussion when more people are present because they believe practical help from a larger group will be more useful.

Due to the interaction between the number of speakers and the work-related implementation topic, we wished to investigate how the number of speakers varied across the conversations. As shown in Figure 4, of the total 1013 conversations, approximately 500 had only one participant, slightly less than 300 had two participants, and less than 150 had three participants – roughly an exponential decrease. This distribution suggests that the number of participants is being driven primarily by the availability of participants who are independently deciding whether or not to participate in the public chat at any given time.

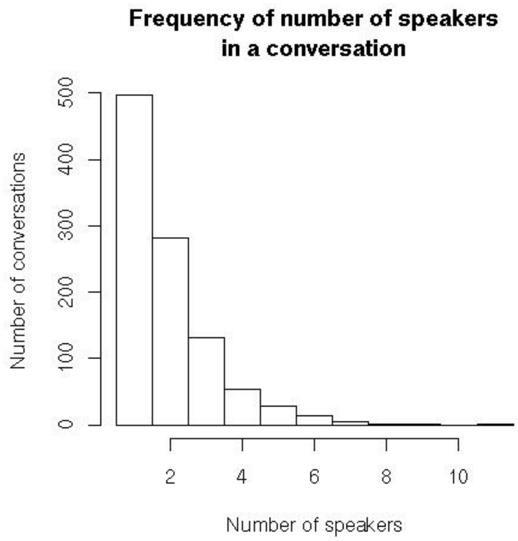

Figure 4: Frequency of Conversations by Number of Participants

Table 3 lists the correlation between the number of speakers participating in a conversation and several other LIWC measures. The lower use of prepositions in larger groups suggests less concrete thought. There is a positive correlation between the number of speakers and social words, question marks, and all punctuation. We believe the question mark correlation is due to the fact that question marks are a leading indicator of participation, i.e. a question acts as an initiator for a conversation. Increasing numbers of questions provide more opportunities for different group members to join the conversation. The social word correlation suggests that in larger discussion groups the conversation tends to be more relationship than work oriented. We hypothesize this represents a intragroup bonding effect. Specific work assignments tended to be tasked out to small subgroups within the overall group structure, so the larger the set of conversationists, the less likely they would have specific work items in common to discuss. In the absence of work topics but under the aegis of a common team, social pleasantries and exchanges are likely.

4.3 Content Variation Over Time

Public chat focused on the topics identified in Section 4.1 to differing degrees over time. Since each of the three main topics (social, theory, implementation) represent important aspects of work for this team, we are interested in tracking how conversation moved between topics over time, from hours to months. Topical content of conversations were quantified by the percentage of words from each of the topic factors.

Table 3: Correlation Between LIWC Categories and Number of Chat Participants

LIWC Category	**Number of Speakers**
All punctuation	0.26
Prepositions	-0.27
Question marks	0.30
Social words	0.28

On a month-based timescale no significant trends in topics emerged, as shown in Figure 5. All three topics were consistently discussed, although social organization conversations occupied a slightly higher percentage of all discussions. (The percentages shown in Figure 5 are relatively low because function words, which are excluded in this content-based analysis, represent the majority of words in chat.) Fluctuations in the degree to which each topic was discussed in a given month were found to be negatively correlated. The strongest inverse relationship was a negative relationship between conversations about in-depth theoretical work and social organization. Since each topic only classifies a small percentage of the total number of words in a given conversation, each less than 10%, in principle both social organization and in-depth work discussion could take place in the same conversation. The emergence of negative relationship therefore suggests that there is a group-level trade-off between these two topics, perhaps representing the 'locus of attention' of the group. In contrast, implementation discussions are independent of both social and theory conversations. We hypothesize this is because these discussions are prompted by the essentially random occurrence of problems arising in the course of executing work.

On a day-based timescale, differences between weekday and weekend chat usage become important. Instant messaging use on the weekend was primarily asynchronous, in that it was used to leave work messages for team members to see when they next logged onto the system. Weekend use was also occasionally affected by intense conversations around issues that needed to be addressed before the start of the next work week. Weekdays, compared to the weekend, had higher assent words, question marks, second person pronouns, and social words. More people participated in public chat during the work week, creating a more socially oriented conversation pattern. Weekend speech has language markers that demonstrate a higher proportion of work related conversation. Figure 6 shows there is a significant decrease in social organization discussions as the week progresses from Monday to Sunday and an increase in work related conversations. Social organization may occur early in the week because the discussion centers around what meetings might take place, and what progress has been made. In depth work conversations may increase as the work week continues as problems arise and individuals settle into a work routine.

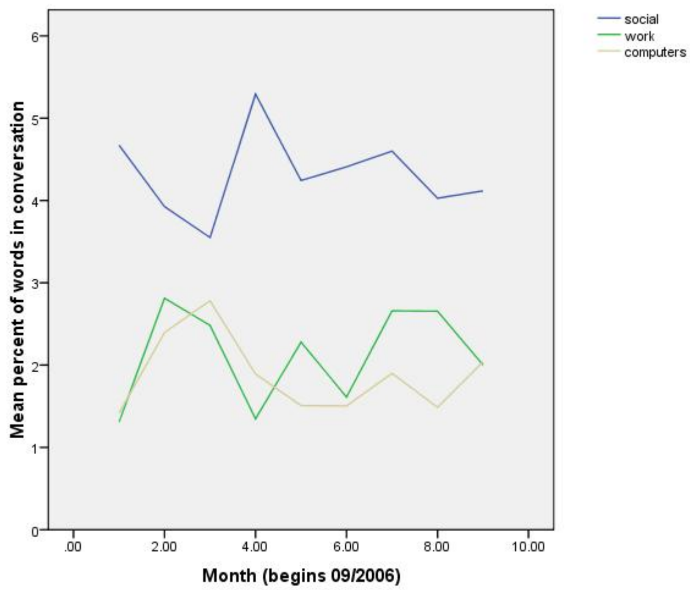

Figure 5: Main Themes in Chat Over Time

We observe the same interplay between conversations centered on work-related discussions and those concerning social/alignment across the hours of the day as we observed across the week. As the day proceeds from around 7 a.m. when participants begin working to the early hours of the next day, conversation shifts from social to work (see Figure 7). This may be driven by the fact that alignment and organization is best accomplished when many people are participating at the beginning of a work day, and conversation lingers into the night only when individuals have important problems that they want to discuss. The repetition of the counter-cyclical pattern of social alignment and work-related conversations over both the course of the day and the week highlights the fact that these two discussion topics represent two important and intertwined steps in the iterative process of completing work.

Comparing words used across the entire day shows that there is a significant increase in the average words per message in later, off-hours and a decrease in third person plural pronouns. These two findings match with language patterns for the days of the week. They suggest that during work hours individuals engage in conversations in which they make reference to others, and during non-work hours messages may act as a message board rather than a conversation. For example, one night at 4 a.m. Person A posted the following two messages when no one else was participating:

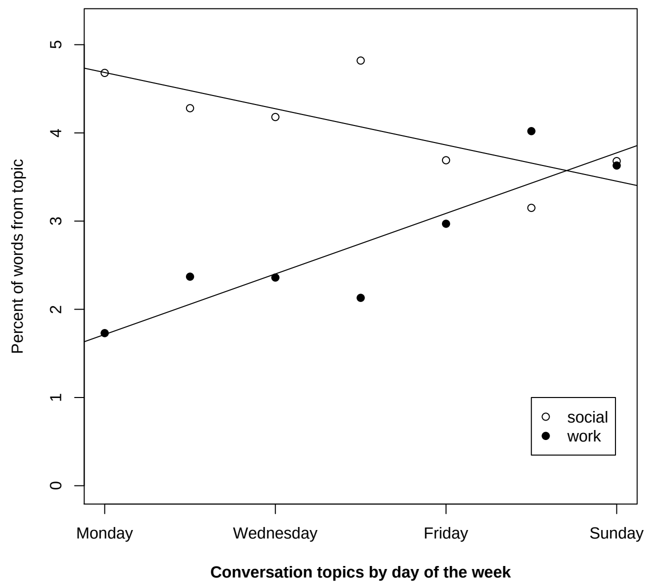

Figure 6: Trends in Social and Work Topics Over Days of the Week

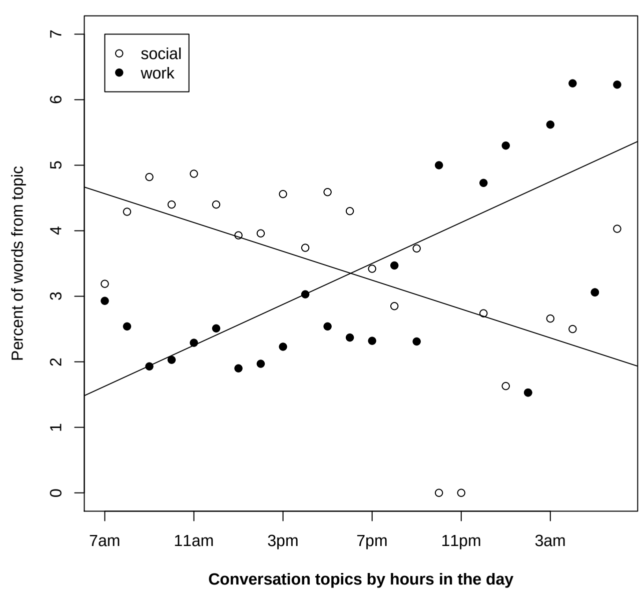

Figure 7: Trends in Social and Work Topics Over Hours of the Day

Person A	Overall however it seems that none of the individual market attributes adequately describe 'criticality' as demonstrated by unmet demand in an N-ABLE simulation.
Person A	I will post some of these analyses on the wiki. If you are logging in on a Friday and any of the above analysis has scrolled out of your review buffer you can send the archivist a message 'replay' and she will send you back all of the text I have typed in this morning.

The second message makes it clear that Person A is posting this message so that others can read about the work he is doing in the morning.

We also found some significant trends in language use over the course of the work day, defined as beginning at 7 a.m. and ending at 6 p.m. Negative emotions, and auxiliary verbs decreased, while positive emotion and non-fluencies increased. (See Figure 8.) Individuals may become more positive and less negative as the day goes on, perhaps because they have made good progress during the day, or because soon they will be done working. Individuals may become more relaxed in their language and therefore use more non-fluencies. The non-fluencies trend may also indicate an increase in cognitive fatigue over the working day.

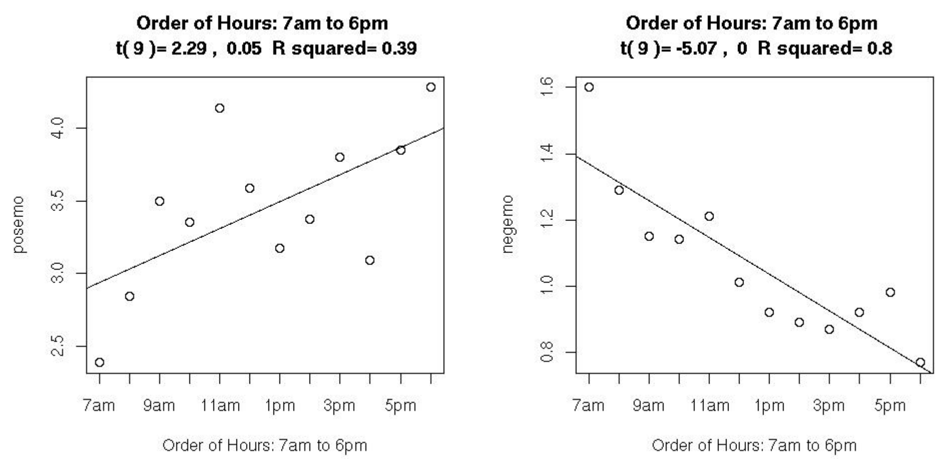

Figure 8: Trends in Positive and Negative Emotion Over Hours of the Day

5 Social Language Network Analysis Methodology

Social Language Network Analysis (SLNA) consists of three interrelated processing steps.

5.1 Preprocessing

The first step, preprocessing, involves preparing communication data for social language analysis. Since subsequent analysis steps assume a network of dyadic ties, each unit of data must be assigned as linking one or more dyadic pairs in the group. For example, for email data, the newly authored portion of each email body forms the data unit, and it is assigned to a series of dyadic links, each from the author to an individual recipient. Once all such data have been assigned to appropriate links between the participants, the preprocessing step is complete.

5.2 Processing

The second step, processing, involves converting the text associated with each link to a quantitative metric. Typically the quantitative metric is constructed according to a particular psychological, social, or emotional theory or stylized fact, such as the observation that the use of the first person plural pronoun 'we' is often used as a marker of in-group belonging, while the use of the pronoun 'they' also is used by groups as defining out-group individuals. Metrics may need to be normalized in some fashion. For example, if the theory guiding the processing step suggests that attention is conserved, the metrics may be normalized such that they sum to unity for each recipient (in-bound normalization). Conversely, if theory suggests energy is a more binding constraint, normalization is done relative to each link originator (out-bound normalization). Ratio metrics are typically computed per data unit, and then averaged as opposed to aggregating the text data first then computing a metric; metric averaging provides results which are more robust to variations in the sizes of the data sets associated with each directed link. The output of this step is a series of valued adjacency matrices,[7] one for each metric computed.

5.3 Postprocessing

The third and final step, post-processing, uses one or more of the quantitative metric matrices (see the friendship example below) in a graph processing algorithm to compute an objective of interest. For reasonably sized graphs, visualization of the results may be helpful. Applicable graph processing includes nodal ranking, total

[7] An adjacency matrix is an n by n square matrix representing the connectivity between the n entities (individuals in this case) in the network. Connectivity is read from row to column, so the value A_{ij} is the connectivity between entity i and entity j.

flow calculations, and clustering algorithms.

5.4 Application to Chat Data

This SLNA approach has been applied to the archive of work-related chat described in Section 2.2. The initial working hypothesis assumed that each person has a certain amount of daily discretionary attention that may be directed towards others. Patterns of interaction in the public forum (such as responses to queries and specific, directed verbal exchanges) can be considered to form pair-wise connections between members of the group. Using concepts of limited attention and the sustainability of strong and weak interpersonal ties, the relative strength of all possible pairings can be computed and compared. Linguistic analysis of the content of these links, moreover, can provide important insight into the richly layered and textured nature of each interpersonal working relationship.

These data were preprocessed into relational conversations based on natural time sequences in the data. Conversations were defined as consecutive messages without more than a 5-minute delay between responses. This value was chosen both by looking at the intervals in the data (note the intersection of the 5 minute line just above the second inflection point in Figure 9) and by consulting similar work in the literature (see (20)). We selected for further analysis only those conversations in which at least two individuals interacted; this was a subset of 517 conversations. Conversations are assumed to be solely between those participants synchronously participating. This is a simplification, since the chat room persisted up to the last 100 lines of chat history for absent clients, but it accurately describes the majority of conversations. The language associated with each relational link was then processed using the LIWC program, resulting in valued adjacency matrices across 80 linguistic dimensions. Post processing in SLNA is application specific, and so is discussed further in the following sections.

