Technical Report 1218

IkeNet: Social Network Analysis of E-mail Traffic in the Eisenhower Leadership Development Program

MAJ Ian McCulloh, 2LT Grace Garcia, 2LT Kelsey Tardieu, 2LT Jennifer MacGibbon, Heather Dye, MAJ Kerry Moores, and LTC John Graham
U.S. Military Academy, West Point NY

Daniel B. Horn
U. S. Army Research Institute

November 2007

United States Army Research Institute for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

MICHELLE SAMS, Ph.D.
Director

Technical review by

John S. Barnett, U.S. Army Research Institute
Nehama E. Babin, U.S. Army Research Institute

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED (from. . . to)
November 2007	Final	August 2006 – May 2007

4. TITLE AND SUBTITLE

IkeNet: Social Network Analysis of E-mail Traffic in the Eisenhower Leadership Development Program

5a. CONTRACT OR GRANT NUMBER

5b. PROGRAM ELEMENT NUMBER
611102

6. AUTHOR(S)

MAJ Ian McCulloh, 2LT Grace Garcia, 2LT Kelsey Tardieu, 2LTJennifer MacGibbon, Heather Dye, MAJ Kerry Moores, LTC John M. Graham (U.S. Military Academy) and Daniel B. Horn (U.S. Army Research Institute)

5c. PROJECT NUMBER
B74F

5d. TASK NUMBER
2903

5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Department of Mathematical Sciences
United States Military Academy
626 Swift Road
West Point, NY 10996

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway
Arlington, VA 22202-3926

10. MONITOR ACRONYM
ARI

11. MONITOR REPORT NUMBER
Technical Report 1218

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Subject Matter POC: Dr. Daniel Horn

14. ABSTRACT *(Maximum 200 words)*:

Social network analysis (SNA) has become an important analytic tool for analyzing terrorist networks, friendly command and control structures, and a wide variety of other applications. In this project we collect social network data from a group of 24 Army officers in a one-year graduate program at Columbia University. In this report we discuss methodological issues associated with collecting e-mail social networks and include source code for an add-in to Microsoft Outlook to aid in this process. These data were investigated for patterns and trends in mutual, asymmetric, and null dyads. Behavioral changes in the group resulting from awareness of one's position in social network were also studied. Additionally, comparisons were made between SNA data derived from e-mail traffic and from questionnaires. The differences between these two types of networks are important concerns when considering the implementation of SNA as a command and control tool for friendly forces.

15. SUBJECT TERMS

Social Network Analysis, e-mail, dyad, command and control

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited		Ellen Kinzer Technical Publications Specialist 703-602-8047

Technical Report 1218

IkeNet: Social Network Analysis of E-mail Traffic in the Eisenhower Leadership Development Program

MAJ Ian McCulloh, 2LT Grace Garcia, 2LT Kelsey Tardieu, 2LT Jennifer MacGibbon, Heather Dye, MAJ Kerry Moores, and LTC John Graham
U.S. Military Academy

Daniel B. Horn
U.S. Army Research Institute

Basic Research Unit
Paul A. Gade

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

November 2007

Army Project Number
611102B74F

Personnel, Performance and Training

ACKNOWLEDGEMENT

The authors would like to thank Dr. Kathleen Carley and the Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University for their support on this project. Their efforts and custom modifications to the Organizational Risk Analyzer (ORA) software were essential to the successful completion of this research.

The authors also would like to thank Richard Freytag of Freytag Industries LLC. Mr. Freytag wrote several custom programs to make the data collection on this project much faster and more efficient. He was always available and provided rapid and excellent support.

This project would never have been possible without the initial efforts of MAJ Steve Henderson of the Department of Systems Engineering at the U.S. Military Academy. While in the process of returning from combat operations in Iraq and starting a Ph.D. at Columbia University, Steve wrote, tested, and debugged the initial patch that was used to collect data from the Army officers who participated in this research.

The authors are equally appreciative of MAJ Dennis O'Neill and the 24 officers in the Eisenhower Leadership Development Program who graciously allowed investigators to monitor their personal and professional e-mail traffic over the course of the academic year.

IKENET: SOCIAL NETWORK ANALYSIS OF E-MAIL TRAFFIC IN THE EISENHOWER LEADERSHIP DEVELOPMENT PROGRAM

EXECUTIVE SUMMARY

Research Requirement:

Network Science is an emerging discipline with many potential applications for the Department of Defense. In the 2005 National Research Council "Network Science" Report, recommendation #1 stated,

> "The federal government should initiate a focused program of research and development to close the gap between currently available knowledge about networks and the knowledge required to characterize and sustain the complex global networks on which the well-being of the United States has come to depend." (p. 4)

Network Science differs from classical scientific methods in that it views the subject matter as being made up of many interacting entities that are called nodes. One application area of Network Science that has become extremely popular is Social Network Analysis (SNA). SNA looks at groups of people and their interactions. Social Network Analysis is a methodology that does a very good job of explaining much of the complex behavior of social groups. This work focuses specifically on the communication patterns of one particular group of Army officers in a one-year graduate program at Columbia University as they prepare to become tactical officers at the U.S. Military Academy (USMA).

This is the first investigation in a five-year strategic social network project that monitors the e-mail activity among officers in this graduate program. This first report focuses on the logistics of data collection, the properties of the observed networks, and the relationship between e-mail networks and those generated through traditional questionnaire methods. The value of SNA as a tool for command and control of friendly forces relies on these techniques. This report represents a first step in their use.

Procedure:

This project collected a rich and innovative network data set by monitoring the e-mail traffic of 24 Army officers in a one-year graduate program at Columbia University. The collected e-mail data formed a dynamic social network, connecting the senders and receivers of e-mail messages. The content of the e-mail messages was not collected, meaning that all findings in this research were based on communication patterns rather than communication content. This report describes the implementation of this data collection, an analysis of the network properties of the e-mail communications, and an evaluation of the relationship between self-reported networks and e-mail based networks.

Findings:

The process of data collection overcame several technical and organizational hurdles, and has been refined to improve both the comprehensiveness and ease of collection. The networks themselves were discovered to be strongly influenced by academic demands, and differed significantly from uniform random networks. Additionally, the patterns of communication in the e-mail networks enabled the detection of formal leaders more effectively than self-reported networks.

Future research on electronic communication will be extremely important to the development of predictive social network models. This project has revealed many lessons in both the technical implementation of this kind of project, as well as the behavior of e-mail networks.

Utilization and Dissemination of Findings:

The e-mail conversion utility is currently implemented in versions of Organizational Risk Analyzer (ORA), which is an ARI funded SNA package maintained by Carnegie Mellon University. This software is free for government use. ORA software allows analysts to run statistics on many social network measures and is a valuable resource for future work in this area. This software, as a result of the modifications from this research, would be extremely useful for monitoring e-mail traffic in a military organization to monitor the social networks of organizational members. Appendix A of this report includes the source code for the software patch that was used for data collection. This will enable other organizations to more easily apply the techniques described in this report.

This research is an important emerging area of Network Science. As such it has been presented at a variety of academic conferences. Some of these conferences include:

- ARI-USMA Network Science Workshop, 18-20 April 2007;
- Service Academy Student Mathematics Conference, 18-19 Apr 2007;
- 8th Annual European Social Network Conference, 30 Apr-4 May 2007;
- ELICIT Tool Set Conference (OSD NII), 4 May 2007;
- DoD Human Factors Engineering Technical Advisory Group Meeting, 16 May 2007; and
- NetSci International Workshop and Conference on Network Science, 20-25 May 2007.

IKENET: SOCIAL NETWORK ANALYSIS OF E-MAIL TRAFFIC IN THE EISENHOWER LEADERSHIP DEVELOPMENT PROGRAM

CONTENTS

LIST OF TABLES

LIST OF FIGURES

IkeNet Data Collection: Technical Implementation

Introduction

E-mail has significantly changed how people communicate and interact. In many ways communication is easier and more reliable with e-mail, however, there are many new communication challenges introduced. In order to integrate the use of this new communication tool into Army protocol, leaders need to be aware of the strengths and weaknesses of e-mail. As a result, Army leaders will be able to use this resource to their advantage to distribute information to subordinates.

E-mail provides several advantages over classic communication methods. Peers that were once unreachable due to distance, conflicting time schedules, or other duties, can now be reached with ease using e-mail. If the intended recipient is not available, the message is stored in an in-box until it can be read. Leaders are able to issue guidance and direction efficiently to a much lower level or rank than before. While it was once infeasible for a Brigade Commander to assemble the brigade and issue guidance, except in rare circumstances, he/she can now e-mail everyone in the organization. Those Soldiers, who wouldn't have been present due to some other duty, are able to read the guidance or directive later when they have time. Soldiers are able to have greater access to headquarters elements such as personnel, finance, and supply, through the use of e-mail and Web based services. All of these advantages can greatly improve the efficiency of military organizations. However, there are new challenges that must be understood before the Army can efficiently modify doctrine to exploit the advantages e-mail offers.

With each advantage e-mail brings, there are challenges that must be understood as well. The increase in peer to peer collaboration allows problems to be solved at a much lower level or rank than before. Unfortunately, military commanders may now be unaware of issues and problems in their organization. While it may be good for subordinates to solve problems, it does not allow commanders to provide guidance or take advantage of their experience. It is great for commanders to reach all of their subordinates at once; however, if that is the case, what role do intermediate commanders now play? How does the commander know how his or her Soldiers are receiving this guidance? How does a commander handle feedback, questions, and recommendations from a large volume of subordinates?

Social Network Analysis (SNA) offers a valuable methodology to understand e-mail communication. A social network is a collection of individuals and their relationships. In SNA, we analyze groups of individuals and their relationships mathematically to gain insight about their behavior. Some examples of the networks that can be studied include those of friendship, respect, migration, biological relation, or in this project e-mail communication. In a network, an individual and his or her actions are dependent on other network members. A person might e-mail someone in response to an e-mail they received. Before e-mail networks are described in detail, a background in social networks is presented.

We describe a social network in terms of its members and their connections. Members of the network can be connected through these relations as groups or individuals. In the following

figures, the shaded circles (or nodes) represent members of a social network. The lines (or edges) between these nodes represent their relations.

The connections between members can be either directional or non-directional relationships. An example of a directional social network is very similar to a military chain of command (see Figure 1). This type of social network has a directed flow of information. The commander gives orders or directions to his subordinate units, so the arrows point down to the lowest level. This diagram is also representative of a connected graph where everyone is in the network. Every node in the diagram receives information from a superior.