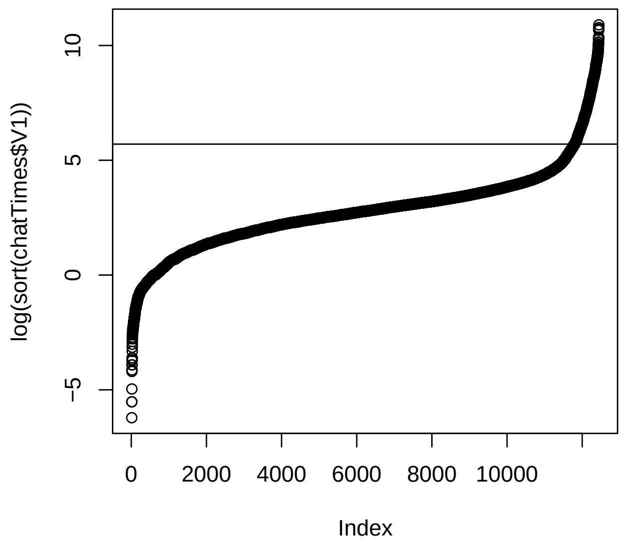

Figure 9: Log Plot of Separation Between Statements in Chat and 5 Minute Line

6 Relational Analysis Results

6.1 Language Use and Group Structure

The pattern of conversations in chat illustrate a dichotomous structure underlying the group interaction that in turn affects language use. This subsection describes the group structure and language use qualitatively first, and then using a statistical measure.

Figure 10 shows the conversation count data clustered using Johnson's hierarchical clustering with weighted average clustering. (The raw data, in the form of the conversation count adjacency matrix, are provided in Appendix A. Note that the data was symmetrized for this clustering algorithm by averaging both directional arcs between two individuals, a_{ij} and a_{ji}.) The conversational count data were assumed to be a measure of similarity. Individuals who are connected to each other near the left hand side of the dendogram shown in Figure 10 were involved in a higher number of conversations with each other. The gray rectangle superimposed on the dendogram divides the group into two subgroups of equal size. The upper subgroup represents the highly connected 'core' of the group. The lower subgroup represents the more loosely interconnected periphery. Although there is some pairwise structure in the peripheral subgroup, these ties mostly occur at weaker levels than the ties in the core subgroup.

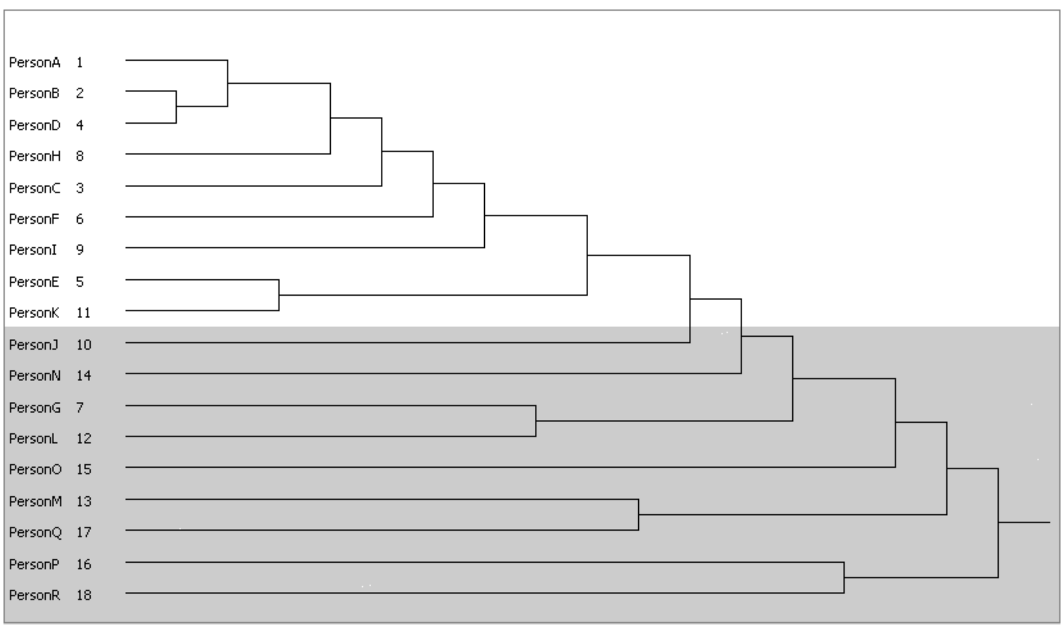

Figure 10: Johnson Hierarchical Clustering of Conversation Count

The 'We' pronoun group includes the pronouns 'we,' 'us,' and 'our' as well as various plural and possessive variants. The LIWC program computes the relative ratios of these words to all words spoken in the recorded conversations as outlined in Section 5.2. Due to variations in speech patterns by age and gender, these metrics are normalized by computing the average of non-zero 'We' pronoun usage percentages for each speaker and deducting this average from those values. Zero values for various individuals indicate no conversations occurred between the speaker and that individual, and so these zeros were left unmodified. The positive values, indicating higher than normal use of the 'We' pronouns, are shown in sorted order in Figure 11.

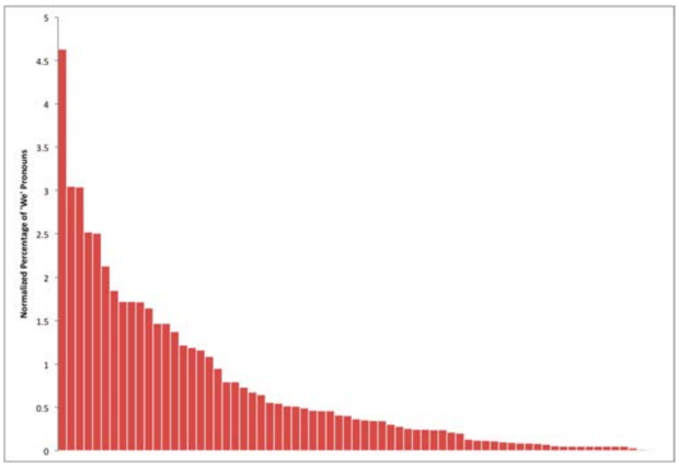

Figure 11: Positive Percentages of 'We' Pronouns in Chat Data

Figure 11 illustrates the range of values in the normalized 'We' data. There are three groups: arcs with values above 2%, arcs with values between 2% and 1%, and arcs with values below 1%. The arcs in the first group are illustrated in Figure 12. Note that all the arcs originate in the core group and link to the peripheral group.

The distribution of less than average (negative) percentages of 'We' pronouns is shown in Figure 13. Although this distribution is both much more gradual than the above average usage and lacks prominent clusters of values, a reasonable cutoff for the group of the most negative percentages is below 0.75%. Examining arcs with percentages below 0.75% in Figure 14, 17 of the 22 arcs (77%) both originate and terminate within the core group. The use of 'We' pronouns appears to be substantially less within the subgroup of individuals who comprise the core of this group.

The essence of this finding, then, is that 'We' pronoun usage is inversely related to the degree to which members belong to the group. Those individuals engaging in the most conversations within the group use pronouns from the 'We' group most infrequently when chatting with other frequent conversation partners. To test this association statistically, the Quadratic Assignment Procedure (QAP) can be applied to measure the degree of correlation between the normalized 'We' pronoun use

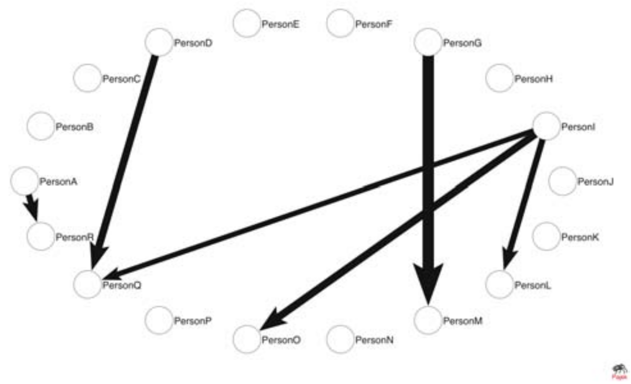

Figure 12: Conversational Pairings with Highest Percentages of ‘We’ Pronouns

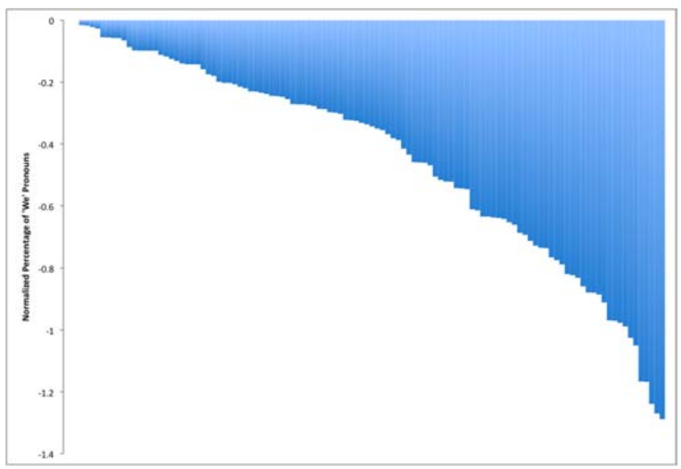

Figure 13: Negative Percentages of ‘We’ Pronouns in Chat Data

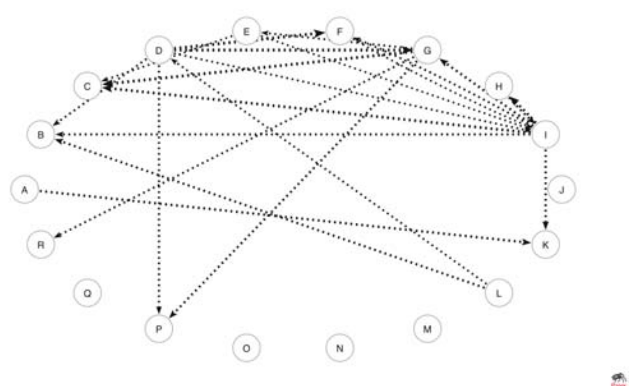

Figure 14: Conversational Pairings with Lowest Percentages of ‘We’ Pronouns

matrix and the conversation count matrix.[8] First, the corresponding cells of the two adjacency matrices are correlated using ordinary Pearson correlation. Second, a large number (50,000) of randomly re-arranged matrices are correlated to assess if the observed match is likely by pure chance. If the proportion of random trials that would generate a coefficient as small as the statistic actually observed is small enough, typically below 0.05, the hypothesis of no association is rejected. Table 4 shows that randomly permuted matrices on average have no correlation whatsoever (Pearson Correlation of 0.000), and therefore the observed inverse correlation of -0.314 is highly significant statistically.

Table 4: QAP Correlation of 'We' use and Conversation Count data

Statistic	Value
Pearson Correlation:	-0.314
Significance:	0.000
Permutation Average (50000 permutations):	0.000
Permutation Standard Deviation:	0.086
Minimum Permuted Value:	-0.291
Maximum Permuted Value:	0.306

A nearly identical result is seen with 'They' pronoun usage (Table 5), which is to be expected since 'We' and 'They' pronoun usage is highly correlated.

Table 5: QAP Correlation of 'They' use and Conversation Count data

Statistic	Value
Pearson Correlation:	-0.244
Significance:	0.000
Permutation Average (50000 permutations):	-0.001
Permutation Standard Deviation:	0.074
Minimum Permuted Value:	-0.240
Maximum Permuted Value:	0.293

To put these findings in context, Table 6 lists these two categories along with the other LIWC categories that are most strongly associated with group structure. All of these associations are negative, suggesting that the core subgroup focuses on these categories primarily in communications with the peripheral subgroup rather

[8] The conversation count data was zero-meaned before being processed, so that the Pearson's coefficient is computed for centered data.

than among themselves. The topics suggest attention to health and wellness (health, body, ingest, bio), non-work issues (family), and minimizing communication misunderstandings (smileys) in these communication channels. As discussed above, there are multiple components suggesting outreach and perhaps attempts to verbally assimilate the periphery into the core, including the 'we', 'they', and 'incl' categories. The 'discrep' group is the only category suggesting a specific work-related focus; discrepancy words are used to differentiate concepts.

Table 6: Pearson Correlation of LIWC Categories and Conversation Count data

Category	Examples	Value	Signif.	Avg.	S. D.	Min.
bio	breakfast, cafeteria, pizza	-0.330	0.000	0.000	0.082	-0.289
smileys	:-)	-0.325	0.000	0.000	0.082	-0.269
body	eye*, face, sleep*	-0.319	0.000	0.000	0.075	-0.252
we	we, us, our	-0.314	0.000	0.000	0.086	-0.291
health	ache*, exercis*, pills	-0.312	0.000	0.000	0.071	-0.249
incl	both, come, inclu*	-0.285	0.000	0.000	0.077	-0.282
ingest	ate, chew*, coffee	-0.269	0.000	-0.001	0.088	-0.305
family	family, husband, wife*	-0.264	0.000	0.000	0.061	-0.272
they	their*, them, they've	-0.244	0.000	-0.001	0.074	-0.240
discrep	besides, if, problem*	-0.237	0.000	0.000	0.077	-0.258

The above discussion illustrates the connection between a particular conversational pattern (core to periphery) and particular LIWC categories. Figure 15 represents a generalization of this mapping. Figure 15 diagrams 78 LIWC categories by clustering[9] the statistical correlations between each of the LIWC quantitative metric matrices. LIWC categories that correspond closely in usage patterns across the group, such as the 'We'-'They'-'Incl' cluster, form branches. Categories connected closer to the left-hand edge of the figure are more similar in use patterns than those connected further to the right. For example, the group patterns of use of the inclusive language of the 'We' and 'Incl' LIWC categories are more closely related to each other than to the exclusive language of the 'They' category, and so the 'We'-'They'-'Incl' branch connects 'We' to 'Incl' to the left of 'They.'

[9]The clustering method used was Johnson's hierarchical clustering with weighted average clustering.

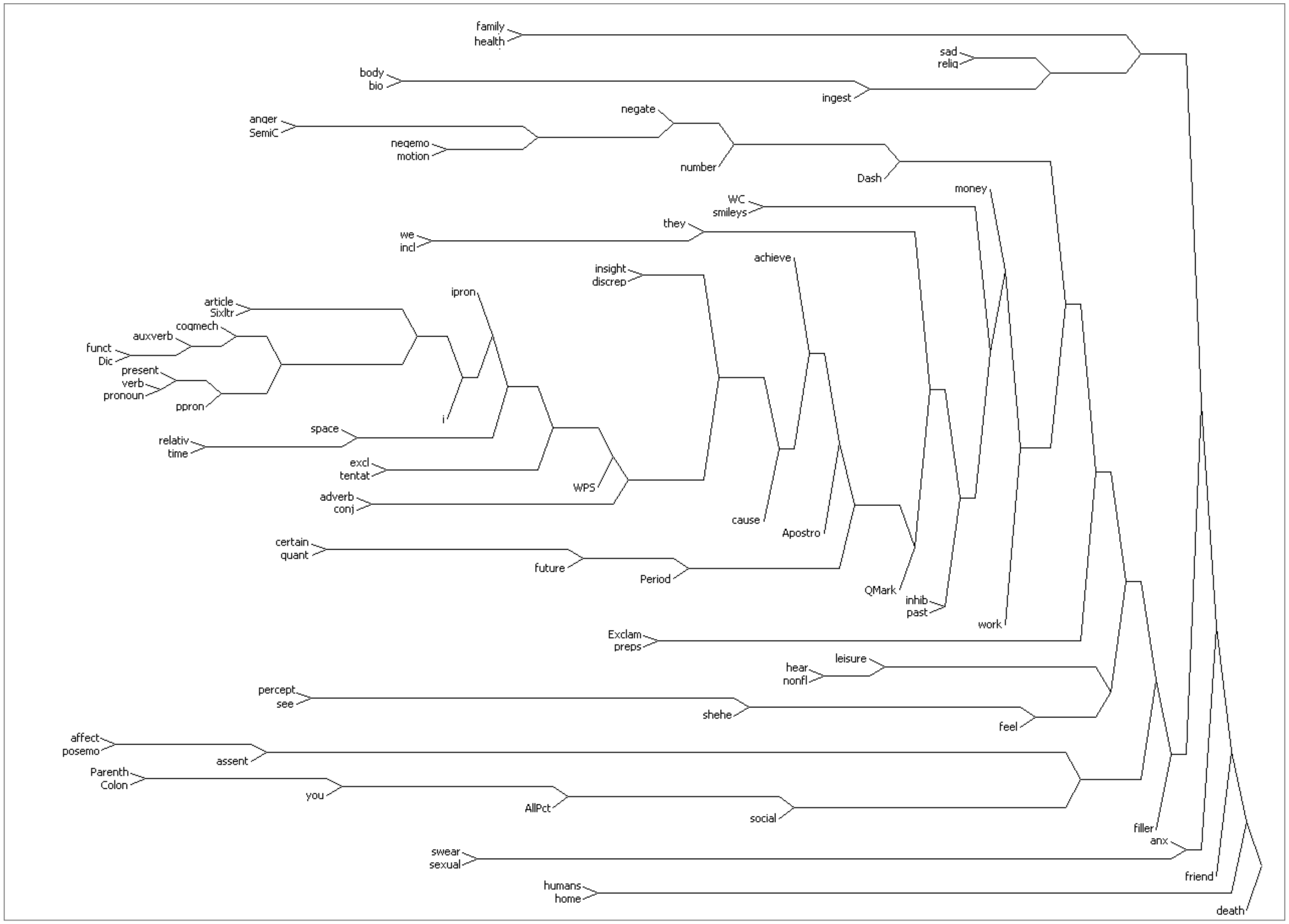

Figure 15: Clustering of LIWC Categories Based on Conversational Use in Chat

6.2 Prestige and Perception

In group work, effective task decomposition, delegation, and result integration depend on shared perceptions of expertise, competence, and engagement (3). Particularly in knowledge work where the total scope of the problem exceeds any individual's knowledge, socially constructed beliefs about relative expertise define how problems are tackled collaboratively.

To assess the group-level attitude toward the expertise of its members, we used a normalized adjacency matrices measuring first person singular pronoun (e.g. "I", "I've", "me", "mine") usage in chat conversations. Out-bound normalization converted the raw LIWC counts to the proportion of personal pronouns used with each conversant. Previous studies (23) have shown that usage of this class of pronouns (unconsciously) increases as a speaker interacts with a person of higher status. Thus the relative value on each arc between team members measures the extent to which the originator of the arc views the receiver of the message as being of higher class. We then post-processed this matrix with the Google PageRank™ algorithm, effectively using each team member's language to 'vote' for the individuals with the highest status. The results of this analysis suggested that the status hierarchy, in terms of roles, is: Group Leads, Programmers, Analysts, Manager, Students and Matrixed Staff. This hierarchy corresponds exactly to 'stylized facts' about the culture of the R&D organization, where technical skill-based roles are prized above the compliance-centric role of management, and working within one's own organizational out ranks cross-organizational work-for-hire roles. It suggests that status is at least in part a function of expertise for this work-based group.

Because the PageRank™ algorithm is a Markov-chain analysis, we can also impose a prior distribution upon it, and evaluate an individual's perception of the expertise hierarchy. This approach is more than just a direct evaluation of who in the group the individual is directly deferential to, as the opinion of the most respected individuals also factors into the final ranking. Evaluating the perspective of the group's manager against that of the entire group (see Figure 16) revealed two interesting insights. One, there is a 'retention bias' – the manager actually overvalues the team's top talent and undervalues the lesser performers, relative to the group. In other words, the manager is more concerned about losing a 'star performer' than rank-and-file members of the group. Two, there were two anomalously low rankings of members of the team (Person G and Person I in Figure 16), again relative to the group norm. Both these individuals experienced value-of-contribution recognition problems with this manager after the period of this study.

6.3 Information Flow

The most basic analysis of information flow within the group begins with a study of the use of question mark punctuation. The normalized 'QMark' network identifies the proportion of questions each individual directed to others in the group. Summing

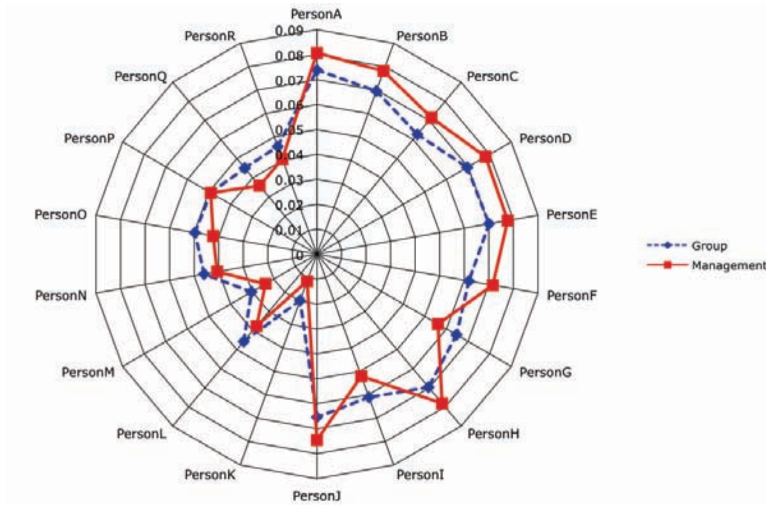

Figure 16: Group and Management Perceptions of Status.

across all in-bound links for each node reveals the individuals who are being asked the most questions, and hence can be considered crucial for information flow within the group. Figure 17 illustrates this relative information flow importance across the various roles within the group. Individuals receiving the most questions are drawn with a larger node size, a higher elevation within the layout, and a higher quantitative ranking (shown in gray next to each node).

Common topics raised in questions are listed in Table 7, which lists the LIWC categories most correlated to question marks. Topics appear to be chiefly related to work rather than social or group functions. The 'Home' category scores highly due to the use of the term 'Household' to describe a functional unit within the N-ABLE™ software.

Table 7: LIWC Categories Most Positively Correlated to Question Marks

Category	Examples	Value
home	apartment*, condos, home, house*, neighbor*	0.290
humans	adult, baby*, child, men, women	0.278
conj	also, because, if, or, though, unless	0.274
time	after, begin, day, finish*, hour*, month*	0.259
cause	affect, because, depends, result*, why	0.237
adverb	about, basically, nearly, really, when	0.210
relativ	above, before, close, near, out, up	0.192

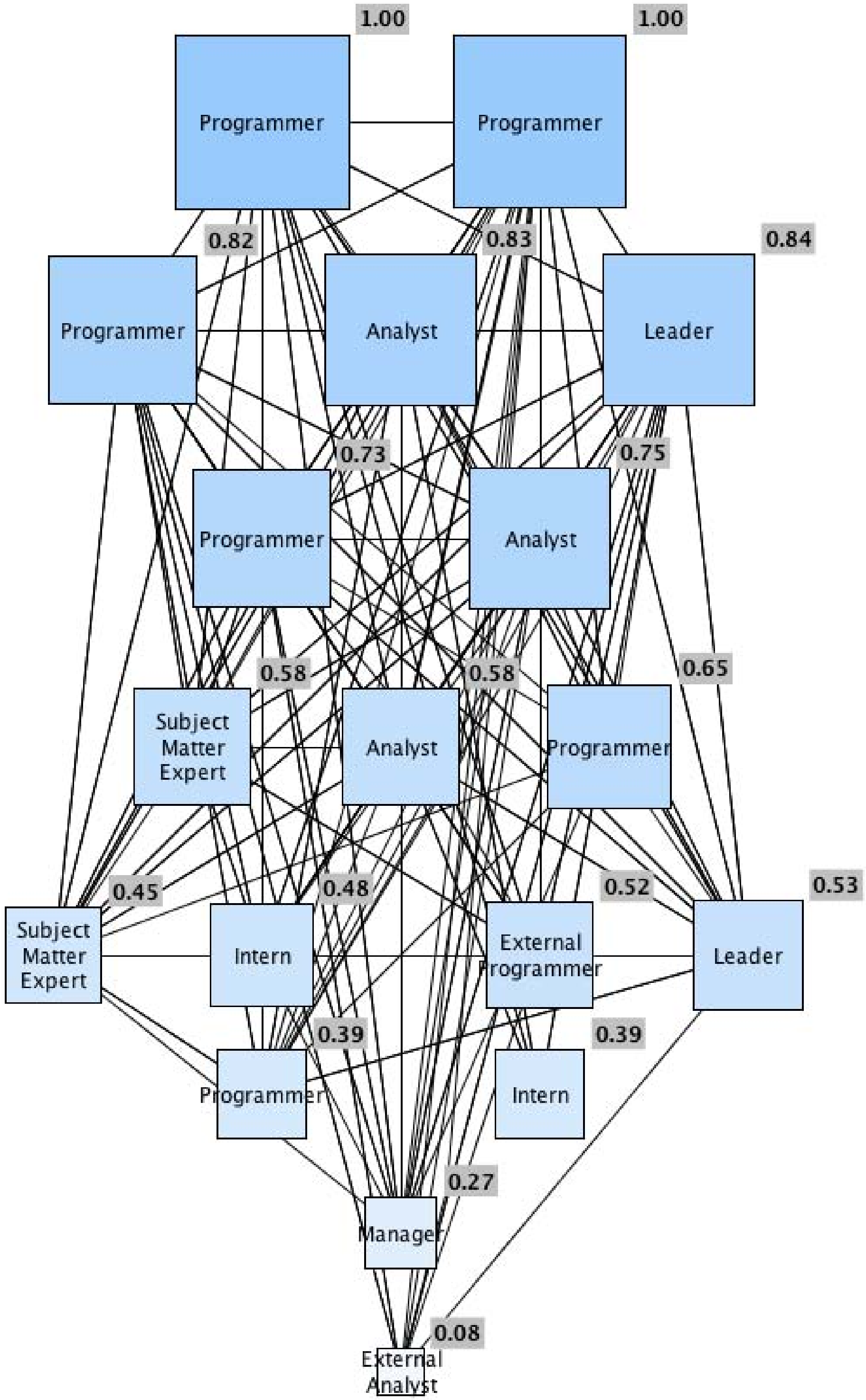

Figure 17: Group Roles Receiving the Most Questions

6.4 Group Support

Groups are known to be a source of social support to their members. We applied SLNA to identify friendship within this group, as these results could be shared with group members for evaluation without substantial risk. We first hand coded (4 coders, Cronbach's alpha for inter-coder reliability 0.821[10]) each two-person conversation in the chat data as overtly friendly or not. We then ran a logistic regression using the coded response as the binary outcome variable and selected LIWC categories as the predictors. With an alpha level for removal of 0.01, we arrive at a model for combining the values of the Number, Dash, and Apostrophe adjacency matrices:

$$A_{ij} = e^{0.358\mathrm{Number}_{ij}} * e^{0.129\mathrm{Dash}_{ij}} * e^{0.219\mathrm{Apostrophe}_{ij}} \tag{1}$$

where e^x represents the exponential function and subscripts 'i' and 'j' represent the position in the adjacency matrixes. (Note that the constant term in the logistic regression equation is discarded, since we are interested only in relative relationships between individuals.) We make the assumption that friendly statements are made more frequently between closer friends, so that higher friendly statement likelihoods correspond to increasing degrees of friendship. Under this assumption, a relative ranking of friendship strength to individuals in the network for each person can then be computed by a weighted number of independent paths algorithm (46) across the graph constructed by the above model. This approach is surprisingly good at identifying the relative strength of friend ties. In a survey-based evaluation[11] with an 82% response rate, 61% of respondents agreed that the ranking provided by this algorithm was accurate. The next closest algorithm, a ranking based solely on frequency of conversation, earned only half as many votes.

In this support model, the strongest predictors are LIWC punctuation categories rather than word categories. The fact that punctuation is a strong predictor of friendly attitudes has been discussed in the literature previously; for example, four "friendly" category codes in Carol Waseleskis work (44) describe 32% of the exclamation marks in her corpus. The computational linguistics literature similarly acknowledges that a primary use of punctuation is as an indicator of the strength of a relation (9).

[10]See Section E.3 in Appendix E for a full discussion of this calculation.
[11]Section 7.2 discusses this survey in greater detail.

7 Analysis of Survey Results

7.1 Psychological Questionnaire

Participants completed several psychological questionnaires about themselves including background information and measures of five factor personality (14), social skills, self-esteem (33), and machiavellianism (4). Participants rated their attitudes toward each of their collegues by evaluating nine different statements on a seven point Likert scale. (See Appendix F for a list of the questions). Eleven of the 18 participants whose messages were recorded in the public chat forum responded to our request to complete surveys. These participants, 4 female and 7 male, ranged in age from 22 to 64.

The attitude ratings provided by respondents to these questions were richly interrelated. As expected for essentially social traits, the degree to which an individual was rated as having higher social status, making decisions for the group, and dominating conversations were all positively correlated. The correlation coefficients are listed in Table 8. Close personal relationships followed a similarly predictable pattern. Ratings of being a close friend correlated with the rated degree of being an effective group member, being well known, and being easy to work with. The degree to which individuals were rated as both effective group members and well known also positively correlated with the degree to which they were rated as easy to work with and easy to communicate with. These correlations are shown in Table 9.

Table 8: Social Trait Correlations

	Social status	Made decisions for group
Made decisions for group	0.34	
Dominated conversation	0.39	0.20

Table 9: Close Personal Relationship Correlations

	Close friends	Effective group member	Well known
Effective group member	0.37		
Well known	0.47		
Easy to work with	0.31	0.25	0.41
Easy to communicate with		0.26	0.24

Correlations between social traits and close personal relationships were much less intuitive. Group members that were rated as being closer friends were also rated as having lower social status. Making decisions for the group was negatively related to effectiveness as a group member, ease of communication, and being difficult to work with. In other words, decision makers were perceived as being ineffective and difficult to communicate with, but their broad, overarching decisions facilitated working together. A negative correlation between being well known and dominating conversation also suggests that group leaders were not socially integrated into the group. The interesting distinction between difficulty in communication versus difficulty in working together was also directly supported by a positive correlation between easy to communicate with and difficult to work with. This would imply that work involved more than a communicative component (for example, collaboration on shared computer code), or that effective communication raised expectations for co-created work which were not met. Respondents may have also felt it was part of their job to work well with others, and may have graded others as easy to work with as an indirect positive reflection on their own interpersonal skills. These correlations are listed in Table 10.