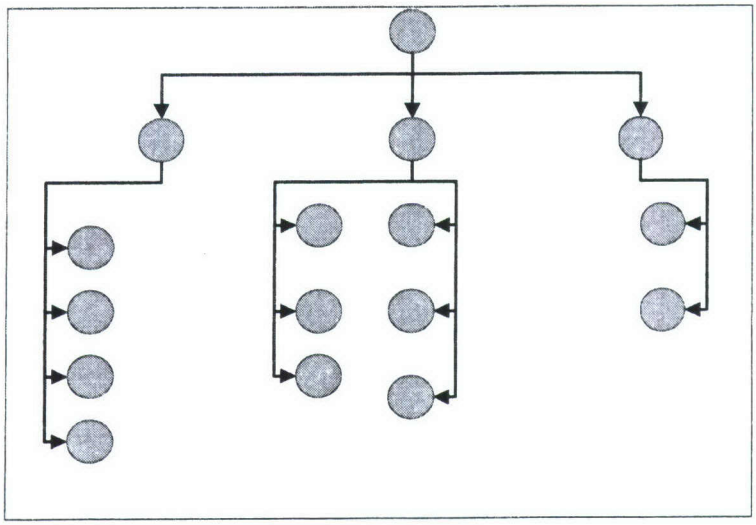


Figure 1. Chain of command network.

The family and friendship network shown in Figure 2 is, by contrast, an example of a non-directional, non-connected graph. In this figure, we can see that there is no predictable structure to the social network because there are no restrictions when talking to family members. For example, the youngest child may talk to anyone in the family whereas in the military a private cannot simply talk to a general. This graph is disconnected because a particular family member may not talk to anyone, such as a disgruntled teenager, yet he/she is still part of the family.

These are only two examples of the many social networks that exist. Applications of social network analysis are beneficial not only in the military, but also in other organizations. For example, a mayor may want to conduct an analysis of a local neighborhood to improve cohesiveness, or a CEO may want to conduct some analysis to improve a department in his company.

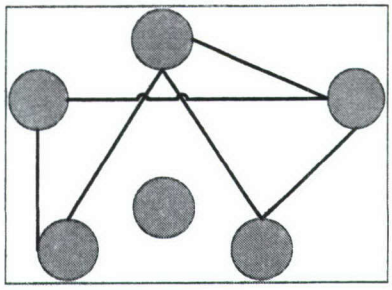


Figure 2. Family or friendship social network.

The present research focuses on the application of SNA techniques to understand a group of US Army officers. This effort focuses on the use of e-mail communication patterns as the basis for these analyses. This report is broken down into three parts: first, we will discuss the implementation of the data collection effort; second, we will explore the characteristics of the collected networks; and third, we will explore the relationship between e-mail networks and networks generated through self-report questionnaires.

This first section focuses on the practical implementation of the data collection and the organizational and technical challenges that must be overcome for such data collection to occur. It concludes with suggestions for efficient, ethical data collection, and serves not only as the basis for the remainder of this report, but also for the continued collection of data for the remainder of the IkeNet project.

Participants

This project monitors the e-mail traffic of 24 mid-grade Army officers, selected to serve as company tactical officers at the United States Military Academy. Before serving as tactical officers, they go through a one-year Master's Degree in Leadership program at Columbia University. Students in the program take some classes on the Columbia main campus and some classes in the Department of Behavioral Sciences and Leadership at the U.S. Military Academy. The program is called the Eisenhower Leadership Development Program (ELDP), thus our project is called IkeNet.

The overall demographics consisted of 21 males, 3 females; all active duty officers. Participants were treated in accordance with ethical standards established by the American Psychological Association. The research methods used in this experiment were approved by the U.S. Military Academy Human Subjects Use Committee.

The ELDP students willing to participate in this investigation permitted us to place a patch onto their Microsoft Outlook e-mail accounts. This patch allowed us to collect information from e-mails in their sent items folder, found on participants' personal computers. The information included: the time each e-mail was sent, the FROM/TO/BC/CC e-mail addresses/names, and the subject line. We were not able to see the content of any e-mail itself. The patch was triggered by two conditions: 1,000 minutes of elapsed time since the previous

trigger and sending an e-mail. Upon triggering, the patch would search the subject's sent mail folder and compile information into a comma-separated values(CSV) spreadsheet from all sent e-mails since the last search. The patch then initiated an e-mail to send the CSV spreadsheet to the principal investigator. The principal investigator would then compile the spreadsheets from all 24 subjects into one master spreadsheet and ensure anonymity of the names. The anonymous names are coded as P1 through P24.

In addition to the e-mail data, the ELDP coordinator provided other sources of important data. The planning calendar for the entire year in which the data were collected was provided. This calendar included graded events and important dates that would potentially have an impact in the e-mail communication of the subjects, such as Army football games, Thanksgiving, and Christmas. Army home football games are big events at USMA in which the entire corps of cadets and tactical officers are required to attend. The ELDP coordinator also administered several surveys to the subjects to compare e-mail communication with self-reported social network data. Finally, several subjects were interviewed to gain greater insight into the activities of the subjects and to identify workload demands on the subjects.

Outlook Visual Basic Patch

The first research task in this project was to create a method for collecting e-mail data for SNA. A Visual Basic for Applications (VBA) program was written that could be installed on a personal computer (PC) in the session window of Microsoft Outlook. The code is included in Appendix A. It is installed on Outlook by opening the program, selecting Alt-F11, and copying the code into the session window. Once the code is installed in this manner, that PC will send regular CSV spreadsheets of e-mail communication to the principal investigator defined in the code.

There were several advantages to the VBA approach to the problem. Microsoft Outlook is proprietary software that causes difficulty in allowing a program such as Organizational Risk Analyzer (ORA) to directly pull information from a subject's outbox. The VBA patch is an easier software implementation. Another advantage is that the patch is installed on each individual subject's computer; therefore, the people who run the e-mail exchange server do not have to be involved in the research project. It often can be difficult, in research, to be dependent upon another group of people for your data; they do not necessarily care about your research. Researchers at the Center for Computational Analysis of Social and Organizational Systems (CASOS) have written software that allows the same e-mail information to be pulled directly from the exchange server as an extension of this project. The final advantage offered by the VBA patch is the control the subject feels in the project. The subject actually installs the patch himself or herself. The subject then feels they have control over the information they send. Most of the subjects knew how to remove the patch when their participation in the project ended. Several subjects said they felt more comfortable knowing that the software sending the principal investigator information was on their computer, and "Big Brother" wasn't pulling their information from somewhere else.

Automation of Data Collection

Data compiling and collection needed to be automated. There were 24 subjects sending daily e-mail to the principal investigator. This meant saving 24 files from e-mail to a hard drive

every day. The principal investigator would have to wait at least a week before compiling the data because there would be an occasional subject who would not log on to e-mail for a day or two. This would cause missing data, until the subject eventually did log in and send a bigger data file. Data were typically compiled once per month throughout the project. Once the files were saved on a hard drive, each file would be opened and the contents copied and pasted into a master file. This process merged all of the data from all 24 subjects for a one-month period into a single data file. With 24 subjects and 25 average files per month, this meant processing through 600 files, which would have taken the principal investigator several hours to complete. Instead, significant time-saving software was developed by Freytag Industries LLC to deal with this complication.[1]

SNA was performed on the data using the proprietary software ORA owned by the CASOS and free for government use. The CASOS created several custom modifications to their software to facilitate this research. The software now has a feature that loads e-mail social network data directly into ORA using the format of the CSV file from the VBA Outlook patch. Once data files were compiled, they were easily loaded into ORA for analysis. Investigators on this project sorted data for different time and date periods and looked at certain key subjects in the group using Microsoft Excel and were then able to easily load these files into ORA. Other software modifications were user-preference and made the software much easier for a novice to use. Several students at the USMA began conducting analysis on social network data with only a couple of hours of familiarization with ORA, thanks to these custom modifications.

Methodology

This research effort was highly successful at collecting a unique sensitive data set for 31 weeks, running from 15 October 2006 through 11 May 2007. We did not collect a full 52 weeks of data for multiple reasons. The subjects began their Master's program in June 2006. There were several delays in debugging the initial Outlook Patch, in gaining approval for the project, and in installing the patch on all subjects' computers. The first week of full data from all 24 participants began 15 October 2006. The subjects graduated from their Master's program on 11 May 2007, thus ending their participation in the project.

Collecting the data in the format of a CSV spreadsheet was excellent. This format allowed multiple future investigators and undergraduates using data for course objectives to easily manipulate the data into whatever format they wanted to for analysis. Some investigators

---

[1] Two software programs were developed. The first program was able to install as an option button in Microsoft Outlook. When this button is selected, the program asks the user to specify a directory to look for data files. Once specified, the program extracts all of the CSV spreadsheet data files from all e-mails in the specified Outlook directory and places copies of all of these files in a directory labeled, C:\todays date. The second program was written in Perl. This program merged all of the CSV files into a single file containing all of the information. These two programs took a two- to three-hour task and reduced the time required to less than five minutes.

The final data processing step was to anonymize the data and remove non-subjects. The principal investigator would open the CSV file in Microsoft Excel. He would then use the 'Replace' feature to replace the subject's actual e-mail name with the anonymous name P#. Finally, e-mails sent to non-subjects were deleted from the data and e-mails addressed to a subject and non-subjects, had the non-subjects deleted. This resulted in a CSV file containing all e-mail correspondence between the subjects only.

looked at the whole data set, while others broke it down into monthly files, and still others broke it down into weekly or daily files. Some investigators looked only at the subjects, while others included some common e-mails sent to people outside the 24 subjects. Some investigators just looked at a subset of subjects.

There was minimal performance impact to the subjects' computers. Because the VBA Outlook Patch would search a subject's sent mail folder, the computer could be slow sending the first e-mail of the day when the sent mail folder had many messages. In two cases, the Patch caused Outlook to crash. This problem was eliminated by creating a "back-up sent" folder in Outlook. In these cases, the subject would periodically need to either move the contents of the "sent mail" folder to the "back-up sent" or delete the contents of the "sent mail" folder.

The biggest problem in data collection was the inconsistency of receiving CSV files from all subjects in the ELDP program. Six of the subjects experienced major problems with their computers, likely caused by a virus or some other software issue. These computers were re-imaged/re-formatted. The subjects did not notify the principal investigator until May 2007, so the Outlook Patch was not re-installed and data were lost on these subjects from the time their computers were re-imaged through the remainder of the project. Another issue in consistency was a result of how individuals logged on to their e-mail. Five subjects left their computers at home, off the network, and used webmail exclusively for their e-mail. Rarely were their computers logged onto the network, so their communication was rarely captured in the data set. Subjects that used a Virtual Private Network (VPN) from home were on the network and their data were captured. Two subjects used their Army Knowledge On-line (AKO) e-mail accounts much more frequently than their USMA accounts, and so much of their data were not captured, unless they received an e-mail from another subject. Several recommendations to improve data collection in future efforts are proposed.