Table 10: Social - Personal Correlations

	Social status	Made decisions for group	Dominated conversation	Difficult to work with
Close friends	-0.29			
Effective group member		-0.26		
Well known			-0.17	
Easy to communicate with		-0.44		0.28
Difficult to work with		-0.24		

The ratings each participant gave to each of the other group members can be considered that person's conceptualization of their dyadic relationship. Chats with those individuals is then an instance of the class of conversations that occurs at that degree of relation. For example, assume Person A rates Person B as a close personal friend at Likert scale degree 7, 'a great deal.' When Person A chats to Person B, then, this language can be considered to be representative of chat between very close friends. If Person E similarly rates Person K as a very close friend, chat from Person E to Person K would also be categorized as being between very close friends. Aggregating all of the language from individuals who rated their conversational partner at each Likert scale level yields a sample of language across the spectrum of sentiment for a given question. We subjected this partitioning of the chat language sample to both LIWC and Meaning Extraction Method content analysis and correlated the results to the scale level.

There was a positive correlation between use of first person plural and rated social status, as shown in Figure 18. This relationship has also been found in past research (37; 23), and confirms that use of first person plural is a good proxy for social status.

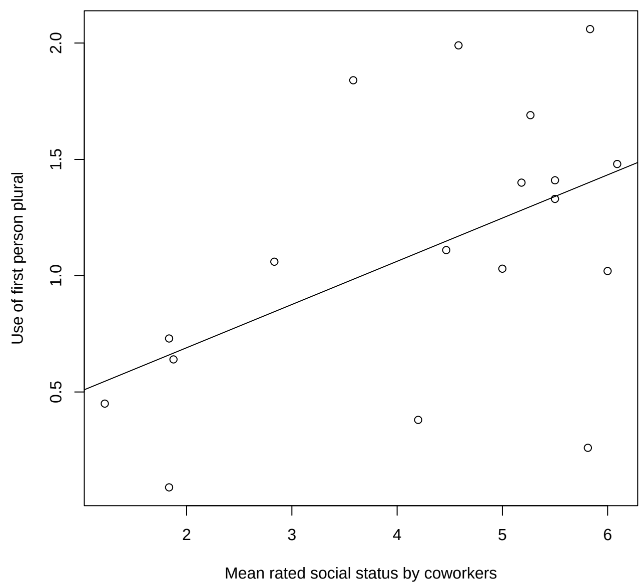

Figure 18: Positive Correlation between Rated Social Status and First Person Plural

A common observation in the organizational studies literature is that negative information is increasingly filtered as it moves up the management chain. Figure 19 provides empirical support for this assertion. The graph on the left side of Figure 19 shows a negative correlation between the LIWC category 'Negemo'[12] and higher social status. People spoke with fewer negative terms when conversing with individuals they perceived to be of higher status. The graph on the right shows a reinforcing effect, namely that people in this group tended to use more positive terms ('Posemo') the less well they knew the person. Hence, management, having both higher social status[13] and being less well known by most staff members than their peers, receive both less negative information and more positive information from staff.

How group members rated the communication skills of their peers strongly pre-

[12] An exhaustive list of words in this category and selected other categories is provided in Appendix G to aid in interpreting these results.

[13] The manager and team leads in this work group were rated as having high social status in the University of Texas questionnaire documented in Appendix F.

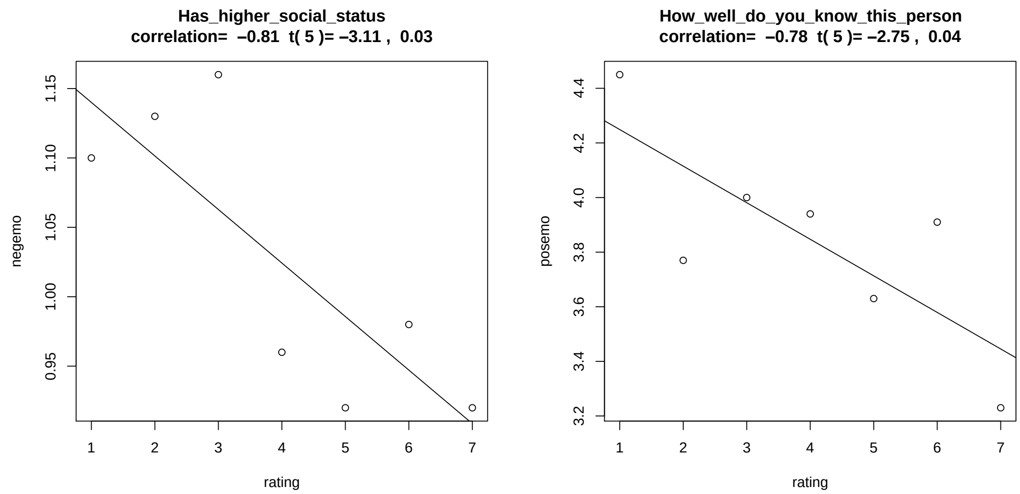

Figure 19: Emotional Filtering as a Function of Status and Familiarity

dicted the types of conversations held with those peers. The graph on the left side of Figure 20 shows a link between how easy to talk to a person is perceived to be and the extent to which conversations with that person contain tentative words. People are willing to share thoughts and interpretations they are less sure of when their conversational partner is easy to talk to. Conversely, the graph on the right side of Figure 20 shows that more difficult to approach individuals are met with more "causative" language. Conversations are initiated with these individuals predominantly when there is a reason to approach them.

7.2 Group Collaborative Work Questionnaire

A second, independently administered survey attempted to both elicit information about the social structure of the group, and provide computer-generated information from various candidate algorithms for evaluation. A representative example of this survey is presented in Appendix C. Based on the social network literature,[14] we discuss here the evaluation feedback on algorithms intended to identify distinct friendship and work-related consultation networks. Table 22 summarizes the algorithms selected by the respondents for these purposes.

Algorithm A wins the popular vote for selecting friends, while Algorithm D wins for selecting those to whom group members turn to for consultation and sensemaking. The average ranking for the adverse affects was 5.077, indicating disagreement

[14]Ibarra and Andrews have noted that "Social network theory distinguishes between the instrumental network links that arise in the course of work-role performance and expressive network relations that primarily provide friendship and social support (Tichy, Tushman, and Fomrun, 1974; Lincoln and Miller, 1979; Fombrun, 1982)." (19),p. 282.

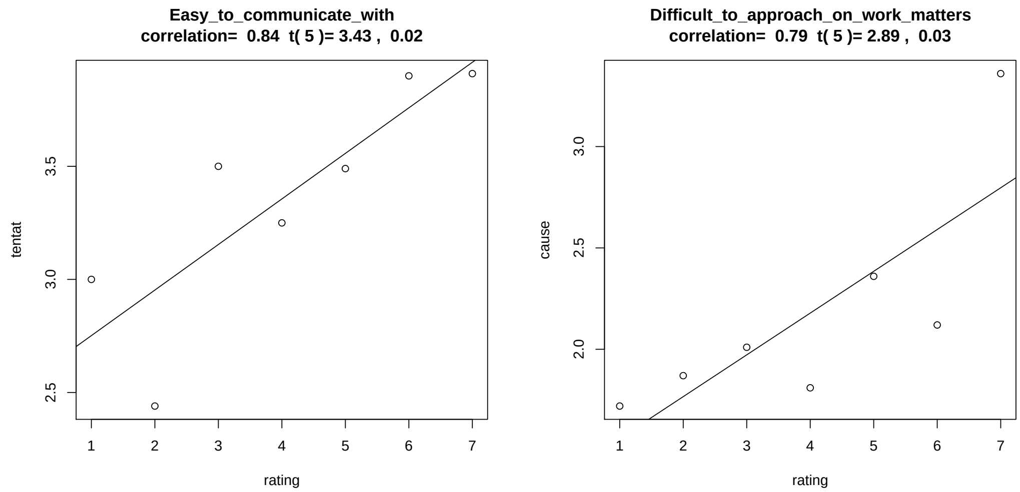

Figure 20: Type of Conversation as Function of Communication Skills

Table 11: Summary of Best Algorithm Choices

	Friendship		Consultation	
	Count	Percent	Count	Percent
Algorithm A	8	61.54%	0	0.00%
Algorithm B	1	7.69%	2	16.67%
Algorithm C	0	0.00%	1	8.33%
Algorithm D	4	30.77%	9	75.00%

with these algorithms. The details of these various candidate algorithms are discussed further below, first for the friendship category, then the consultation rankings, and finally for the adversity detection.

The friendship prediction models were as follows. As noted in Section 6.4 previously, Algorithm A is an algorithm operating over a network constructed from the small multiplicative model for combining the values of the (normed) Number, Dash, and Apostrophe adjacency matrices:

$$A_{ij} = e^{0.358\text{Number}_{ij}} * e^{0.129\text{Dash}_{ij}} * e^{0.219\text{Apostrophe}_{ij}} \tag{2}$$

where e^x represents the exponential function and subscripts 'i' and 'j' represent the position in the adjacency matrixes. A relative ranking of strength friendship to individuals in the network for each person can then be computed by a weighted number of independent paths algorithm (46) across this combined model graph. Algorithm parameters are $\lambda = 2.0$ and $N = 17$.

Algorithm B is a simple ranked ego-to-alter link algorithm operating on a network constructed from the following large multiplicative model.

$$\begin{aligned} A_{ij} \quad = \quad & e^{0.216\text{You}_{ij}} * e^{-0.164\text{Present}_{ij}} * e^{0.511\text{Number}_{ij}} * e^{0.773\text{Inhib}_{ij}} * \\ & e^{0.726\text{SemiColon}_{ij}} * e^{0.158\text{QMark}_{ij}} * e^{0.161\text{Exclam}_{ij}} * e^{0.159\text{Dash}_{ij}} * \\ & e^{0.365\text{Apostrophe}_{ij}} * e^{-0.650\text{Parenth}_{ij}} \end{aligned} \tag{3}$$

This model was arrived at by the same method as described in Section 6.4, namely back selection in a logistic regression using the coded response as the binary outcome variable, but with a less restrictive alpha level for predictor term removal of 0.05. The matrices are again normed based on a conservation of attention principle (outgoing links sum to unity).

Algorithm C is also a simple ranked ego-to-alter link algorithm, but operating on an additive model network.

$$\begin{aligned} A_{ij} \quad = \quad & \text{Affect}_{ij} + \text{We}_{ij} + \text{Friend}_{ij} + \text{Humans}_{ij} + \text{Health}_{ij} + \text{Ingest}_{ij} + \\ & \text{Leisure}_{ij} + \text{Posemo}_{ij} + \text{Assent}_{ij} + \text{Social}_{ij} \end{aligned} \tag{4}$$

The constituent normed LIWC adjacency matrices were chosen based on psychological plausibility after consultation with psychologists from the University of Texas.

Algorithm D uses distances in a graph constructed by clustering conversations. The graph, shown in Figure 21, is constructed by using Johnson's Hierarchical Clustering on the conversation count data (see Table 14 in Appendix A) with weighted averaging to form clusters. Distances are computed as the number of links away from the ego, with links further away indicating less friendliness.

The algorithms for consultation and sensemaking are all based on the 'I' pronoun matrix (Table 16 in Appendix A). Algorithm A is the simple ranked ego-to-alter link

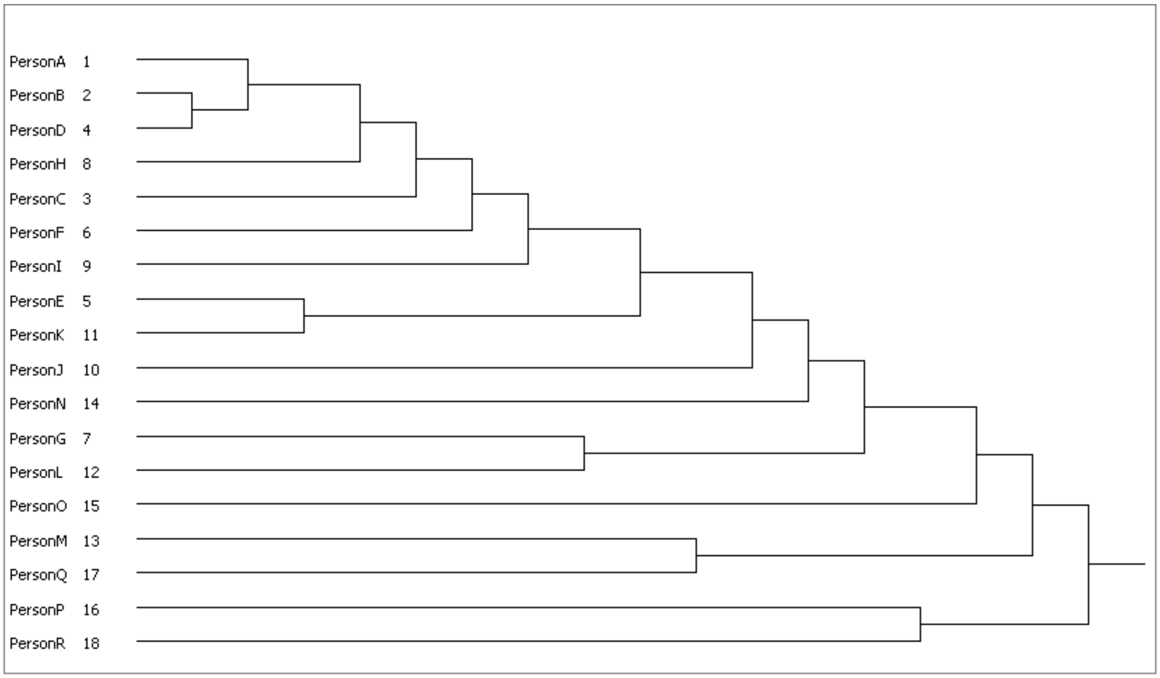

Figure 21: Johnson's Hierarchical Clustering Network on Conversation Count Data

algorithm. Algorithm B is K Step Markov model. This approach takes random walks of fixed-length K out from the root set and then computes the stationary probability of being at each individual's node. Hence, it computes the relative probability for each alter individual, given a start on the ego and an end after K steps. K was set to 3 for this survey. Algorithm C is Google PageRank™ with the prior node set to be the ego and $\beta = 0.15$. Algorithm D is the weighted number of independent paths algorithm (46) with parameters $\lambda = 2.0$ and $N = 17$.

It is worth noting that the algorithms judged most accurate in both categories (friendship and consultation/sensemaking) utilize global rather than local network information. Apparently the full graph encodes information the localized connections to an individual do not, a point we will revisit in Section 9.

To predict avoidance behaviors indicative of negative relations within the group, we examined whether the frequency of conversations between members was randomly distributed in proportion to the amount speakers participated in conversations. If conversations do not happen randomly between individuals it may suggest that individuals are discriminatory in who they collaborate with in public chat. We examined both conversations with two or more members and conversations between exactly two members, which we refer to as exclusive conversations. There were 1679 conversations with two or more participants, and there were 265 exclusive conversations. Individuals J, K, L, M, N, O, P, Q, and R were excluded from the exclusive pair analyses because they had fewer than 5 conversations (J:3, K:2, L:3, M:1, N:1, O:2, P:3, Q:0, R:1). The number of conversations that each pair had ranged from 37 between Person A and Person B to 0 between several pairs, for example between Person D and Person I, and between Person G and Person F. For the conversations among two or more participants, only Person Q, who spoke in 3 conversations, and Person R, who spoke in 4 conversations, were excluded. The number of pairs participating in these group conversations ranged from 154 (for Persons A and B) to zero for pairs who were never a part of the same conversation (for example, Person C and Person L). These observed frequencies were compared to expected frequencies assuming that individuals randomly chose who to have conversation with. The expected probability of any pair (i, j) is given by:

$$P(i, j) = P(i) * P(j|not\ i) + P(j) * P(i|not\ j) \tag{5}$$

where $P(i)$ is the fraction of all conversations i participated in, and $P(i|not\ j)$ is the ratio of the number of conversations i participated in to the sum of all conversations exclusive of those with j participating. The expected frequency is simply the probability of the pair (i, j) times the total number of conversations each pair participated in. We used the Chi Squared goodness of fit test to compare the observed and expected frequencies. For conversations with two or more speakers there was a significant deviation from the null hypothesis that individuals participated in conversations with each other at random, $\chi^2(119) = 243.56$, $p < 0.001$. For conversations limited to two speakers there was also a significant deviation from the null hypothesis, $\chi^2(35) = 54.11$, $p = 0.021$. These findings suggested that both in conversations that include at least two speakers and those that include only two speakers individuals selectively converse with some partners more than others.

Examining the exclusive pair results first, Table 12 shows that four of the top five absolute deviations from a random conversation model were positive associations between individuals with similarity relations. The pairs (Person B, Person D), (Person A, Person F), and (Person A, Person C) are all good friends and share equivalent employment roles. Person B and Person H work on similar tasking within the group.

Table 12: Top 5 Absolute Differences Exclusive Conversation Pairs

Person	Person	Observed Count	Expected Count	Difference	Absolute Difference
B	H	23	11.91	11.09	11.09
A	F	22	12.92	9.08	9.08
B	D	24	16.61	7.39	7.39
B	F	6	13.15	-7.15	7.15
A	C	28	21.40	6.60	6.60

Table 13 shows the opposite result with four of the top five absolute deviations from a random conversation model potentially attributable to avoidance.

Table 13: Top 5 Absolute Differences All Conversations

Person	Person	Observed Count	Expected Count	Difference	Absolute Difference
A	B	154	189.32	-35.32	35.32
B	H	100	77.61	22.39	22.39
B	F	38	53.25	-15.25	15.25
C	D	34	49.14	-15.14	15.14
C	B	85	96.83	-11.83	11.83

If a significant avoidance behavior was noted for any given pair, both participants were provided with this information and asked to evaluate the accuracy of this observation. The average ranking was 5.077 on a scale from 1 to 7, where 7 indicated strong disagreement with the evaluation.[15] It is possible this fairly robust rejection of this algorithm is because decisions to join group conversation depended solely on topic and similar work schedules rather than avoidance of the individuals present. Alternately, the avoidance rankings may be accurate but feelings of annoyance or exasperation in conversations occurring over nearly two years earlier may be difficult to recall. There may also be a social stigma associated with admitting the avoidance of certain individuals on these attributable (albeit coded) surveys, inhibiting a true assessment of avoidance.

[15] The actual evaluations provided are listed in Table 22 in Appendix D.

8 Discussion

8.1 Future Work

It is perhaps characteristic of all research that in addition to answering some of the initial questions conceived when approaching the subject matter, new questions and new lines of inquiry are also generated. In particular, this work as an exploratory investigation leaves many areas open for further investigation.

An important limitation of the current work is that it analyzes exchanges only within a single workgroup. As Baldwin et. al. (2) and others have noted, relationships within and between groups have significant effects on both perceptions of effectiveness and objective performance. Future work should focus on larger data sets to enable the study of inter-group effects. Additionally, while we have sought to focus on aspects of group behavior the literature suggests are generalizable, we would be remiss in not acknowledging that our sample size is one group. As Postmes et. al. point out, "Results show that norms prescribing a particular use of technology are socially constructed over time at the level of locally defined groups and also show that the influence of these norms is limited to the boundaries of the group." (32)

This work was relatively unsuccessful at eliciting information on the hindrance network, or negative relations between group members. We believe this is symptomatic at least in part of a non-confrontational culture at Sandia that discourages explicit disparagement of co-workers. Work by Sparrowe et. al. (39) suggests negative behaviors among members are perhaps more important for group performance than cooperative behaviors, and so accurately capturing this dimension is an important future goal.

Section 4.3 showed there are important temporal components this data. A new class of network models called exponential random graph models is being developed to model underlying and slowly changing structure, such as friendship or alliances, in network data considered to be stochastic or noisy. We believe these models can be usefully applied to the linguistic data we have examined here.

9 Conclusions

Social language processing is an acknowledged probabilistic approach (5), and recent work suggests that any given communication medium – email, phone, instant message, videoconferencing, face-to-face meetings – carries only a portion of the total discourse on any given topic (31). Both example SLNA applications discussed above, however, were able to reconstitute a sufficiently holistic approximation of the underlying processes to match external accuracy measures by leveraging the network. In other words, the use of the whole network reconstitutes sampling gaps at an individual level precisely because social networks are not random networks. Clustering, transitive closure, shared perceptions and views among close friends and other well-known group-based social phenomena provide redundant information that appropriate algorithms can leverage. This means that the ability of social language processing to access information only partially under the conscious control of the speaker gives insights into whole group and organizational dynamics not otherwise obtainable.

It should be noted, however, that development of these SLNA metrics is non-trivial. The size of the set of networks created by the combination of social language metrics is constrained only by the imagination and creativity of the researcher. Even when the number of candidate networks can be constrained by theoretical considerations or data fitting as described above, the plethora of network algorithms provides another source of combinatoric explosion in solution space. The discovery of new explanatory SLNA methods is an inherently explorative process.

As Weick (45) noted with the quote, "How can I know what I mean until I see what I say?," communication is central to negotiating meaning out of the events around us. Lave and Wagner (25) interpret on-going dialog across and within a spectrum of expertise as central to legitimate participation in communities of practice. Social language processing suggests, however, that this same communication is also richly layered with information about the relative social, psychological, and emotional connections that connect and situate us within a community. Social network approaches can construct higher order structures from these attributional and dyadic data. This report argues for the importance of fusing these theories into a new methodology, Social Language Network Analysis (SLNA), and demonstrates how the application of SLNA to a real world knowledge-intensive collaborative work communication corpus (22) highlights and makes explicit important components of organizational functioning, such as information exchange and evaluation (a function of perceived expertise) and social support.

Appendices

A Social Network Analysis Data

The conversation count matrix analyzed in Section 6.1 is listed in Table 14. It represents the percentage of all conversations each person had directed toward each participant. As indicated in Section 5.4, we defined a conversation to be chat synchronously occurring within 5 minutes, based on both an empirical analysis of the gaps in the chat data and for consistency with other published work in chat analysis (20). Conversations in which no other team member participated provided weight to the diagonal value for the speaker.

Table 14: Normalized Conversation Count Matrix

Person:	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
PersonA	0.206	0.201	0.112	0.119	0.086	0.064	0.027	0.089	0.035	0.017	0.004	0.012	0.008	0.007	0.004	0.005	0.004	0.001
PersonB	0.189	0.166	0.104	0.141	0.094	0.047	0.031	0.123	0.044	0.017	0.000	0.015	0.004	0.007	0.002	0.010	0.001	0.005
PersonC	0.226	0.223	0.144	0.089	0.076	0.071	0.016	0.066	0.034	0.018	0.003	0.000	0.000	0.013	0.005	0.010	0.000	0.005
PersonD	0.211	0.266	0.079	0.019	0.088	0.056	0.046	0.100	0.051	0.021	0.000	0.023	0.009	0.005	0.007	0.012	0.007	0.002
PersonE	0.189	0.220	0.083	0.109	0.114	0.034	0.029	0.103	0.043	0.014	0.011	0.029	0.000	0.006	0.006	0.006	0.003	0.003
PersonF	0.216	0.167	0.119	0.106	0.053	0.159	0.018	0.075	0.035	0.022	0.000	0.009	0.009	0.004	0.000	0.004	0.000	0.004
PersonG	0.153	0.182	0.044	0.146	0.073	0.029	0.131	0.051	0.066	0.015	0.000	0.051	0.007	0.007	0.007	0.015	0.007	0.015
PersonH	0.185	0.272	0.068	0.117	0.098	0.046	0.019	0.103	0.030	0.014	0.000	0.011	0.008	0.011	0.005	0.008	0.003	0.003
PerosnI	0.173	0.231	0.083	0.141	0.096	0.051	0.058	0.071	0.000	0.019	0.013	0.032	0.000	0.006	0.019	0.000	0.006	0.000
PersonJ	0.181	0.194	0.097	0.125	0.069	0.069	0.028	0.069	0.042	0.014	0.000	0.014	0.000	0.028	0.028	0.014	0.014	0.014
PersonK	0.250	0.000	0.083	0.000	0.333	0.000	0.000	0.000	0.167	0.000	0.167	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PersonL	0.136	0.182	0.000	0.152	0.152	0.030	0.106	0.061	0.076	0.015	0.000	0.030	0.000	0.000	0.030	0.015	0.015	0.000
PersonM	0.300	0.150	0.000	0.200	0.000	0.100	0.050	0.150	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.050	0.000
PersonN	0.161	0.194	0.161	0.065	0.065	0.032	0.032	0.129	0.032	0.065	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.032
PersonO	0.120	0.080	0.080	0.120	0.080	0.000	0.040	0.080	0.120	0.080	0.000	0.080	0.000	0.000	0.080	0.000	0.040	0.000
PersonP	0.111	0.222	0.111	0.139	0.056	0.028	0.056	0.083	0.000	0.028	0.000	0.028	0.000	0.000	0.000	0.111	0.000	0.028
PersonQ	0.200	0.067	0.000	0.200	0.067	0.000	0.067	0.067	0.067	0.067	0.000	0.067	0.067	0.000	0.067	0.000	0.000	0.000
PersonR	0.059	0.235	0.118	0.059	0.059	0.059	0.118	0.059	0.000	0.059	0.000	0.000	0.000	0.059	0.000	0.059	0.000	0.059

The following measures of centrality were calculated from the conversation count data. Betweenness values map fairly well to the core / periphery split discussed in Section 6.1, except for Person J. The previous discussion put PersonJ in the periphery subgroup, in part to enforce equal sizes across both subgroups. The Betweenness data suggest PersonJ is marginally in the core group.

Table 15: Normalized Centrality Measures of Conversation Count Matrix

Individual	**Degree**	**Closeness**	**Betweenness**	**Eigenvector**
PersonA	100.000	100.000	5.443	39.857
PersonB	94.118	94.444	2.257	39.122
PersonC	82.353	85.000	2.510	34.485
PersonD	94.118	94.444	2.257	39.122
PersonE	94.118	94.444	3.675	38.561
PersonF	82.353	85.000	1.361	35.187
PersonG	94.118	94.444	2.257	39.122
PersonH	94.118	94.444	2.257	39.122
PersonI	82.353	85.000	2.512	34.545
PersonJ	88.235	89.474	1.224	37.825
PersonK	23.529	56.667	0.000	10.790
PersonL	70.588	77.273	0.394	31.442
PersonM	41.176	62.963	0.074	19.015
PersonN	64.706	73.913	0.074	29.734
PersonO	64.706	73.913	0.149	29.382
PersonP	64.706	73.913	0.149	29.507
PersonQ	64.706	73.913	0.543	28.328
PersonR	64.706	73.913	0.074	29.391

Several results relating to prestige and deference were derived from the use of 'I' pronoun across the group. Table 16 provides the normalized data that underlies these results.

Table 16: Normalized 'I' Pronoun Usage Matrix

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
A		0.06757	0.06297	0.06339	0.05810	0.06192	0.05021	0.06548	0.05628	0.05439	0.07908	0.05439	0.06004	0.06339	0.0483	0.04854	0.05084	0.05418
B	0.07156		0.07156	0.07186	0.07229	0.06151	0.05568	0.06865	0.06501	0.05786	0	0.05553	0.07243	0.05218	0.06850	0.06049	0.05888	0.03600
C	0.08592	0.07065		0.09177	0.07390	0.08754	0.06610	0.06464	0.08706	0.06545	0	0	0	0.05782	0.07065	0.10833	0	0.070164
D	0.05523	0.05938	0.06246		0.05246	0.06954	0.04308	0.06246	0.04508	0.04385	0	0.04046	0.11369	0.09108	0.03569	0.09185	0.03108	0.10262
E	0.05657	0.05707	0.05740	0.06868		0.05574	0.07150	0.06038	0.09837	0.06652	0.11994	0.06453	0	0.08062	0.02190	0.06121	0.02770	0.03185
F	0.07922	0.07361	0.06121	0.08342	0.08522		0.06801	0.06281	0.07101	0.04881	0	0.0800	0.10002	0.04781	0	0.09102	0	0.04781
G	0.05484	0.06143	0.06510	0.05862	0.05520	0.07071		0.05557	0.05471	0.07560	0	0.04861	0.04067	0.10564	0.04604	0.05300	0.04604	0.10821
H	0.04904	0.05536	0.04761	0.06643	0.05283	0.03306	0.04128		0.04382	0.04619	0	0.05647	0.06596	0.05252	0.11041	0.08795	0.16941	0.02167
I	0.10109	0.11030	0.05948	0.06060	0.04189	0.03993	0.09662	0.10919		0.07456	0.12706	0.06562	0	0	0.05557	0	0.05808	0
J	0.07477	0.07034	0.08220	0.05890	0.06748	0.08778	0.09535	0.07691	0.07820		0	0	0	0.08163	0.08935	0.04117	0	0.09593
K	0.22942	0	0.21433	0	0.37757	0	0	0	0.17867	0		0	0	0	0	0	0	0
L	0.10745	0.08502	0	0.10274	0.08862	0.11548	0.08502	0.08640	0.08280	0.04265	0		0	0	0.04237	0.11880	0.04265	0
M	0.13556	0.13007	0	0.14266	0	0.13533	0.10900	0.16167	0	0	0	0		0	0	0	0.18571	0
N	0.06716	0.06616	0.06883	0.05525	0.06193	0.13021	0.13021	0.06427	0.15193	0.07385	0	0	0		0	0	0	0.13021
O	0.07952	0.09382	0.10277	0.08035	0.09382	0	0.10300	0.08095	0.08798	0.08095	0	0.09382	0	0		0	0.10300	0
P	0.20556	0.15649	0.10380	0.13636	0.15920	0.03709	0	0.16441	0	0.03709	0	0	0	0	0		0	0
Q	0.09937	0.10016	0	0.09937	0.10016	0	0.10016	0.10016	0.10016	0.10016	0	0.10016	0	0	0.10016	0		0
R	0.10189	0.07642	0.09729	0.08231	0.05491	0.10189	0.09729	0.10189	0	0.10189	0	0	0	0.10189	0	0.08231	0	

B Uses of Social Network Analysis in Organizations

Table 17: Common Social Network Analysis Applications.