Recommendations for Future Efforts

Recommendations for future efforts focus on continuing to collect data on all subjects without a subject "falling out" of the sample. "Falling out" refers to a subject who does not log on to the network where the exchange e-mail server is located, or a subject who has his or her computer re-imaged. One recommendation is to include a "heart-beat" into the Outlook Patch so that the principal investigator knows within minutes when the Outlook Patch is no longer present or working. If each subject was given a CD that contained an install feature for the patch, then a simple automatic e-mail request from the principal investigator asking the subject to re-install the software would be a one-click task. This would get lost subjects back into the project quickly. Subjects that keep their computers at home are slightly more difficult to handle.

There are several potential solutions to subjects that use webmail to check their e-mail. From the perspective of the principal investigator, the simplest would be to coordinate with the office that runs the e-mail exchange server and use the software under development at the CASOS at Carnegie Mellon to pull e-mail data directly off the exchange server. If this is not possible, four other alternatives are proposed.

1. Have subjects who check e-mail via webmail leave their computers on and Outlook running but configure the Outlook options so the e-mail is not deleted from the server.

Instead the e-mail would be downloaded. Using this "heart-beat," the principal investigator would know if the subjects were complying. This solution would not require any additional support or special permissions, because it is a configuration change to Outlook on the client-side.

2. Create a modified Outlook Patch that would equip the subjects' "remote" Web browser to capture the e-mails they browse and only the e-mails they browse. While this could be difficult to implement, it does not require any outside support from those that manage the exchange server.

3. Run a version of Outlook on a locked-down machine (no one can access it after it is set up because it is in a secure location). This machine would log in to each of the subjects' accounts and download a copy of their e-mail, leaving the original on the server. This version of Outlook on the locked-down machine would delete the body of the e-mail and keep the rest. Then, the locked-down machine would process the e-mail using the original Outlook Patch and send the CSV report to the principal investigator. Again, this avoids outside support from those managing the exchange server.

4. Similar to Recommendation 3, one could use a Web browser on a central, locked-down machine to download a copy of subjects' e-mail data. This is not necessarily better or worse than Recommendation 3, but offers another alternative.

There have been many lessons learned from this initial project. Future projects will continue to improve in terms of completeness and better methods for collecting data. The remainder of this technical report describes the results of data analysis conducted on the IkeNet data set. While these two sub-projects provide interesting findings, the investigators hope others will be able to conduct additional research using this data set and the data scheduled for collection over the next four years.

Descriptive Analysis of IkeNet E-Mail Data

E-Mail Data

Several descriptive statistics were used to explore the IkeNet e-mail data. Comparisons are drawn from this project and historical findings of similar social networks in the early 1990s. Understanding some of the findings of the IkeNet e-mail data is important to determine how social network analysis of e-mail data can be used to improve command and control systems.

IkeNet e-mail data were divided into week-long intervals because most academic requirements were assigned at the beginning of the week and due at the end of the week. After sorting the e-mails into weeks, we further narrowed our dataset into four categories: academics, administrative (chain of command and class administrative information), social, and blank subject lines. This allowed isolation of those e-mails that were related to academics. Additionally, we assumed that e-mails without a subject line were not related to academics and thus eliminated them from the dataset. Similarly, only e-mails within the network were included, so if Student P1 sent an e-mail to his wife, or his wife sent an e-mail to him, that communication would be removed from the data set. This limited our network to 24 nodes or students.

Furthermore, we assumed that six weeks gave a representative spread of activities. The six weeks did not include summer break nor administrative weeks during which our academic network group would have no cohesive goal. Finally, we assumed that participants did not change their e-mail communication habits as a result of our data collection.

Analysis and Results

Initial analysis focused on investigating the network properties of the data. It was not clear at the beginning of the investigation whether e-mail communication within a homogenous group of people would appear random, if it would remain relatively consistent from week to week, or if there were identifiable factors that would affect changes in network structure. The data were sorted, and we created digraphs for each week. Wasserman and Faust (1994) define a digraph as "[A] directional relation [that] can be represented by a directed graph, or digraph for short. A digraph consists of a set of nodes representing the actors in a network, and a set of arcs directed between pairs of nodes representing directed ties between actors."

For example, if we study the social network consisting of two friends—Sue and Mel—for any given day, we can visually represent this social network as a digraph (see Figure 3). Suppose Sue initiates a conversation with Mel eight times in a given day, and Mel initiates a conversation with Sue eleven times on the same day. Sue and Mel are the nodes or circles in this graph. The lines between the nodes are the directed edges that represent Sue initiating a conversation with Mel and Mel initiating a conversation with Sue. We can weight a digraph by assigning each edge a numerical label. This label indicates the number of times Sue speaks to Mel. Otherwise the graph is unweighted.

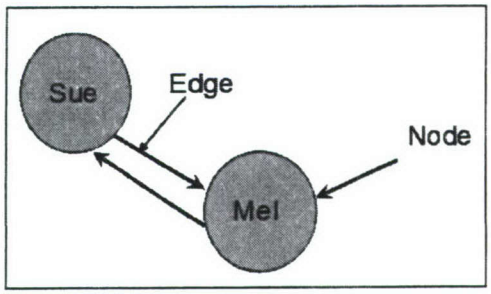


Figure 3. Unweighted social network digraph for Mel-Sue example.

To perform qualitative analysis on this scenario, we can obtain a matrix representation of the digraph (See Figure 4). The matrix on the left in Figure 4 shows a weighted matrix where the actual number of conversations initiated by a given actor is recorded in the matrix. The matrix on the right is an unweighted matrix or binary matrix. In this matrix, we record a "0" if no conversation was initiated or a "1" if one or more conversations were initiated. In our analysis of the academic data, we used a binary matrix.

$$\begin{array}{c|cc} & \text{Sue} & \text{Mel} \\ \text{Sue} & 0 & 8 \\ \text{Mel} & 11 & 0 \end{array} \quad \Rightarrow \quad \begin{array}{c|cc} & \text{Sue} & \text{Mel} \\ \text{Sue} & 0 & 1 \\ \text{Mel} & 1 & 0 \end{array}$$

Figure 4. Weighted and unweighted matrices for Mel-Sue example.

We used the social network analysis program, Organizational Risk Analyzer (ORA), to analyze the data. ORA facilitated our computations by creating the binary matrices and digraphs for every week in the dataset (see Appendix B for digraphs of all six weeks). We labeled these binary matrices, X, for various calculations later. In the matrix X, the term X_{ij} (the entry in the ith row and jth column) indicates e-mail communication between Student i and Student j. Note that it is possible to send an e-mail to yourself, so that $X_{ii} \neq 0$ in some matrices. However, in some of our considerations this is considered an invalid communication and we set $X_{ii} = 0$. Each matrix has dimensions 24 x 24. As shown in Equation 1, we define X as follows:

$$X_{ij} = \begin{cases} 1 & \text{if Student } i \text{ sent email to Student } j \\ 0 & \text{if Student } i \text{ sent no email to Student } j \end{cases} \tag{1}$$

First, we calculated the proportion of academic e-mails out of the total e-mails sent each week. Figure 5 shows a visual representation of the academic e-mails sent for the six weeks.

Knowing the percentage of academic e-mails compared to the total number of e-mails sent in a week allows us to assess how important academics are to the current sample in question. The majority of e-mails sent over the six-week period were related to academics, indicating that academics were a high priority for e-mail communication in this sample of students (see Figure 5).

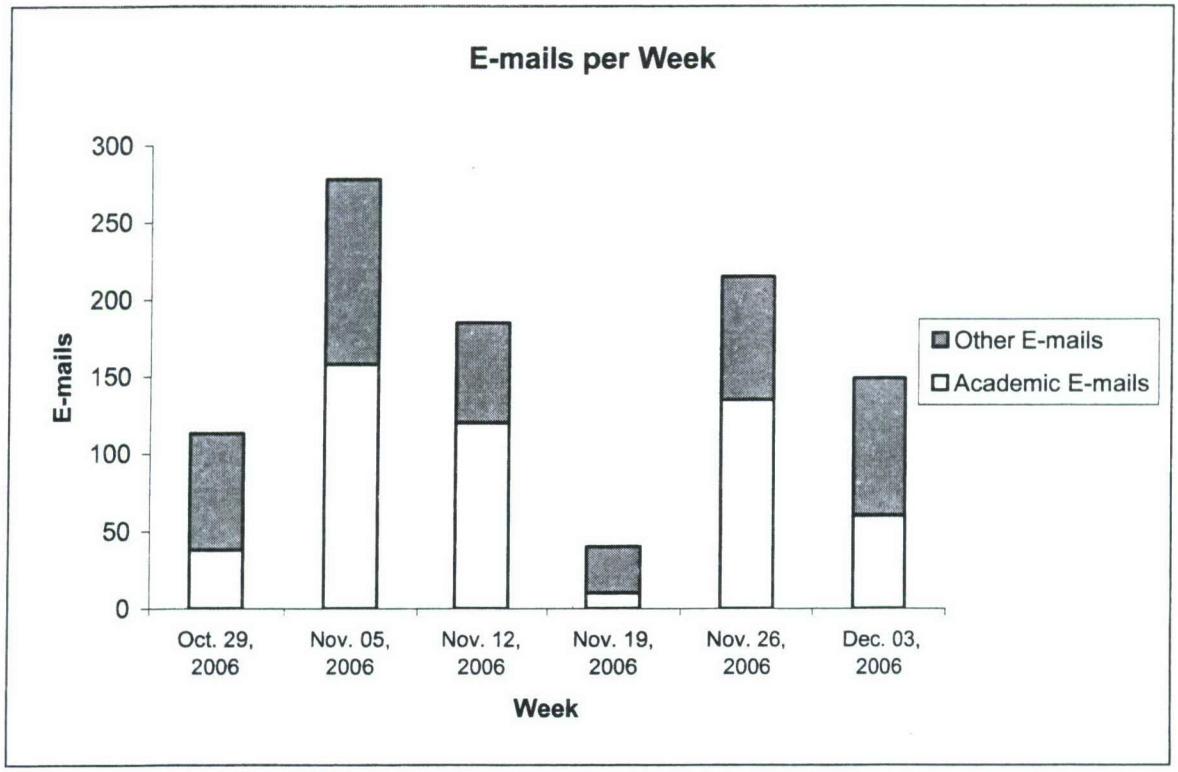


Figure 5. Proportion of total e-mails that were related to academics.