Supporting partnerships and alliances	Executives are increasingly employing cross-organizational initiatives such as alliances or other forms of strategic partnerships to leverage their organizations' unique capabilities. Social network analysis can illuminate the effectiveness of such initiatives in terms of information flow, knowledge transfer, and decision making.
Assessing strategy execution	Core competencies or capabilities in knowledge-intensive work are usually a product of collaboration across functional or divisional boundaries. Social network analysis allows executives to determine whether the appropriate cross-functional or department collaborations are occurring to support strategic objectives.
Improving strategic decision making in top leadership networks	A core function of top executive teams is to acquire information, make sound decisions, and convey those decisions effectively to the broader organization. Social network analysis, when done with both the top leadership team and the next layer down, can provide valuable diagnostic information to leadership. Not only can it help assess connections within a top leadership team, but it can also reveal how information is entering and leaving this group.
Integrating networks across core processes	Informal networks across core processes are often fragmented by functional boundaries. Both cognitive and organizational barriers often keep groups from effectively integrating unique expertise, which can damage quality, efficiency, and innovation. As the process map did for reengineering, social network analysis provides a diagnostic assessment of information and knowledge flow both within and across functions critical to a core process.

continued on next page

Promoting innovation	Most innovation of importance is a collaborative endeavor. Whether concerned with new-product development or process improvement initiatives, social network analysis can be particularly insightful in assessing both how a team is integrating its expertise and the effectiveness with which it is drawing on the expertise of others within the organization.
Ensuring integration post-merger or large-scale change	Particularly in knowledge-intensive settings, large-scale change is fundamentally an issue of network integration. Social network analysis, done before a change initiative, can help inform the change process as well as identify central people within the network whom a sponsor might want to engage in design because of their ability to convey information to others. Social network analysis can also be done as a follow-up six to nine months after implementation. Quite often these assessments reveal significant issues that leaders need to address for the initiative to be successful.
Developing communities of practice	Communities of practice are usually not formally recognized within an organization but can be critical to an organization's ability to leverage expertise distributed by virtue of physical location or organizational design. Social network analysis can be used to uncover the key members of the community as well as assess overall health in terms of connectivity.

Source: (8), p. 8

C Example Survey

The following survey was administered to members of the Computational Economic Group who agreed to participate in this study. Sixteen of the eighteen eligible participants responded. This example has been modified to conceal the names of the participants.

The survey was designed in two parts. The first part of the survey elicited generic free-form responses, whereas a second individually customized part asked for respondents' reactions to computer-generated suggestions specific to them. A final section in the survey included 200 lines of randomly selected chat from the respondent to remind and orient them to the conversations which took place over the period of study.

Part I Elicitation

Friendship and Support

Conversations in the public chat room involved both social and work-related conversations. The first question we have addresses the social component. Please rank the persons you interacted with during the period in terms of closeness of friendship, starting with your closest friends first. Ties in ranking are permissible. Use N/A for any persons you do not recognize.

Your Ranking	Name
_______	PersonA
_______	PersonB
_______	PersonC
_______	PersonD
_______	PersonE
_______	PersonF
_______	PersonG
_______	PersonH
_______	PersonI
_______	PersonJ
_______	PersonK
_______	PersonL
_______	PersonM
_______	PersonN
_______	PersonO
_______	PersonP
_______	PersonQ
_______	PersonR

Communal Sensemaking

Conversation in chat also served to locate, filter, and make sense of information necessary to accomplish work. In the broad and multidisciplinary CEG community, different individuals provided interpretations based on their own perspectives and experience. Please rank the persons you interacted with by the degree to which you would be willing to allow their comments to shape your thinking. Ties in ranking are permissible. Use N/A for any persons you do not recognize.

Your Ranking	**Name**
_______	PersonA
_______	PersonB
_______	PersonC
_______	PersonD
_______	PersonE
_______	PersonF
_______	PersonG
_______	PersonH
_______	PersonI
_______	PersonJ
_______	PersonK
_______	PersonL
_______	PersonM
_______	PersonN
_______	PersonO
_______	PersonP
_______	PersonQ
_______	PersonR

Adverse Relations

In high stress situations, conversations can sometimes turn angry or convey frustration. Please list below any persons who you recall exhibiting such language in chat in your presence.

_______________ _______________ _______________ _______________

_______________ _______________ _______________ _______________

One common strategy for dealing with adverse relations is avoidance. Please list below any persons whom you might choose to avoid joining in conversation with.

_______________ _______________ _______________ _______________

Overall Evaluation

Please indicate the degree to which you agree with the following statements.
1. Following and contributing to communal chat required substantial attention and thought.

1	2	3	4	5	6	7
Strongly Agree			Neither Agree nor Disagree			Strongly Disagree

2. Participation in communal chat was worth the effort.

1	2	3	4	5	6	7
Strongly Agree			Neither Agree nor Disagree			Strongly Disagree

3. I participated in chat primarily for the benefit of others.

1	2	3	4	5	6	7
Strongly Agree			Neither Agree nor Disagree			Strongly Disagree

4. Group chat increased the sense of community within the group.

1	2	3	4	5	6	7
Strongly Agree			Neither Agree nor Disagree			Strongly Disagree

5. I used group chat to accomplish work.

1	2	3	4	5	6	7
Strongly Agree			Neither Agree nor Disagree			Strongly Disagree

6. Group chat provided an important awareness of events and activities in the environment that affected me and my work.

1	2	3	4	5	6	7
Strongly Agree			Neither Agree nor Disagree			Strongly Disagree

General Comments:

Part II Evaluation

Friendship and Support

The following results have been generated specifically for you by four different candidate algorithms. Please select the algorithm you feel best approximates how close, in terms of friendship and support, you were to the listed individuals during the period from September 2006 thru November 2007. If appropriate, feel free to add any comments, such as the accuracy of the selected approximation, contrasts with your ranking in Part I above, or any other relevant information.

1. The best algorithm for ranking closeness (friendship and support) is:

A B C D None of them

Algorithm A	**Algorithm B**	**Algorithm C**	**Algorithm D**	**Any Comments?**	**Name**
1	4	1	1		PersonA
N/A	N/A	N/A	N/A		PersonB
2	2	7	3		PersonC
2	5	2	2		PersonD
1	3	4	2		PersonE
5	7	8	8		PersonF
6	7	8	8		PersonG
4	7	8	6		PersonH
2	6	6	4		PersonI
8	7	8	5		PersonJ
8	7	8	7		PersonK
5	7	8	4		PersonL
6	7	8	4		PersonM
2	1	5	2		PersonN
5	7	8	6		PersonO
6	7	8	7		PersonP
3	7	3	1		PersonQ
7	7	8	6		PersonR

Communal Sensemaking

The following results have been generated specifically for you by four different candidate algorithms. Please select the algorithm you feel most closely approximates your preferences for individual consultation on work matters (indicative of the degree to which you would be willing to allow their comments to shape your thinking on the topics discussed in chat). If appropriate, feel free to add any comments, such as the accuracy of the selected approximation, contrasts with your ranking in Part I above, or any other relevant information.

1. The best algorithm for ranking my preferences in professional consultation is:

A B C D None of them

Algorithm A	**Algorithm B**	**Algorithm C**	**Algorithm D**	**Any Comments?**	**Name**
7	3	1	1		PersonA
N/A	N/A	N/A	N/A		PersonB
10	8	8	4		PersonC
4	2	2	1		PersonD
5	4	2	1		PersonE
3	7	9	5		PersonF
13	14	10	6		PersonG
11	7	5	2		PersonH
6	6	3	2		PersonI
6	12	10	6		PersonJ
1	1	11	6		PersonK
14	15	13	8		PersonL
9	9	4	2		PersonM
12	11	6	3		PersonN
5	11	10	6		PersonO
4	13	12	7		PersonP
8	10	7	4		PersonQ
2	5	9	5		PersonR

Adverse Relations

The following results have been generated specifically for you by a candidate algorithm. Please indicate if you agree with this listing. Feel free to provide any relevant comments.

For the purposes of these calculations, adverse relations have been split into three separate sub-areas: anger, negative emotions, and anxiety. For each of the three sub-areas, the algorithm estimates whom you might list as having exhibited these behaviors in chat. Note that a listing of a person's name here is not necessarily a negative connotation; good friends, for example, will often discuss external negative events as a coping mechanism. We believe, however, that dealing with these topics in conversations with others requires additional psychological resources (such as attentiveness and empathy) above work-related chat, and as such, may be considered 'adverse.'

Persons Vocalizing Anger

PersonF, score: 38.2%

1. I recall these individuals expressing this emotion in chat.

1	2	3	4	5	6	7
Strongly Agree			Neither Agree nor Disagree			Strongly Disagree

2. I believe this is the correct subset of individuals expressing this emotion in chat.

1	2	3	4	5	6	7
Strongly Agree			Neither Agree nor Disagree			Strongly Disagree

General Comments:

Persons Vocalizing Negative Emotions

No person expressed this behavior to a notable degree in front of you.

1. I recall no individuals expressing this emotion in chat.

1	2	3	4	5	6	7
Strongly Agree			Neither Agree nor Disagree			Strongly Disagree

2. I believe this is the correct subset of individuals expressing this emotion in chat.

1	2	3	4	5	6	7
Strongly Agree			Neither Agree nor Disagree			Strongly Disagree

General Comments:

Persons Vocalizing Anxiety

No person expressed this behavior to a notable degree in front of you.

1. I recall no individuals expressing this emotion in chat.

1	2	3	4	5	6	7
Strongly Agree			Neither Agree nor Disagree			Strongly Disagree

2. I believe this is the correct subset of individuals expressing this emotion in chat.

1	2	3	4	5	6	7
Strongly Agree			Neither Agree nor Disagree			Strongly Disagree

General Comments:

A common method of dealing with the additional psychological demands of adverse relations is to put off conversing with the individual, perhaps until a more convenient time. Psychologists call this avoidance behavior. Absences of communication may also simply represent different work assignments or other non-psychological explanations. A likelihood of conversation algorithm identified the following individuals as noticeably absent from conversations involving you. Please indicate if you believe this was related to a potential desire to avoid them based on their conversational content.

Computer Identified Avoidance with:	Due to chat content? (Y/N)	Comments
PersonK (100% lower)		

You are welcome to add any general comments on this entire survey here:

Thank you for taking this survey! Please return the completed survey to `ajschol@sandia.gov`.

D Survey Responses

The following tables present survey results from two independent surveys. Answers to the questions listed in Table 18 were collected anonymously as part of a survey concerning the use of color and avatar images in early 2007.

Table 18: Group-Related Anonymous Survey Responses

Statement	Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5	Average
(23) Participating in collaboration provides me with a sense of belonging and group identity within the CEG team			■	■■■■■	■	4.0
(24) I use collaboration to contact team members in preference to other methods (email, phone, face to face)			■	■■■■■■		3.857
(25) Software-mediated collaboration makes CEG a better team				■■■■■	■■	4.286

The second survey is the survey documented in Appendix C. It was conducted in 2009 with respondent attribution, although the results are presented in the following tables in coded form to preserve the privacy of the participants. Note that a dash in the following tables means no survey was returned from this individual, whereas an empty cell means the individual did not provide a value.

Table 19: Elicited Friendship and Support Rankings

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**
PersonA	N/A	2	1	1	3	1	3	3	2	3	2	2	4	2	2	2	4	3
PersonB	4	N/A	7	1	9	8	11	2	4	3		5		10	10	11		6
PersonC	1	3	N/A	3	2	1	4	3	4	3		4		1	2	2		4
PersonD	1	2	3	N/A	4	3	5	2	1	2	4	1	4	4	3	5	4	4
PersonE	3	4	4	1	N/A	3	5	4	4	4	4	4	5	2	5	5	5	4
PersonF	1	12	3	6	2	N/A	10	11	17	8	13	9	14	7	4	5	15	16
PersonG	-	-	-	-	-	-	N/A	-	-	-	-	-	-	-	-	-	-	-
PersonH	2	2	3	1	6	4	6	N/A	5	5		5	6	5	4	4	6	6
PersonI	4	3	10	1	8	11	7	6	N/A	5	4	2		9	5	12		
PersonJ	2	2	3	1	4	3	2	3	2	N/A		3		2	1	3		4
PersonK	1	7		3	6	5		7	2		N/A			4		8		
PersonL	2	3		1	8	7	10	4	1	5		N/A			6	9		
PersonM	4	4		2	4	5	6	1	4				N/A			5	3	
PersonN	1	2	1	2	1	1	4	2	3	3	3	4		N/A	3	4		3
PersonO	3	7	2	6	9	4	1	12	10	5		11		8	N/A			13
PersonP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N/A	-	-
PersonQ	4			2	6	7	5		3				1			8	N/A	
PersonR	2	2	3	3	3	4	4	2		4	1			3		4		N/A

Table 20: Elicited Sensemaking Rankings

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**
PersonA	N/A	2	1	1	1	1	6	3	4	5	6	5	4	2	4	7	4	2
PersonB	2	N/A	2	3	5	4	6	2	6	4		6		5	3	6		3
PersonC	1	3	N/A	2	1	1	4	3	3	2		3		1	2	3		1
PersonD	1	2	2	N/A	3	3	5	1	4	1	4	4	4	3	2	5	4	4
PersonE	1	5	2	4	N/A	1	6	5	5	2	6	6	7	3	3	6	7	6
PersonF	1	9	3	5	2	N/A	8	10		7				4	6	11		
PersonG	-	-	-	-	-	-	N/A	-	-	-	-	-	-	-	-	-	-	-
PersonH	2	1	2	1	4	4	7	N/A	5	3		5	6	4	3	4	6	1
PersonI	2	4	5	1	9	6	8	6	N/A	3	10	5		9	3	7		
PersonJ	1	1	1	1	1	1	2	3	3	N/A		3		3	1	2		3
PersonK	1	7		3	4	5		7	2		N/A			6		5		
PersonL	1	6		2	8	4	9	5	3	3		N/A			4	7		
PersonM	2	3		2	3	3	5	1	4				N/A			4	2	
PersonN	1	2	1	2	1	1	3	2	3	2	3	3		N/A	2	4		3
PersonO	2	6	3	7	4	1	10	9	12	8		13		5	N/A			11
PersonP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N/A	-	-
PersonQ				1	4	3	4		6				2			5	N/A	
PersonR	1	2	1	4	2	5	2	1		3	6			2		6		N/A

Table 21: Overall Evaluation Responses

Statement	Strongly Agree 1	2	3	Neither Agree nor Disagree 4	5	6	Strongly Disagree 7	Average
Following and contributing to communal chat required substantial attention and thought	I, O	D, E, Q, R	C, J, L	B, K, N	M	H	F	3.267
Participation in communal chat was worth the effort	C, F, I, K	B, D, H, L, N	E, O	M, Q	J	R		2.6
I participated in chat primarily for the benefit of others		B, O		H, I	C, E, M, N, Q, R	D, J, K, L	F	4.867
Group chat increased the sense of community within the group	B, F, I, K	D, E, H, L, N, O	J, M, R	C	Q			2.267
I used group chat to accomplish work	B, F, N	D, H, I, K, L, Q	C, M, R	J	E	O		2.6
Group chat provided an important awareness of events and activities in the environment that affected me and my work	F	B, C, E, H, I, K, L	D, M, N, O, Q	R	J			2.6

Table 22: Best Algorithm Choices

Name	Best Friend	Best Consult	Rate Adverse
PersonB	D	D	2
PersonC	D	D	5
PersonD	A	D	6
PersonE	B	D	7
PersonH	D	D	5
PersonI	A	D	6
PersonJ	A	D	7
PersonK	A		6
PersonL	A	B	7
PersonM		B	4
PersonN	A	D	2
PersonO	A		
PersonQ	D	D	7
PersonR	A	C	2

E Manual Friendliness Coding

As discussed in Section 6.4, conversations between pairs of individuals were manually coded for degree of friendliness. This appendix documents the coding instructions and examples of applying the guidelines given to the coders.

E.1 Instructions

These conversations took place over the course of a year between a group of engineers, economists, and computer programmers working together on building and analyzing simulations of economic crises (like the aftermath of Hurricane Katrina). The conversations were taken from a public chat room where anyone who was working on the project could send a message, and anyone, even if they were not sending messages, could read other sent messages. Over the course of the year IMs were sent at all times of day and on all days of the week. From this collection of messages we separated out conversations between two people, by finding times when only two people sent messages to each other and they IMed without taking more than a 5 minute break. There are several reasons that each conversation may be hard to follow or confusing. Because of the way we defined a conversation, sometimes a conversation will be a continuation of a previous conversation and sometimes a conversation will end abruptly; this can be confusing. Also, often speakers refer to images that they are both able to see while they were chatting on IM, and use very technical words to describe their work. Don't worry about understanding every word that is said. Instead, focus on the relationship between the two speakers, how they are talking to each other, and your general perceptions of the conversation. Some conversations will be as short as only two lines, others will be long. Try not to be influenced by the length of the conversation; instead, concentrate on the content.

We are interested in your perceptions of how the two individuals are relating to each other in the conversation. Are they friends? How helpful are they to the other? Do they express negative feelings toward each other? Is one more of an expert than the other? Please read each of the conversations carefully. For each conversation, answer the following questions as accurately as you can based on your perceptions from that conversation alone. You can answer yes to multiple questions for the same conversation; you can answer no to all the questions for the same conversation.

1. Do the speakers joke, use nicknames, or in some way demonstrate that they are friendly beyond the politeness you would express to a coworker?
Yes No/Can't Tell

2. Are the speakers actively solving work problems?
Yes No

3. Is one of the speakers more knowledgeable than he or she was before as a result of the conversation?

Yes No
If yes, which speaker(s) gained knowledge?

4. If the conversation is about work, who would you say has more expertise on the topic at hand? Write down the code of the person or write can't tell

5. Based on this conversation, do either of the speakers seem at all frustrated with the other?
Yes No

E.2 Examples

The following first five conversations were analyzed and coded by Yla Tausczik and provided to the coders as examples of how to answer the coding questions.

id	speaker	message
2	PersonB	whatcha think?
2	PersonF	about what?
2	PersonB	the screenboards
2	PersonF	let me get a screenboard up.
2	PersonF	very nice! Good you kept the NViz name too; alludes ot NISAC and YOU.
2	PersonB	you like my super-slinky-sexy Bezier curves? ;-)

Questions
1. Do the speakers joke, use nicknames, or in some way demonstrate that they are friendly beyond the politeness you would express to a coworker? **Yes**
2. Are the speakers actively solving work problems? **Yes**
3. Is one of the speakers more knowledgeable than he or she was before as a result of the conversation? Yes No If yes, which speaker(s) gained knowledge? **Person B**
4. If the conversation is about work, who would you say has more expertise on the topic at hand? Write down the code of the person or write can't tell **Can't tell**
5. Based on this conversation, do either of the speakers seem at all frustrated with the other? **No**

id	speaker	message
8	PersonB	I think it is astounding that the same guy who wrote Alice in Wonderland and Through the Looking Glass also wrote An elementary treatise on determinants : with their application to simultaneous linear equations and algebraical geometry
8	PersonA	that's probably why he needed to write Alice in Wonderland and Through the Looking Glass
8	PersonB	hehe. He wrote Alice in Wonderland first

Questions

1. Do the speakers joke, use nicknames, or in some way demonstrate that they are friendly beyond the politeness you would express to a coworker? **No/Can't Tell**

2. Are the speakers actively solving work problems? **No**

3. Is one of the speakers more knowledgeable than he or she was before as a result of the conversation? **No**

4. If the conversation is about work, who would you say has more expertise on the topic at hand? **Can't tell**

5. Based on this conversation, do either of the speakers seem at all frustrated with the other? **No**

id	speaker	message
9	PersonH	Question for Everyone: I need some feedback about the WhiteBoard does anyone use it?
9	PersonB	I use it when in N2... but obviously not in NJC.... The only time I use it is when I want to illustrate something but have no image to load....
9	PersonB	PersonD PersonJ and I have had several conversations on it
9	PersonH	So adding back into NJC/N2 would be of value.
9	PersonB	Dunno what others think... You should ask someone who is working with CEG right now ;-P
9	PersonB	seems useful to me though. Or maybe lift the restriction on the screenboard about drawing inside a image
9	PersonH	Yeah that's a bit more of a technical challenge right now.
9	PersonB	understandable. You should ask the others for sure... I'd say add it but I may be one of the only people who uses it.
9	PersonH	Wouldn't be that hard. You'd only have one fixed WB not multiples like SB.
9	PersonB	that seems reasonable
9	PersonB	who in our group has a non-mac that they can try to watch a quicktime I just made?
9	PersonH	PersonJ? PersonA?
9	PersonB	ok both away ;-)
9	PersonH	If you mail it to me I can try and dig up a WinBox...
9	PersonB	lol - not a big deal. I am just curious if this codec is Windows friendly

Questions

1. Do the speakers joke, use nicknames, or in some way demonstrate that they are friendly beyond the politeness you would express to a coworker? **No/Can't Tell**

2. Are the speakers actively solving work problems? **Yes**

3. Is one of the speakers more knowledgeable than he or she was before as a result of the conversation? **Yes Person H**

4. If the conversation is about work, who would you say has more expertise on the topic at hand? Write down the code of the person or write can't tell **Person H**

5. Based on this conversation, do either of the speakers seem at all frustrated with the other? **No**

id	speaker	message
22	PersonB	dude those bongos have got to go.... When I load this thing up it sounds like a bull in a drum shop.
22	PersonB	or perhaps a 4 year old with a drum set...
22	PersonA	PersonF did volunteer to use Garage Tunes to trim them back a bit. The sounds that were affordable were all long. Sorry!
22	PersonA	We do have a volume knob RFE in the pipeline
22	PersonB	yeah I saw that someone beat me to the request ;-)
22	PersonA	PersonX had actually done some field observations that helped set that one in motion - thanks PersonX!

Questions

1. Do the speakers joke, use nicknames, or in some way demonstrate that they are friendly beyond the politeness you would express to a coworker? **No/Can't Tell**

2. Are the speakers actively solving work problems? **Yes**

3. Is one of the speakers more knowledgeable than he or she was before as a result of the conversation? **No**

4. If the conversation is about work, who would you say has more expertise on the topic at hand? **Person A**

5. Based on this conversation, do either of the speakers seem at all frustrated with the other? **No**

id	speaker	message
25	PersonF	hey all - for today's collab session could you please try the ShapeFileViewer on nisac-g5...its my first application of njc to something other than N2 directly. Thanks!
25	PersonC	PersonF–file received!

Questions

1. Do the speakers joke, use nicknames, or in some way demonstrate that they are friendly beyond the politeness you would express to a coworker? **No/Can't Tell**

2. Are the speakers actively solving work problems? **Yes**

3. Is one of the speakers more knowledgeable than he or she was before as a result of the conversation? **No**

4. If the conversation is about work, who would you say has more expertise on the topic at hand? **Can't tell**

5. Based on this conversation, do either of the speakers seem at all frustrated with the other? **No**

E.3 Analysis of Coding Results

The reliability of the manual coding process was assessed using Cronbach's alpha (7). Cronbach's alpha is an index of reliability, given by the equation

$$\alpha = (\frac{k}{k-1})(\frac{s_y^2 - \Sigma s_i^2}{s_y^2}) \tag{6}$$

where k is the number of coders, s_y^2 is the variance of the combined scores for each conversation, and s_i^2 is the variance of each coder individually, summed across all coders. This alpha coefficient ranges in value from 0 to 1, and can be interpreted as the expected correlation between scores assigned by coders on conversations.

Cronbach's alpha for each of the five questions is shown in Table 23. Cronbach's alpha for Questions 3 and 4 were only computed when all coders indicated that the conversation was about work, a subset of 102 conversations. Questions 3 and 4 were also evaluated only on whether or not the coders indicated a speaker had gained knowledge or was an expert, rather than by matching the identified person across all coders. Question 5 did not receive enough yes responses to calculate a reliability statistic, as shown in Table 24, which lists the ratio of yes responses to each of the questions for each coder. Question 5 was therefore eliminated from further analysis. Questions 1 and 2 are at a reasonable level of reliability, but Questions 3 and 4 are below the normal range of acceptable reliabilities.

Table 23: Inter-coder Reliability

Question	α
Question 1 (Friendly?)	0.821
Question 2 (Work Related?)	0.865
Question 3 (Knowledge Transferred?)	0.592
Question 4 (Identify Expert)	0.391
Question 5 (Frustration?)	N/A

Table 25 shows the counts for those conversations in which there was complete agreement between all four coders. These unanimously coded conversations were

Table 24: Frequency of ‘Yes’ Responses by Coder

Question	Coder1	Coder2	Coder3	Coder4
Question 1 (Friendly?)	0.26	0.19	0.24	0.41
Question 2 (Work Related?)	0.54	0.74	0.50	0.46
Question 3 (Knowledge Transferred?)	0.75	0.58	0.75	0.73
Question 4 (Identify Expert)	0.72	0.40	0.57	0.83
Question 5 (Frustration?)	0.01	0.01	0.01	0.00

Table 25: Count of Unanimously Coded Conversations

Question	Coded Yes	Coded No	Total
Question 1 (Friendly?)	29	147	176
Question 2 (Work Related?)	24	38	62
Question 3 (Knowledge Transferred?)	29	6	35
Question 4 (Identify Expert)	15	4	19

then used to investigate predictive models for each question based on LIWC categories.

Based on the variation in the total number of conversations available for predictive modeling and the low Cronbach's alpha values for Questions 3 and 4, two different approaches were taken. Question 1 was addressed using a logistic regression with the coded response as the binary outcome variable and backwards selected LIWC categories as the predictors, as discussed in Section 6.4. Depending on the selectiveness of the criteria for predictor removal (α), two different logistic regression models were calculated. The coefficients of these models are given in Table 26. Note that the constant (intercept) terms of these regression models are not shown, since for relative rankings between individuals the constant terms are irrelevant.

Table 26: Friendliness Logistic Regression Coefficients

	LIWC Category									
α	Apostro	Dash	Exclam	Inhib	Number	Parenth	Present	QMark	SemiC	You
5%	0.365	0.159	0.161	0.773	0.511	-0.650	-0.164	0.158	0.726	0.216
1%	0.219	0.129			0.358					

A more approximate approach was taken to isolate LIWC predictors for the other three questions. A t-test was used to identify the LIWC categories with statistically significant mean differences. Although this approach will not obtain the best regression model, it will isolate the predictors that show the most dramatic differences between the types of conversations. These predictors may be the easiest to work with in future models. A backward selection logistic regression is then run to reduce the number of categories down to a manageable set. Note that due to the number of predictors being considered there may be very inflated Type 1 error rate. For this reason only the most significant predictors were selected where possible, namely the Wald p-value removal was set at 0.10. Table 27 lists these predictors and the sign of their influence. These predictors are potential categories of interest for future modeling efforts.

Table 27: Potential LIWC Predictors

Question	Categories
Question 2 (Work Related?)	Anger (-), Exclam (-), Filler (-), I (+), Past (-), Posemo (-), Social (-)
Question 3 (Knowledge Transferred?)	I (-), Quant (-)
Question 4 (Identify Expert)	Dash (-)

F University of Texas Psychological Questionnaire

Each survey respondent was asked to rate every other group member on the degree to which the following nine statements were true on a Likert scale from 1 to 7, where 1 means 'not at all,' and 7 means 'a great deal.'

Easy to work with
Made decisions for the group
Effective group member
Dominated conversations
Easy to communicate with
Has higher social status
How well do you know this person
Difficult to approach on work matters
Close personal friend

We present here the aggregated ratings for the group leadership (Line Management and Team Leads: Persons C, F, and P) on the question 'Has higher social status' to support the argument on information filtering in Section 7.1. The weighted average of those who provided a numeric ranking (that is, excluding the 'N/A' responses) is 5.947, indicating that group leadership was definitely perceived as having higher social status.

Table 28: Higher Social Status Responses for Leadership

Ranking	Total Number of Responses
N/A	8
1 - Not at all	
2	
3	1
4 - Neutral	6
5	3
6	12
7 - A great deal	16

G Selected LIWC Category Content Words

To aid in the interpretation of some of the results presented in this report, we reproduce here the word stems in several LIWC categories.

G.1 Cause

activat* affect affected affecting affects aggravat* allow* attribut* based bases basis because boss* caus* change changed changes changing compel* compliance complie* comply* conclud* consequen* control* cos coz create* creati* cuz deduc* depend depended depending depends effect* elicit* experiment force* foundation* founded founder* generate* generating generator* hence how hows how's ignit* implica* implie* imply* inact* independ* induc* infer inferr* infers influenc* intend* intent* justif* launch* lead* led made make maker* makes making manipul* misle* motiv* obedien* obey* origin originat* origins outcome* permit* pick produc* provoc* provok* purpose* rational* react* reason* response result* root* since solution* solve solved solves solving source* stimul* therefor* thus trigger* use used uses using why

G.2 I

i Id I'd I'll Im I'm ive I've me mine my myself

G.3 Negemo

abandon* abuse* abusi* ache* aching advers* afraid aggravat* aggress* agitat* agoniz* agony alarm* alone anger* angr* anguish* annoy* antagoni* anxi* apath* appall* apprehens* argh* argu* arrogan* asham* assault* asshole* attack* aversi* avoid* awful awkward* bad bashful* bastard* battl* beaten bitch* bitter* blam* bore* boring bother* broke brutal* burden* careless* cheat* complain* confront* confus* contempt* contradic* crap crappy craz* cried cries critical critici* crude* cruel* crushed cry crying cunt* cut cynic* damag* damn* danger* daze* decay* defeat* defect* defenc* defens* degrad* depress* depriv* despair* desperat* despis* destroy* destruct* devastat* devil* difficult* disadvantage* disagree* disappoint* disaster* discomfort* discourag* disgust* dishearten* disillusion* dislike disliked dislikes disliking dismay* dissatisf* distract* distraught distress* distrust* disturb* domina* doom* dork* doubt* dread* dull* dumb* dump* dwell* egotis* embarrass* emotional empt* enemie* enemy* enrag* envie* envious envy* evil* excruciat* exhaust* fail* fake fatal* fatigu* fault* fear feared fearful* fearing fears feroc* feud* fiery fight* fired flunk* foe* fool* forbid* fought frantic* freak* fright* frustrat* fuck fucked* fucker* fuckin* fucks fume* fuming furious* fury geek* gloom* goddam* gossip* grave* greed* grief griev* grim* gross* grouch* grr* guilt* harass* harm harmed harmful* harming harms hate hated hateful* hater* hates hating hatred heartbreak* heartbroke* heartless* hell hellish helpless* hesita* homesick* hopeless* horr* hostil* humiliat* hurt* idiot ignor* immoral* impatien* impersonal impolite* inadequa* indecis* ineffect* inferior* inhib* insecur* insincer* insult* interrup* intimidat* irrational* irrita* isolat* jaded jealous* jerk jerked jerks kill* lame* lazie* lazy liabilit* liar* lied lies lone* longing* lose loser* loses losing loss* lost lous* low* luckless* ludicrous* lying mad maddening madder maddest maniac* masochis* melanchol* mess messy miser* miss missed misses missing mistak* mock mocked mocker* mocking mocks molest* mooch* moodi* moody moron* mourn* murder* nag* nast* needy neglect* nerd* nervous* neurotic* numb* obnoxious* obsess* offence* offend* offens* outrag* overwhelm* pain pained painf* paining pains panic* paranoi* pathetic* peculiar* perver* pessimis* petrif* pettie* petty* phobi* piss* piti* pity* poison* prejudic* pressur* prick* problem* protest protested protesting puk* punish* rage* raging rancid* rape* raping rapist* rebel* reek* regret* reject* reluctan* remorse* repress* resent* resign* restless* revenge* ridicul* rigid* risk* rotten rude* ruin* sad sadde* sadly sadness sarcas* savage* scare* scaring scary sceptic* scream* screw* selfish* serious seriously seriousness severe* shake* shaki* shaky shame* shit* shock* shook shy* sicken* sin sinister sins skeptic* slut* smother* smug* snob* sob sobbed sobbing sobs solemn* sorrow* sorry spite* stammer* stank startl* steal* stench* stink* strain* strange stress* struggl* stubborn* stunk stunned stuns stupid* stutter* submissive* suck sucked sucker* sucks sucky suffer suffered sufferer* suffering suffers suspicio* tantrum* tears teas* temper tempers tense* tensing tension* terribl* terrified terrifies terrify terrifying terror* thief thieve* threat* ticked timid* tortur* tough* traged* tragic* trauma* trembl* trick* trite trivi* troubl* turmoil ugh ugl* unattractive uncertain* uncomfortabl* uncontrol* uneas* unfortunate* unfriendly ungrateful* unhapp* unimportant unimpress* unkind unlov* unpleasant unprotected unsavo* unsuccessful* unsure* unwelcom* upset* uptight* useless* vain vanity vicious* victim* vile villain* violat* violent* vulnerab* vulture* war warfare* warred warring wars weak* weapon* weep* weird* wept whine* whining whore* wicked* wimp* witch woe* worr* worse* worst worthless* wrong* yearn*