To investigate the structure of the network, we computed the dyad count. A dyad is the communication between two nodes. There are three different types of dyad communication: asymmetric, mutual, and null. Figure 6 shows a visual representation of the different types of dyad communication. In an asymmetric dyad, one node talks to another, but does not receive a response. In Figure 6, there is an asymmetric dyad with nodes 1 and 2. This type of communication could be an example of a group that has members who are sending out information. A mutual dyad would be two nodes communicating such as nodes 3 and 4 in Figure 6. This type of communication might occur in a group that collaborates equally, or one in which subordinates verify or clarify directives. Finally, a null dyad is when two nodes are part of the network and do not have any communication activity such as nodes 1 and 5 in Figure 6. In a dyad count, we conduct a census and tabulate the number of null, mutual, and asymmetric dyads. There are 24 members in the network resulting in 276 combinations of possible dyads of all types.

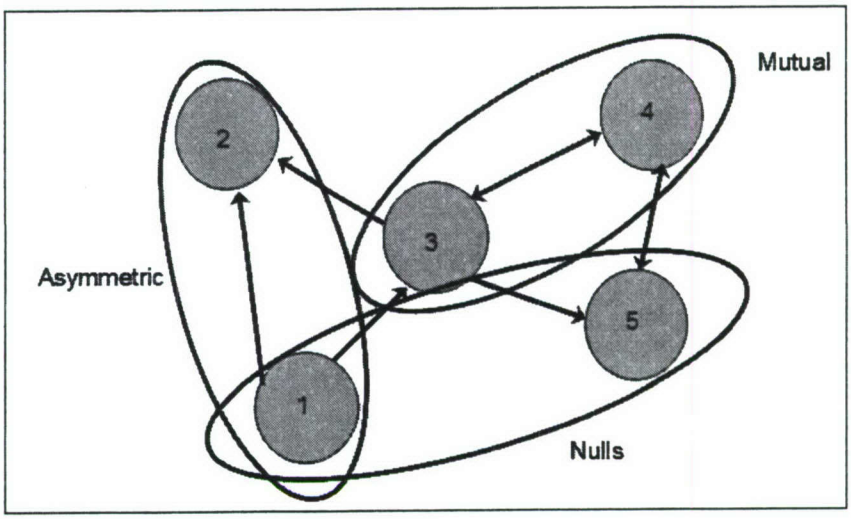


Figure 6. Types of dyad communication.

Given the matrix X for every week, we used *Mathematica,* a mathematical computing program, to compute the dyad count and determine the number of mutual, null and asymmetric dyads. From Wasserman and Faust (1994) we obtained a sequence of formulas to compute the dyad count (Equations 2-4).

Let X denote the binary matrix derived from the digraph of a given week as before and let M denote the number of mutual dyads in a given week. Then,

$$M = \frac{1}{2}Trace[X^2] - \frac{1}{2}Trace[X] \tag{2}$$

Now that we have the mutual count, we can calculate the asymmetric count. The asymmetric count will give us some insight on the structure of the class. A higher asymmetric count indicates that the class might be more organized. Let L denote the number of arcs in a given binary matrix derived from the diagraph (which is the total number of e-mail communications). Let A denote the number of asymmetric dyads in a given week. Then

$$A = L - 2M \tag{3}$$

From asymmetric count, we can calculate the number of nulls.

Let Md denote the number of all possible mutual communications in a given week with 24 agents. Recall that in a network with 24 members, there are 276 possible mutuals. Let N denote the total number of nulls in a week. Then

$$N = Md - M - A \tag{4}$$

Table 1 shows a tabulation of null, mutual and asymmetric counts for every week in the dataset.

Table 1
Mutual, Asymmetric, Null Count Per Week

Week	Mutual	Asymmetric	Null
Oct. 29, 2006	11	91	174
Nov. 05, 2006	71	137	68
Nov. 12, 2006	21	104	151
Nov. 19, 2006	2	10	264
Nov. 26, 2006	24	75	177
Dec. 03, 2006	25	87	164

Combining information from the planning calendar data (Appendix C) and the data from Table 1, we can make some inferences about when certain types of dyad communication were prevalent over a given week (Figure 7). For example, we can see that on the week of Nov 19th, Thanksgiving break, null dyads compose the majority of the dyads. The students only had one class to attend during that week and they may not have had to prepare any work. In addition, over Thanksgiving break, the students did not attend class. From Figure 7, we observe that the week of Nov 5th seems to have the most mutual and asymmetric communication. In this week, the class had many academic requirements due and the Army-Air Force football game was played that weekend. Students may have been planning ahead to attend tailgates or were working together to complete all their requirements before Thanksgiving leave. When we see a large number of asymmetric dyads, it may indicate that leaders are doing most of the talking/coordinating in the network. This chart also allows us to see that there were a larger number of asymmetric dyads than mutual dyads.

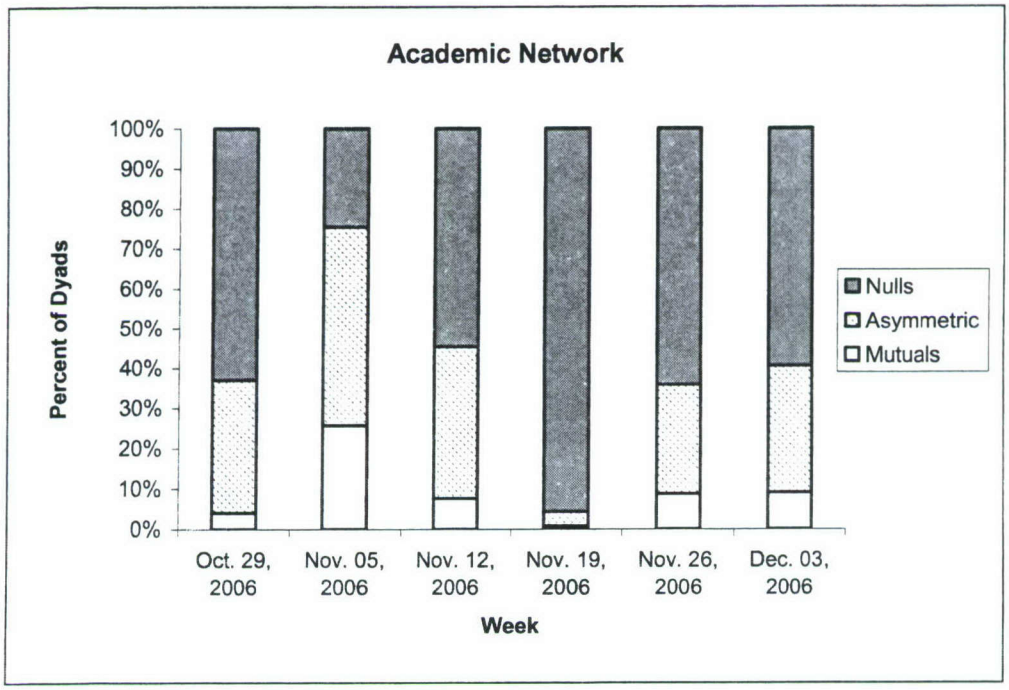


Figure 7. Plot of proportion of different dyad communications by week.

We determined the expected number of mutuals for a random network, $G(24, 0.5)$. In this class of random graph, each directed arc between two nodes occurs with a probability of 0.5. We then compared this expected number to the actual number of mutual dyads present per week from Table 1. Again, let L denote the sum of total arcs in the binary matrix, X, for a given week. Let L_2 denote the sum of squares (*SSE*, Equation 5) for the arcs/conversations initiated by each TOEP. Let g denote the number of subjects in the network, 24. Then:

$$SSE = \sum_{i=1}^{g} \left(\sum_{j=1}^{g} X_{ij} \right)^2 \tag{5}$$

$$E[M] = \frac{L^2 - L_2}{2(g-1)^2} \tag{6}$$

The results are shown in Figure 8. From this we observe that there are fewer mutual dyads than would be expected by chance; that is, a greater number of asymmetric dyads indicate that communication is being directed by leaders. Although some of these are probably determined by professor selections of an S3 (operations officer or class leader), some of the communication may be directed by naturally emergent leaders.

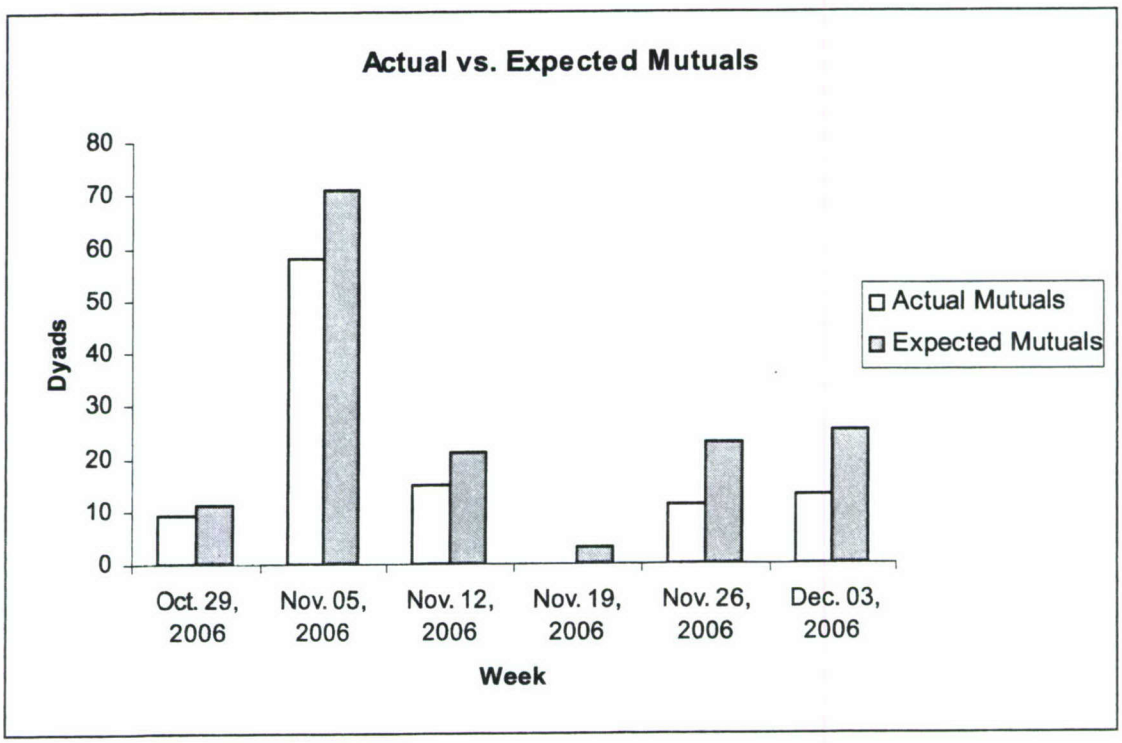


Figure 8. Expected number of mutuals vs. actual number of mutuals.