G.4 Posemo

accept accepta* accepted accepting accepts active* admir* ador* advantag* adventur* affection* agree agreeab* agreed agreeing agreement* agrees alright* amaz* amor* amus* aok appreciat* assur* attachment* attract* award* awesome beaut* beloved benefic* benefit benefits benefitt* benevolen* benign* best better bless* bold* bonus* brave* bright* brillian* calm* care cared carefree careful* cares caring casual casually certain* challeng* champ* charit* charm* cheer* cherish* chuckl* clever* comed* comfort* commitment* compassion* compliment* confidence confident confidently considerate contented* contentment convinc* cool courag* create* creati* credit* cute* cutie* daring darlin* dear* definite definitely delectabl* delicate* delicious* deligh* determina* determined devot* digni* divin* dynam* eager* ease* easie* easily easiness easing easy* ecsta* efficien* elegan* encourag* energ* engag* enjoy* entertain* enthus* excel* excit* fab fabulous* faith* fantastic* favor* favour* fearless* festiv* fiesta* fine flatter* flawless* flexib* flirt* fond fondly fondness forgave forgiv* free free* freeb* freed* freeing freely freeness freer frees* friend* fun funn* genero* gentle gentler gentlest gently giggl* giver* giving glad gladly glamor* glamour* glori* glory good goodness gorgeous* grace graced graceful* graces graci* grand grande* gratef* grati* great grin grinn* grins ha haha* handsom* happi* happy harmless* harmon* heartfelt heartwarm* heaven* heh* helper* helpful* helping helps hero* hilarious hoho* honest* honor* honour* hope hoped hopeful hopefully hopefulness hopes hoping hug hugg* hugs humor* humour* hurra* ideal* importan* impress* improve* improving incentive* innocen* inspir* intell* interest* invigor* joke* joking joll* joy* keen* kidding kind kindly kindn* kiss* laidback laugh* libert* like likeab* liked likes liking livel* LMAO LOL love loved lovely lover* loves loving* loyal* luck lucked lucki* lucks lucky madly magnific* merit* merr* neat* nice* nurtur* ok okay okays oks openminded* openness opportun* optimal* optimi* original outgoing painl* palatabl* paradise partie* party* passion* peace* perfect* play played playful* playing plays pleasant* please* pleasing pleasur* popular* positiv* prais* precious* prettie* pretty pride privileg* prize* profit* promis* proud* radian* readiness ready reassur* relax* relief reliev* resolv* respect revigor* reward* rich* ROFL romanc* romantic* safe* satisf* save scrumptious* secur* sentimental* share shared shares sharing silli* silly sincer* smart* smil* sociab* soulmate* special splend* strength* strong* succeed* success* sunnier sunniest sunny sunshin* super superior* support supported supporter* supporting supportive* supports suprem* sure* surpris* sweet sweetheart* sweetie* sweetly sweetness* sweets talent* tehe tender* terrific* thank thanked thankf* thanks thoughtful* thrill* toleran* tranquil* treasur* treat triumph* true trueness truer truest truly trust* truth* useful* valuabl* value valued values valuing vigor* vigour* virtue* virtuo* vital* warm* wealth* welcom* well* win winn* wins wisdom wise* won wonderf* worship* worthwhile wow* yay yays

G.5 Tentat

allot almost alot ambigu* any anybod* anyhow anyone* anything anytime anywhere apparently appear appeared appearing appears approximat* arbitrar* assum* barely bet bets betting blur* borderline* chance confus* contingen* depend depended depending depends disorient* doubt* dubious* dunno fairly fuzz* generally guess guessed guesses guessing halfass* hardly hazie* hazy hesita* hope hoped hopeful hopefully hopefulness hopes hoping hypothes* hypothetic* if incomplet* indecis* indefinit* indetermin* indirect* kind (of) kinda kindof likel* lot lotof lots lotsa lotta luck lucked lucki* luckless* lucks lucky mainly marginal* may maybe might mightve might've most mostly myster* nearly obscur* occasional* often opinion option or overall partly perhaps possib* practically pretty probable probablistic* probably puzzl* question* quite random* seem seemed seeming* seems shaki* shaky some somebod* somehow someone* something* sometime sometimes somewhat sort sorta sortof sorts sortsa spose suppose supposed supposes supposing supposition* tempora* tentativ* theor* typically uncertain* unclear* undecided* undetermin* unknow* unlikel* unluck* unresolv* unsettl* unsure* usually vague* variab* varies vary wonder wondered wondering wonders

G.6 We

lets let's our ours ourselves us we we'd we'll we're weve we've

References

[1] Artz, A. F., and Armour-Thomas, E. Development of a cognitive-metacognitive framework for protocol analysis of mathematical problem solving in small groups. *Cognition and Instruction 9*, 2 (1992), 137 – 175.

[2] Baldwin, T. T., Bedell, M. D., and Johnson, J. L. The Social Fabric of a Team-based M.B.A. Program: Network Effects on Student Satisfaction and Performance. *The Academy of Management Journal 40*, 6 (1997), 1369–1397.

[3] Borgatti, S. P., and Cross, R. A Relational View of Information Seeking and Learning in Social Networks. *Management Science 49*, 4 (2003), 432–445.

[4] Christie, R., and Geis, F. *Studies in Machiavellianism*. Academic Press, New York, NY, 1970.

[5] Chung, C. K., and Pennebaker, J. W. The psychological functions of function words. In *Social communication: Frontiers of social psychology*, K. Fiedler, Ed. Psychology Press, 2007, pp. 343–359.

[6] Chung, C. K., and Pennebaker, J. W. Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality 42* (2008), 96–132.

[7] Cronbach, L. J. Coefficient Alpha and the Internal Structure of Tests. *Psychometrika 16*, 3 (1951), 297–334.

[8] Cross, R., and Parker, A. *The Hidden Power of Social Networks*. Harvard Business School Press, Boston, MA, 2003.

[9] Dale, R. The role of punctuation in discourse structure. In *AAAI Fall Symposium on Discourse Structure in Natural Language Understanding and Generation* (Menlo Park, CA, USA, 1991), Association for the Advancement of Artificial Intelligence, pp. 13–14.

[10] DiMicco, J. M., and Yankelovich, N. Constellation: Using visualization to find the path to experts. In *GROUP '07: Proceedings of the 2007 International ACM Conference on Supporting Group Work* (New York, NY, USA, 2007), ACM.

[11] Eckert, P. Variation, convention, and social meaning. In *Annual Meeting of the Linguistic Society of America* (2005).

[12] Ehrlich, K., Lin, C.-Y., and Griffiths-Fisher, V. Searching for experts in the enterprise: combining text and social network analysis. In *GROUP '07: Proceedings of the 2007 International ACM Conference on Supporting Group Work* (New York, NY, USA, 2007), ACM, pp. 117–126.

[13] Fischer, U., McDonnell, L., and Orasanu, J. Linguistic Correlates of Team Performance: Toward a Tool for Monitoring Team Functioning During Space Missions. *Aviation Space and Environmental Medicine 78* (2007), 86–95.

[14] Gosling, S. D., Rentfrow, P. J., and Jr., W. B. S. A very brief measure of the big five personality domains. *Journal of Research in Personality 37* (2003), 504–528.

[15] Hansen, M. T. Knowledge Networks: Explaining Effective Knowledge Sharing in Multiunit Companies. *Organization Science 13*, 3 (2002), 232–248.

[16] Hirokawa, R. Y. Group communication and problem-solving effectiveness: An investigation of group phases. *Human Communication Research 9*, 4 (1983), 291–305.

[17] Hossain, L., Wu, A., and Chung, K. K. S. Actor centrality correlates to project based coordination. In *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (New York, NY, USA, 2006), ACM, pp. 363–372.

[18] Huffaker, D., Swaab, R., and Diermeier, D. The language of coalition formation in online multiparty negotiations. In *Proceedings of the International Communication Association (ICA) Conference* (2008), pp. 22–26.

[19] Ibarra, H., and Andrews, S. B. Power, Social Influence, and Sense Making: Effects of Network Centrality and Proximity on Employee Perceptions. *Administrative Science Quarterly 38* (1993), 277–303.

[20] Issacs, E., Walendowski, A., Whittaker, S., Schiano, D. J., and Kamm, C. The character, functions, and styles of instant messaging in the workplace. In *Proceedings of CSCW02* (New York, NY, USA, 2002), ACM, pp. 11– 20.

[21] Jonassen, D. H., and Kwon, H. I. Communication patterns in computer-mediated vs. face-to-face group problem solving. *Educational Technology: Research and Development 49*, 10 (2001), 35–52.

[22] Julsrud, T. E. Core/periphery Structures and Trust in Distributed Work Groups: A comparative case study. *Structure and Dynamics: eJournal of Anthropological and Related Sciences 2*, 2 (2007), 1–28.

[23] Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., and Graesser, A. C. The language of status hierarchies. *In Review*.

[24] Knoke, D., and Yang, S. *Social Network Analysis*. Sage Publications, Los Angeles, CA, 2008.

[25] Lave, J., and Wenger, E. *Situated learning: Legitimate peripheral participation*. Cambridge University Press, New York, NY, 1991.

[26] Leshed, G., Hancock, J. T., Cosley, D., McLeod, P. L., and Gay, G. Feedback for guiding reflection on teamwork practices. In *GROUP '07: Proceedings of the 2007 International ACM Conference on Supporting Group Work* (New York, NY, USA, 2007), ACM, pp. 217–220.

[27] Linebarger, J. M., Scholand, A. J., Ehlen, M. A., and Procopio, M. J. Benefits of synchronous collaboration support for an application-centered analysis team working on complex problems: a case study. In *GROUP '05: Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work* (New York, NY, USA, 2005), ACM, pp. 51–60.

[28] Maass, A., Salvi, D., Arcuri, L., and Semin, G. Language Use in Intergroup Contexts: The Linguistic Intergroup Bias. *Journal of Personality and Social Psychology 57*, 6 (1989), 981–993.

[29] Milroy, L., and Milroy, J. Social network and social class: towards an integrated sociolinguistic model. *Language in Society 21*, 1 (1992), 1–26.

[30] Paolillo, J. C. Language variation on Internet Relay Chat: A social network approach. *Journal of Sociolinguistics 5*, 2 (2001), 180–213.

[31] Pentland, A. *Honest Signals: How They Shape Our World*. MIT Press, Cambridge, MA, 2008.

[32] Postmes, T., Spears, R., and Lea, M. The Formation of Group Norms in Computer-Mediated Communication. *Human Communication Research 26*, 3 (2000), 341–371.

[33] Rosenberg, M. *Society and the Adolescent Self-Image*. Princeton University Press, Princeton, NJ, 1965.

[34] Rubini, M., and Semin, G. R. Language use in the context of congruent and incongruent in-group behaviours. *British journal of social psychology 33* (1994), 355–362.

[35] Scholl, J., McCarthy, J., and Harr, R. A Comparison of Chat and Audio in Media Rich Environments. In *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (New York, NY, USA, 2006), ACM, pp. 323–332.

[36] Scott, J. *Social Network Analysis - a handbook*. Sage Publications, Los Angeles, CA, 2000.

[37] Sexton, J. B., and Helmreich, R. L. Analyzing cockpit communications: The links between language, performance, and workload. *Human Performance in Extreme Environments 5* (2000), 63–68.

[38] Shin, N., Jonassen, D. H., and McGee, S. Predictors of well-structured and ill-structured problem solving in an astronomy simulation. *Journal of Research in Science Teaching 40*, 1 (2001), 6–33.

[39] Sparrowe, R. T., Liden, R. C., Wayne, S. J., and Kraimer, M. L. Social Networks and the Performance of Individuals and Groups. *The Academy of Management Journal 44*, 2 (April 2001), 316–325.

[40] Stubblefield, W. A., Liszka, A. H., and Kleban, S. D. A collaborative knowledge management system for concurrent design and manufacturing. In *Proceedings10th Annual ISPE International Conference on Concurrent Engineering: Research and Applications* (2003), International Society for Productivity Enhancement.

[41] Stubblefield, W. A., Liszka, A. H., and Kleban, S. D. Preserving multiple expert voices in scientific knowledge management. In *Proceedings DUX 2003: Designing User Experiences* (New York, NY, USA, 2003), ACM.

[42] Tausczik, Y. R. Linguistic analysis of workplace computer-mediated communication. Master's thesis, The University of Texas at Austin, August 2009.

[43] Tichy, N. M., Tushman, M. L., and Fombrun, C. Social Network Analysis For Organizations. *The Academy of Management Journal 4*, 4 (October 1979), 507–519.

[44] Waseleski, C. Gender and the use of exclamation points in computer-mediated communication: An analysis of exclamations posted to two electronic discussion lists. *Journal of Computer-Mediated Communication 11*, 4 (2006).

[45] Weick, K. E. *Sensemaking in organizations*. Sage Publications, Los Angeles, CA, 1997.

[46] White, S., and Smyth, P. Algorithms for Discovering Relative Authority in Graphs. In *KDD '03: The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2003), ACM, pp. 266–275.

Distribution

1		Professor and Chair James W. Pennebaker Department of Psychology University of Texas at Austin Austin, TX 78712
1		Yla Tausczik Department of Psychology University of Texas at Austin Austin, TX 78712
1	MS 0576	John Ganter
1	MS 0576	David Melgaard
1	MS 1318	Bruce Hendrickson
1	MS 9159	Philip Kegelmeyer
2	MS 9018	Central Technical Files, 8944
2	MS 0899	Technical Library, 4536
1	MS 0161	Legal Intellectual Property, 11500
1	MS 0188	D. Chavez, LDRD Office, 1011