Then we performed a hypothesis test on the distribution of the edges and of the graphs. The null hypothesis was that edges occurred randomly with a 50% probability. Given this

hypothesis, the graph with 24 nodes should follow a binomial distribution. We therefore tested the hypotheses:

H_o: Edges ~ U(0,1)
H_a: Edges ≠ U(0,1)

H_o: Graph ~ Bin(24*23,0.5)
H_a: Graph ≠ Bin(24*23,0.5)

The test statistic for the first hypothesis is *l*, the sample number of directed edges in the network. If the hypothesis is true, *l* will follow a binomial distribution with parameter $n = 24*23$ (i.e., the total number of possible edges in the network) and parameter $p = 0.5$, which is the assumption of a uniform distribution (Wasserman and Faust, 1994). Table 2 shows the test statistic and p-values for each week.

The test statistic for the second hypothesis is given by:

$$z = \frac{l - E(L)}{\sqrt{Var(L)}} = \frac{l - g(g-1)/2}{\sqrt{g(g-1)/4}} = \frac{l - 276}{11.75} \qquad (7)$$

where *l* is the sample number of directed edges in a social network, E(*L*) is the number of edges that should be present in a network of this size on average if the hypothesis were true, and Var(*L*) is the variance of the edges in the network under the hypothesis (Wasserman and Faust, 1994). It can be seen that *z* conforms to a normally distributed random variable. Table 2 shows the test statistic and p-values for each week.

Table 2
Test Statistics and P-Values for Hypothesis Tests

Week	l	p-value	z	p-value
Oct. 29, 2006	102	0.0000	-14.81	0.0000
Nov. 05, 2006	208	0.0000	-5.79	0.0000
Nov. 12, 2006	125	0.0661	-12.85	0.0000
Nov. 19, 2006	12	0.0000	-22.47	0.0000
Nov. 26, 2006	99	0.0000	-15.07	0.0000
Dec. 03, 2006	112	0.0010	-13.96	0.0000

We reject the null hypothesis that e-mail messages are randomly sent with a probability of 0.5, and thus conclude that the subjects' e-mails are directed to a certain subset of subjects in the group. The class was split up into groups for their projects, yet we can also see other individuals as leaders (one who sends more asymmetric e-mails) in the network. This suggests that there is some organization in the class. This was confirmed from interviews with the subjects. Some classes required the subjects to work together outside of the classroom, and instructors assigned a class leader to disseminate information. All of these reasons support the findings that this social network is not a random network.

In determining that the IkeNet social network is not a uniformly distributed random network, we face the question: How is the IkeNet social network distributed? If we assume that the network follows a binomial distribution, we can create a confidence interval on the probability of two nodes communicating. Confidence intervals that do not overlap will then show statistically significant changes in e-mail communication over the six-week time span. The equation for determining the confidence interval is:

$$\hat{P} \pm z_{\alpha/2}\sqrt{\hat{P}(1-\hat{P})/g(g-1)} \tag{8}$$

where $\hat{P}$ is the maximum likelihood estimate of the unknown parameter p in the assumed binomial distribution and is:

$$\hat{P} = \frac{l}{g(g-1)} \tag{9}$$

and z is the critical value from the standard normal distribution (Wasserman and Faust, 1994). Table 3 shows the confidence intervals for the six weeks investigated above.

Table 3
Confidence Intervals on Parameter P of a Binomial Distributed Network

Week	l	$\hat{P}$	LCL	UCL
Oct. 29, 2006	102	0.37	0.31	0.43
Nov. 05, 2006	208	0.75	0.70	0.80
Nov. 12, 2006	125	0.45	0.39	0.51
Nov. 19, 2006	12	0.04	0.02	0.07
Nov. 26, 2006	99	0.36	0.30	0.42
Dec. 03, 2006	112	0.41	0.35	0.46

The confidence intervals in Table 3 are roughly between the 0.30s and 0.40s with no significant difference. This is seen in the weeks of Oct. 29, Nov. 12, Nov. 26, and Dec. 3. The week of Nov. 05 shows a statistically significant increase in the probability of sending e-mail to more people. This was the week of the Army-Air Force game, which was the largest sporting event held at West Point during the academic year. This week included the most resource intensive social event of the year. In addition, there were heavy academic demands on the subjects. The week of Nov. 19 shows a statistically significant decrease in e-mail traffic. This was Thanksgiving week; it was a short school week. Many subjects left town and were not accessing their computers, and there were very few academic requirements.

Conclusions

We determined that the proportion of academic e-mails seemed to dominate the e-mail communication in the social network. This indicates that the subjects' primary use of e-mail was to communicate regarding academic activities.

Subjects e-mailed each other more frequently as the number of academic demands placed on them increased (see Table 4). We also observed that as classes demanded more group work, their communication increased. Additionally, we discovered that the subjects had more cliques and leaders in the group than we imagined, because they had to work together to complete assignments.

Table 4
Academic Demands Per Week

Week	Classes	Number of Graded assignments	Group assignments	Total Mutual	Total Asymmetric
Oct. 29, 2006	3	1	3	11	91
Nov. 05, 2006	3	1	3	71	137
Nov. 12, 2006	4	0	4	21	104
Nov. 19, 2006	1	0	0	2	10
Nov. 26, 2006	2	1	1	24	75
Dec. 03, 2006	3	0	3	25	87

Our rejection of the hypothesis that the network is a uniformly distributed random network led us to explore other distributions to describe the network. Assuming the subjects' e-mail communication follows a binomial distribution, statistically significant changes in the binomial parameter p can be detected. These changes can easily be related to changing demands in workload and days off.

Finally, it was discovered that the most prevalent form of communication in this time span was asymmetric communication. This tells us that there was some organizational structure to the e-mail communication, possibly resulting from working on projects together.

Behavioral Changes Due to Social Awareness

Social Awareness

In today's technologically advanced world, social networks are less visible. E-mail, instant messaging, text-messaging, and even phone calls make it difficult to have clear knowledge of the social network around the work environment. This social network awareness or workspace awareness is an individual's current understanding of other's interaction within a work environment network. It is important for individuals to be aware of and understand the networks they are embedded in, in order to understand how to work effectively in their workspaces.

According to Gutwin and Greenberg (2004), awareness can be broken down into "who," "what," and "where" components. "Who" awareness is an understanding of who is in the workspace. "What" awareness is an understanding of what the people in the workspace are doing. "Where" awareness relates to where the person works and what they can see from their workspace. Additionally, one must focus on the "when" awareness. "When" awareness is an understanding of what is going on in the workspace over time or at key points in time. Thus, in order to have a basic representation of one's workspace, it is imperative to be aware of the "who, what, when, and where" components.

While awareness of one's workspace is one factor, a person's perception of that workspace is also a factor. In a series of papers, Bernard, Killworth, and Sailer (1980, 1982; Bernard & Killworth, 1977; Killworth & Bernard, 1976, 1979) demonstrated that individuals were generally inaccurate in recalling with whom they interacted. This work has been taken as a demonstration that self-reported social networks are not representative of the 'real' patterns of interaction. While there has been some debate about these conclusions (e.g., Kashy & Kenny, 1990), understanding the perceptions of social networks is important in that it informs our understanding of individuals' beliefs and behaviors.

Lewin referred to an individual's perception of the workspace as a 'psychological field.' He states that full understanding of a person's "psychological field thus cannot result from an 'objective' description by others of what surrounds the person. The crucial factor is the person's own interpretation…the person's own reports typically provide better clues than do the researcher's intuitions." (Fiske and Taylor, 1991). Furthermore, Krackhardt (1990, p. 344)) states that "relationships are often based on people's perception and interpretation and not necessarily on observable, behavioral fact." Therefore, it is important to take individuals' perceptions into consideration in order to gain an accurate view of the workspace.

It is also important to note that in workspaces several people are usually in positions of power. According to Krackhardt, "formal position is significantly related to power and advice centrality." (p. 356) Additionally, "centrality in the informal network itself predicts power" (p. 345). People in positions of power are forced to interact more with others within their workspace, which creates a better understanding of the network.

In addition to understanding the types of awareness that apply to the workspace and an individual's perception of the workspace, one must understand the workspace itself. In today's society, the workspace relies heavily upon technology. The most common forms of workspace technology are e-mail, bulletin board systems, and computer conferencing. This "electronic workplace [is] an organization wide system that integrates information processing and communication activities." (Ellis, Gibbs, & Rein, 1991, p. 39). By using an electronic workplace, face-to-face interaction decreases. The decrease in face-to-face interaction and the increase in electronic interaction transforms once highly visible social networks to less visible networks.

Regarding e-mail interaction alone, just how well does tracking this form of communication accurately depict the workspace social network? In environments where a good deal of face-to-face or telephone interaction occurs, e-mail may not sufficiently capture the patterns of interaction, as only certain types of communications will be tracked.

Method

A between subjects experiment was conducted to determine if there was a significant difference between e-mail communication networks and networks created from surveys. Even more important is to determine if monitoring e-mail communication is an effective method for collecting social network data. E-mail traffic was collected as described in the first section of this report. In addition, a series of three surveys was conducted. The surveys contained 10-12 questions. The first survey consisted mainly of background information on experiences, career, and family. The other two surveys focused more on relationships with other members of the group. Both surveys asked the subjects to identify the three people who they communicated with the most during the current week, the preceding week, and which people in the group they spend time with outside of work. The second survey attempted to define an advice network, by asking the subjects who the top three group members were that they would seek for help with proofreading papers, statistics, military advice, and who they thought were influential members of the group. Survey #3 included a social network built from Survey #2 and asked the subjects to make subjective assessments of which group members were represented by central nodes appearing in the graph. All surveys were given to the students by their instructors and then made anonymous by the principal investigator after the fact. The survey included questions on the e-mail network given to them and created a self-reported network from their responses.

Once data were collected, we conducted a repeated measures analysis in order to monitor behavioral change. We also conducted a pairwise T-test on the Hamming distance between the e-mail and survey networks. Hamming distance is a metric used to measure the difference between two networks of the same size and is calculated as the number of substitutions required to change one matrix into the other (Hamming, 1950). In the current effort, we report standardize Hamming distances by dividing the raw Hamming distance by the total number of possible substitutions, i.e., N(N-1).

Results

The surveys administered to the subjects contained a variety of questions. However, analysis in this report is focused on one survey question, "Which three peers have you

communicated the most with this week?" Social networks were constructed based on the responses to this question and compared to the corresponding social network constructed from the monitored e-mail communication. The first survey administered in October, 2006 did not contain this question. The social network constructed from the second survey administered March 27, 2007 is shown in Figure 9. The corresponding social network of e-mail communication for the week beginning March 25, 2007 is shown in Figure 10.

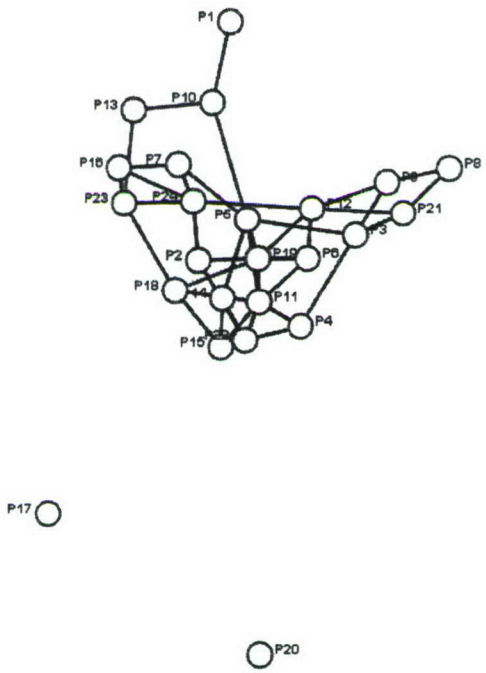


Figure 9. Social network constructed from Survey #2.

The Hamming distance between the self reported network in Survey #2 and the e-mail network is 0.0687. Hamming distance is a binary distance measure between two square matrices. As a binary measure, it does not detect any differences in the quantity of messages sent between two nodes. This makes the measure easily biased by the density of the network. Denser networks will have smaller, or closer, Hamming distances. Therefore, the Hamming distances between e-mail networks and the Hamming distance between the two self reported networks were investigated. The social network constructed from Survey #3, administered April 12, 2007, is shown in Figure 11.

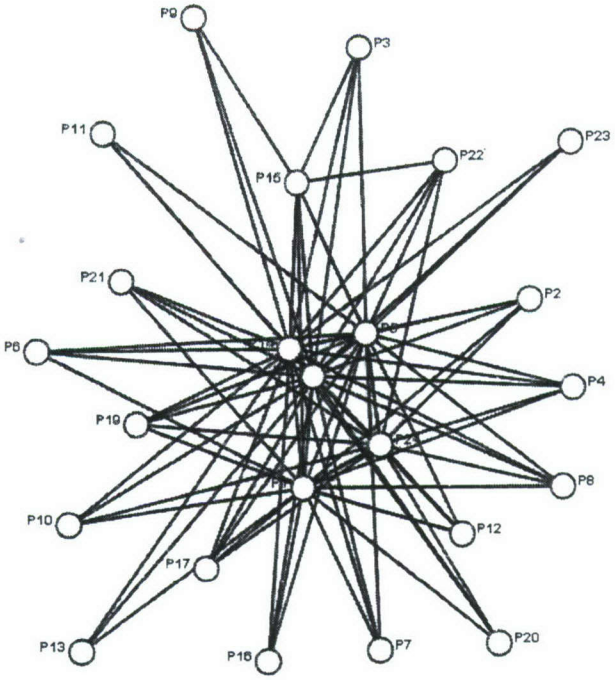


Figure 10. Social network constructed from e-mail communication.

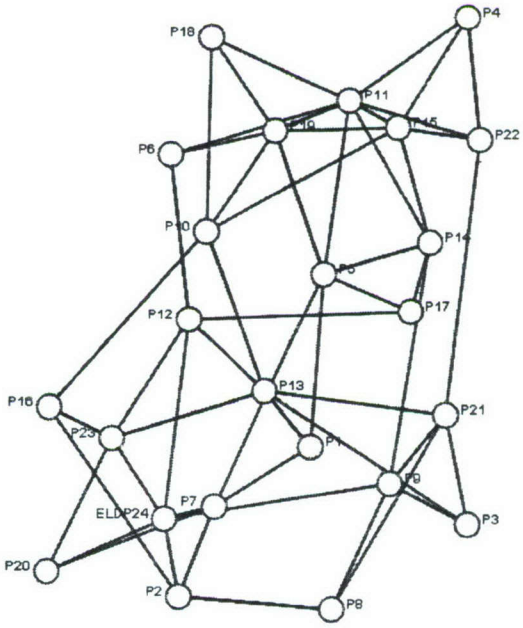


Figure 11. Social network constructed from Survey #3.

The Hamming distance between the social networks constructed for Survey #2 and Survey #3 is 0.1014. This is a 150% increase in distance compared to the difference between the e-mail network and the self-reported network. Hamming distances between weekly e-mail social networks are displayed in Table 5.

Table 5
Hamming Distance Between Weekly E-mail Social Networks

	20070121	20070128	20070204	20070211	20070218	20070225	20070304	20070311
20070121	-							
20070128	0.00763013	-						
20070204	0.01078367	0.00804698	-					
20070211	0.00973249	0.00685080	0.01029433	-				
20070218	0.01085617	0.00793823	0.01130927	0.01025808	-			
20070225	0.01350225	0.01058431	0.01377410	0.01283167	0.01359287	-		
20070304	0.01174424	0.00871756	0.01216108	0.01107366	0.01208859	0.01462592	-	
20070311	0.00652458	0.00357039	0.00701392	0.00589024	0.00683268	0.00951501	0.00768450	-
20070325	0.01216108	0.00917065	0.01239669	0.01141801	0.01236045	0.01478904	0.01328476	0.00810135

The average Hamming distance between e-mail networks is 0.010253, which is 15% of the distance between the self-reported network and the corresponding week of e-mail communication. The 95% confidence interval on the Hamming distance is 0.0093, 0.0112.

Additional comparisons were performed using social network measures and rankings of the most influential members of different networks. The four measures of influence investigated were Betweenness Centrality, Closeness Centrality, Eigenvector Centrality, and Total Degree Centrality. Betweenness is a measure of the frequency a node lies on the shortest path between two other nodes. A node high in betweenness is often a broker of information or a gate keeper. Closeness is a measure of how close a node is to all other nodes in the network. A node high in Closeness can pass information quickly to all members of the network and is likely to be more situationally aware than other nodes. Eigenvector Centrality measures how connected a node is to other connected nodes. A node high in Eigenvector Centrality tends to be the center of a clique or subgroup. Total Degree measures how many other nodes are directly connected. Analysis of these measures for different networks was focused on the top third (8) of nodes in the network. Many times nodes on the periphery of the network were tied in their scores for certain measures. These ties would bias correlation and other statistics. Therefore, focusing on the top third allows more meaningful analysis. Tables 6-9 show the measures and scores for the top third of nodes in each measure.

Table 6
Rankings of Betweenness in Social Networks

Node Rank	Survey #2		E-mail		Survey #3	
1	P24	0.195690	P5	0.057971	P13	0.140679
2	P12	0.192754	P15	0.043874	P23	0.120487
3	P19	0.161881	P1	0.043808	P24	0.095191
4	P11	0.148758	P18	0.026087	P2	0.091634
5	P3	0.143874	P14	0.026087	P12	0.082675
6	P4	0.133634	P24	0.021146	P7	0.064229
7	P6	0.132449	P12	0.000395	P10	0.064229
8	P5	0.122911	Others	N/A	P5	0.061989

Note. There were only 7 individuals with positive betweenness centrality scores.

Table 7
Rankings of Closeness in Social Networks

Node Rank	Survey #2		E-mail		Survey #3	
1	P10	0.176923	P18	1.000000	P3	0.188525
2	P2	0.159722	P14	1.000000	P20	0.181102
3	P9	0.155405	P5	0.793103	P17	0.181102
4	P21	0.155405	P1	0.718750	P9	0.164286
5	P6	0.153333	P24	0.657143	P21	0.163121
6	P24	0.153333	P15	0.575000	P16	0.163121
7	P3	0.152318	P12	0.442308	P18	0.161972
8	P23	0.152318	P9	0.377049	P12	0.159722

Table 8
Rankings of Eigenvector Centrality in Social Networks

Node Rank	Survey #2		E-mail		Survey #3	
1	P11	0.134754	P14	0.094679	P11	0.10023
2	P15	0.104754	P18	0.086491	P22	0.074859
3	P22	0.099019	P1	0.075135	P19	0.072739
4	P18	0.077848	P5	0.074707	P15	0.070100
5	P4	0.067207	P24	0.064460	P5	0.067554
6	P14	0.064349	P15	0.039183	P4	0.056891
7	P19	0.063649	P19	0.037443	P14	0.050516
8	P5	0.062414	P12	0.037443	P7	0.046380

There is no statistically significant correlation found in any of the rankings of measures using either the ordinal rankings or the social network measures. The number of students who were among the highest ranked in *both* the Survey #2 and the E-mail networks was 3 of 8 for Betweenness, 2 of 8 for Closeness, 5 of 8 for Eigenvector Centrality, and 4 of 8 for Total Degree.

Table 9
Rankings of Total Degree in Social Networks

Node Rank	Survey #2		E-mail		Survey #3	
1	P11	0.739130	P18	0.565217	P11	0.673913
2	P24	0.543478	P14	0.565217	P5	0.565217
3	P22	0.500000	P5	0.478261	P24	0.565217
4	P19	0.456522	P1	0.434783	P22	0.543478
5	P15	0.456522	P24	0.369565	P15	0.543478
6	P6	0.413043	P15	0.217391	P19	0.521739
7	P5	0.391304	P12	0.152174	P7	0.500000
8	P3	0.391304	P19	0.108696	P13	0.500000

Two sample T-tests and pairwise T-tests were performed on the average measures of each node in a social network to compare the Survey #2 network with the E-mail network and to compare the Survey #2 network with the Survey #3 network. Table 10 shows the test statistics and p-values for each test for Survey #2 compared to the e-mail network, while Table 11 shows the same information for Survey #2 compared to Survey #3.

Table 10
Statistical Analysis Comparing Average Network Measures Between Survey #2 and E-mail Social Networks

Measure	Two Sample T	p-value	Pairwise T	p-value
Betweenness	3.9142	0.0003	3.8772	0.0008
Closeness	- 1.8983	0.0639	- 1.9104	0.0686
Eigenvector	0.0000	1.0000	0.0000	1.0000
Total Degree	2.6309	0.0115	2.6654	0.0138

It can be seen in Table 10 that there is a statistically significant difference in the rankings of the social network measures for Betweenness and Total Degree; there is marginal difference for Closeness; however, there is not enough evidence to show a significant difference in the rankings of Eigenvector Centrality.

Table 11
Statistical Analysis Comparing Average Network Measures Between Survey #2 and Survey #3 Social Networks

Measure	Two Sample T	p-value	Pairwise T	p-value
Betweenness	1.1733	0.2467	1.2147	0.2368
Closeness	- 0.3361	0.7383	- 0.3017	0.7656
Eigenvector	0.0000	1.0000	0.0000	1.0000
Total Degree	- 1.8198	0.0753	-3.2641	0.0034

Table 11, on the other hand, shows a statistically significant difference between the pairwise difference in Total Degree. This indicates that there is a greater difference between a self-reported network and a corresponding e-mail network, than there is between the self-reported networks from two different time periods. While this might be expected, due to the fact that the two survey networks were generated based on self-report data, while the e-mail network was based on a behavioral measure (e.g. Bernard & Killworth, 1977; Bernard, Killworth, & Sailer 1980, 1982; Killworth & Bernard, 1976, 1979), it is not clear, however, if this difference is a true difference or a bias resulting from the density of the two types of networks. Density is a measure of how many edges are present in a network. It is a percentage of the total possible edges that could exist between a given number of nodes, in this case 24. The e-mail network has a density of 0.1793, while the Survey #2 network has a density of 0.1014 and the Survey #3 network has a density of 0.1304. The fact that the Survey #2 and Survey #3 networks have more similar densities, may make them appear more similar than they actually are. Since they are being described by social network measures that can be biased by density (i.e., Betweenness and Total Degree measures), these two networks may not be as alike as the network measures indicate.

Given the difference between the self-reported network and the e-mail network, one might wonder which is a more accurate representation of the 'true' social network. One might argue that e-mail is a more objective measure, while others may argue that the subjects themselves are better able to know the dynamics of their organization. One way to answer this question was to interview the subjects, ask them what causes a person to be central in their organization and who those people might be. This is a different question than the survey, which asked the subjects who they specifically communicated with in the previous week. The unanimous response from the subjects was that the key leaders were more central in communicating information to the group. There were seven key leaders in the group who had specific responsibilities. These leaders are identified in Table 12.

Table 12
Key Leaders in the IkeNet Group

Subject	Position
P5	Leader. Overall in charge of the group.
P18	XO. Second in charge.
P19	S3. Operations Officer. Responsible for the calendar and notifying the group of upcoming events.
P15	Transportation Officer. Coordinates transportation for weekly trips to the main Columbia campus. Also was responsible for planning a promotion party during the last weekend in March (Survey #2 and E-mail timeframe).
P24	Communication Officer. Responsible for helping subjects with computer problems.
P12	Social Officer. Planned and notified subjects of group social events.
P11	Assistant S3. Assists the S3, Operations Officer.

When the list of key leaders is compared to the rankings of social network measures in Tables 6-9, it can be seen that the list from the e-mail network consistently contains more key leaders than the self-reported networks. The only key leader not reported in any of the e-mail networks is P11, the Assistant Operations Officer. Upon closer examination, it was discovered that P19 and P11, having computer problems, had their computers re-imaged; thus data was not collected on either of their sent e-mail traffic. This would clearly bias the e-mail social network. It is interesting, however, that P19 still appears in the rankings of Eigenvector Centrality and Total Degree. This would only be caused by a sufficiently large number of e-mails being sent to that individual by others in the group. This suggests that e-mail communication might be a more accurate representation of the true social network for communication than self-reported surveys.

Conclusions

Social networks constructed from monitoring e-mail traffic were different than self-reported social networks constructed from surveys. It is unclear if there is a true difference or if the difference results from a bias caused by network density. The density of a self-reported network is affected by the design of the survey. The surveys used in this project were fixed choice, meaning that each subject was asked to name three and only three other subjects, whom they communicated with the most. If a subject communicated with five others, he or she could not include the other two. The e-mail traffic, on the other hand was open choice. This difference in design can create significant differences in density.

This problem can be corrected in future projects. It is recommended that investigators take a complete week of e-mail data and build a social network. They can then calculate the density of that network. Using the density of the e-mail network, they should design the fixed choice survey to target the e-mail network density. For example, if the surveys asked the subjects to list the four people they communicated with most instead of the three, the expected density of the fixed choice social network would be 0.1897. This is much closer to the density of the e-mail network of 0.1793 than the expected density of a three question network, 0.1423. The carefully designed survey should then be given to the subjects the very next day when they are asked with whom they communicated with the most in the previous week. Another option would be to have

the subjects rank-order those with whom they communicated and truncate the list to create a network that most closely approximates the density of the e-mail network. By creating social networks that are closer in density, they can be more easily compared for differences and similarities.

Social networks created from e-mail traffic were better at identifying key leaders in a group than the self reported social networks. For this group, the e-mail traffic was better at creating a social network that represented the actual group leadership structure. However, one major shortcoming resulted from the loss of data from two formal leaders. Several potential improvements to the data collection were suggested earlier in this report and may further improve the performance of monitoring e-mail traffic. Future projects may better show the value of collecting social network data from e-mail.

This investigation does show that network data collected from e-mail is effective at identifying key individuals in the organization. Collecting e-mail data is much less obtrusive and disruptive than surveys. Therefore, organizations can use methods described in this report to monitor their e-mail networks. These methods can identify key communicators in the organization, as well as when communication throughout the organization has increased or decreased and among what sub-elements these events occur. The full extent of benefits to command and control has yet to be determined. These findings suggest that continued research on e-mail networks is promising for enhanced command and control techniques in this new technical age of our Armed Forces.

References

Bernard, H. R., & Killworth, P. D. (1977). Informant accuracy in social network data II. *Human Communication Research, 4*, 3-18.

Bernard, H. R., Killworth, P. D., & Sailer, L. (1980). Informant accuracy in social network data IV: A comparison of clique-level structure in behavioral and cognitive network data. *Social Networks, 2*, 191-218.

Bernard, H. R., Killworth, P. D., & Sailer, L. (1982). Informant accuracy in social network data V: An experimental attempt to predict actual communication from recall data. *Social Science Research, 11*, 30-66.

Ellis, C.A., Gibbs, S.J., and Rein, G.L. (1991). Groupware: Some Issues and Experiences. *Communications of the AMC,* 34(1), 39-58.

Fiske, S. T., and Taylor, S. E. (1991) *Social Cognition*, 4-5, New York: McGraw-Hill, Inc.

Freeman, L. C. (1979) Centrality in social networks: I. Conceptual clarification. *Social Networks, 1*, 215-239.

Gutwin, C., and Greenberg, S. (2004) The Importance of Awareness for Team Cognition in Distributed Collaboration. In E. Salas and S.M. Fiore (Editors) *Team Cognition: Understanding the Factors that Drive Process and Performance*, 177-201, Washington: APA Press.

Hamming (1950) ''Error Detecting and Error Correcting Codes,'' Bell System Technical Journal, 29, 147-160.

Kashy, D. A. & Kenny, D. A. (1990). Do you know whom you were with a week ago Friday? A re-analysis of the Bernard, Killworth, and Sailer studies. *Social Psychology Quarterly, 53*, 55-61.

Killworth, P. D. & Bernard, H. R. (1976). Informant accuracy in social network data. *Human Organization, 35*, 269-96.

Killworth, P. D. & Bernard, H. R. (1979). Informant accuracy in social network data III: A comparison of triadic structure in behavioral and cognitive data. *Social Networks, 2*, 19-46.

Krackhart, D. (1990) *Assessing the Political Landscape: Structure, Cognition, and Power in Organizations*. Administrative Science Quarterly, Vol. 35, 342-369.

ORA: Organizational Risk Analyzer v.1.7.8. (2007). [Network Analysis Software] Pittsburgh: Carnegie Mellon University.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.

APPENDIX A: VISUAL BASIC CODE FOR THE OUTLOOK PATCH

```
'*************************************************************
' SOCIAL NETWORK EMAIL BOT
'
' V1.0
'
' To set paramters - see sendReport and Application_ItemSend
'
'*************************************************************
Sub sendReport(startDate As Date)

  Dim reportTo As String
  Dim stringOut As String

  '*************************************************************
  'PARAMETER - Set who you want the report to go to below
  '*************************************************************
  'Report recipient
  reportTo = "ai6873@usma.edu"

  'This macro checks the Outlook Sent Items folder for messages
  'On Error GoTo SendReport_err

  Dim ns As NameSpace
  Dim sentBox As MAPIFolder
  Dim tempStore As MAPIFolder

  Dim Item As Object
  Dim objCopy As Object
  Dim i As Integer
  Dim subjectID As String

  Set ns = GetNamespace("MAPI")
  Set sentBox = ns.GetDefaultFolder(olFolderSentMail)
  subjectID = Replace(ns.CurrentUser, ",", " ")

  'Iterate through sent items, tagging everything since last report
  If sentBox.Items.Count = 0 Then
    Exit Sub
  End If

  'Make a file to hold the report
  Dim FS
  Set FS = CreateObject("Scripting.FileSystemObject")

  shortDate = Replace(startDate, "/", "-")
  shortDate = Replace(shortDate, " ", "_")
  shortDate = Replace(shortDate, ":", "_")

  Dim OutStream
  Dim outFileName As String
  Dim myBody As String
```

```
outFileName = "Social_Network_Report_[" & subjectID & "]_[" & shortDate & "].csv"

myBody = "Report covers all new arrivals since " & startDate
'MsgBox "startDate = " & startDate

Set OutStream = FS.OpenTextFile(outFileName, 2, True)

'MsgBox "Checking for messages newer than " & startDate
'Write Header
OutStream.Write "Subject,Body,From: (Name),From: (Address),From: (Type)," & _
        "To: (Name),To: (Address),To: (Type)," & _
        "CC: (Name),CC: (Address),CC: (Type)," & _
        "BCC: (Name),BCC: (Address)" & vbCr

'Add each message since the startDate to the report
'Remove attachments
'Substitute date send for body
For Each Item In sentBox.Items

  If Item.SentOn >= startDate Then
    'MsgBox "got one"
    Dim goodSubject As String
    goodSubject = Replace(Item.Subject, ",", " ")
    stringOut = addToString(goodSubject, "SENT " & Item.SentOn)
    Dim goodName As String
    goodName = Replace(Item.SenderName, ",", " ")
    stringOut = addToString(stringOut, goodName)
    'MsgBox "added " & goodName
    stringOut = addToString(stringOut, Item.SenderEmailAddress)
    stringOut = addToString(stringOut, Item.SenderEmailType)
    Dim goodTo As String
    goodTo = Replace(Item.To, ",", " ")
    stringOut = addToString(stringOut, goodTo)
    stringOut = addToString(stringOut, "")
    stringOut = addToString(stringOut, "")
    Dim goodCC As String
    goodCC = Replace(Item.CC, ",", " ")
    stringOut = addToString(stringOut, goodCC)
    stringOut = addToString(stringOut, "")
    stringOut = addToString(stringOut, "")
    Dim goodBCC As String
    goodBCC = Replace(Item.BCC, ",", " ")
    stringOut = addToString(stringOut, goodBCC)
    stringOut = addToString(stringOut, "")
    stringOut = addToString(stringOut, Chr(10))

    OutStream.Write stringOut
    i = i + 1
  End If
Next Item

If i > 0 Then
  'Send the report
  success = FnSendMailSafe(reportTo, outFileName, myBody, outFileName)

End If
```

```
'Clear memory
SendReport_exit:
   Set Atmt = Nothing
   Set Item = Nothing
   Set ns = Nothing
   Exit Sub
'Handle errors
SendReport_err:
   MsgBox "An unexpected error has occurred." _
      & vbCrLf & "Please note and report the following information." _
      & vbCrLf & "Code Module:  OutLookSession.sendReport()" _
      & vbCrLf & "Error Number: " & Err.Number _
      & vbCrLf & "Error Description: " & Err.Description _
      , vbCritical, "Error!"
   Resume SendReport_exit
End Sub

Function addToString(oldString As String, addedString As String) As String
   addToString = oldString & "," & addedString
End Function
'FnSendMailSafe
'--------------
'Simply sends an e-mail using Outlook/Simple MAPI.
'Calling this function by Automation will prevent the warnings
''A program is trying to send a mesage on your behalf...'
'Also features optional HTML message body and attachments by file path.
'
'The To/CC/BCC/Attachments function parameters can contain multiple items by seperating
'them by a semicolon. (e.g. for the strTo parameter, 'test@test.com; test2@test.com' is
'acceptable for sending to multiple recipients.
'
'Read more here:
'http://www.everythingaccess.com/tutorials.asp?ID=Outlook-Send-E-mail-without-Security-Warning
'
Public Function FnSendMailSafe(strTo As String, _
                     strSubject As String, _
                     strMessageBody As String, _
                     strAttachments As String) As Boolean

'(c) 2005 Wayne Phillips - Written 07/05/2005
'http://www.everythingaccess.com
'
'You are free to use this code within your application(s)
'as long as the copyright notice and this message remains intact.

On Error GoTo ErrorHandler:

   Dim MAPISession As Outlook.NameSpace
   Dim MAPIFolder As Outlook.MAPIFolder
   Dim MAPIMailItem As Outlook.MailItem
   Dim oRecipient As Outlook.Recipient

   Dim TempArray() As String
   Dim varArrayItem As Variant
```

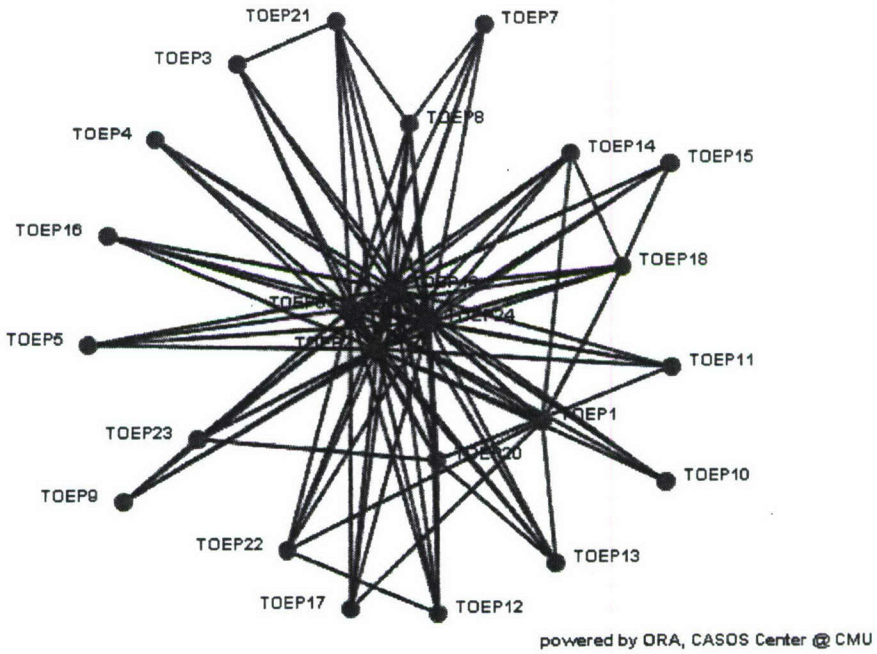
```
Dim blnSuccessful As Boolean

'Get the MAPI NameSpace object
Set MAPISession = Application.Session

If Not MAPISession Is Nothing Then

 'Logon to the MAPI session
 MAPISession.Logon , , True, False

 'Create a pointer to the Outbox folder
 Set MAPIFolder = MAPISession.GetDefaultFolder(olFolderOutbox)
 If Not MAPIFolder Is Nothing Then

   'Create a new mail item in the "Outbox" folder
   Set MAPIMailItem = MAPIFolder.Items.Add(olMailItem)
   If Not MAPIMailItem Is Nothing Then

    With MAPIMailItem

      'Create the recipients TO
        TempArray = Split(strTo, ";")
        For Each varArrayItem In TempArray
          'nextGuy = Chr(34) & CStr(Trim(varArrayItem)) & Chr(34)
          'MsgBox "next guy = " & nextGuy
          'Set oRecipient = .Recipients.Add(nextGuy)
          Set oRecipient = .Recipients.Add(CStr(Trim(varArrayItem)))

          oRecipient.Type = olTo
          Set oRecipient = Nothing

        Next varArrayItem


      'Set the message SUBJECT
        .Subject = strSubject

      'Set the message BODY (HTML or plain text)
        If StrComp(Left(strMessageBody, 6), "<HTML>", vbTextCompare) = 0 Then
          .HTMLBody = strMessageBody
        Else
          .Body = strMessageBody

        End If

      'Add any specified attachments
        TempArray = Split(strAttachments, ";")
        For Each varArrayItem In TempArray

          .Attachments.Add CStr(Trim(varArrayItem))

        Next varArrayItem

      .Send 'No return value since the message will remain in the outbox if it fails to send

      Set MAPIMailItem = Nothing
```

```
        End With

        End If

        Set MAPIFolder = Nothing

      End If

      MAPISession.Logoff

    End If

    'If we got to here, then we shall assume everything went ok.
    blnSuccessful = True

ExitRoutine:
    Set MAPISession = Nothing
    FnSendMailSafe = blnSuccessful

    Exit Function

ErrorHandler:
    MsgBox "An error has occured in the user defined Outlook VBA function FnSendMailSafe()" & vbCrLf &
vbCrLf & _
          "Error Number: " & CStr(Err.Number) & vbCrLf & _
          "Error Description: " & Err.Description, vbApplicationModal + vbCritical
    Resume ExitRoutine

End Function

Sub WriteLastReportTime(lastTime)
    Dim stringOut

    stringOut = "[LAST_SOCIAL_NETWORK_REPORT_TIME]=" & lastTime

    Dim FS
    Set FS = CreateObject("Scripting.FileSystemObject")

    Dim OutStream
    Set OutStream = FS.OpenTextFile("snReport.ini", 2, True)
    OutStream.Write stringOut

End Sub

Function ReadLastReportTime()

    Dim FS
    Set FS = CreateObject("Scripting.FileSystemObject")

    Dim OutStream

    Dim fileContents
    fileContents = "[LAST_SOCIAL_NETWORK_REPORT_TIME]=7/7/1994 11:00:00"

    On Error Resume Next
```

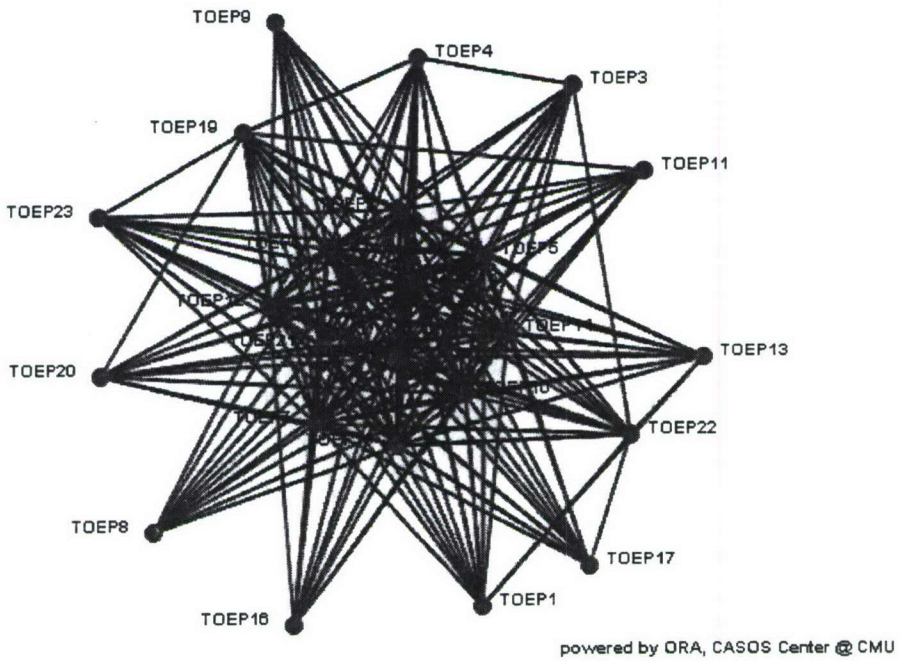
```
    fileContents = FS.OpenTextFile("snReport.ini").ReadAll

    startPos = InStr(1, fileContents, "=", 1)

    Dim lastTime

    lastTime = Mid(fileContents, startPos + 1)

    ReadLastReportTime = lastTime

End Function

Private Sub Application_ItemSend(ByVal Item As Object, Cancel As Boolean)

    'Interval between reports (minutes)
    '*************************************************************
    'PARAMETER - Set the time (min) between reports below
    '*************************************************************
   Dim reportInterval
   reportInterval = 1000

   'See how long its been since the last report was created
   Dim currentDateTime As Date

    currentDateTime = Now

    Dim lastReportTime As Date
    lastReportTime = ReadLastReportTime
    Dim elapsedTime

    elapsedTime = DateDiff("n", lastReportTime, currentDateTime)
    'MsgBox "Elapsed Time = " & elapsedTime
    If elapsedTime > reportInterval Then

        'Build Report Here
        sendReport (lastReportTime)

        'Record current time as last report time
        WriteLastReportTime (currentDateTime)
    Else
        'Do nothing
    End If

End Sub
```

APPENDIX B: SOCIAL NETWORK VISUALIZATIONS FROM THE SIX WEEKS OF E-MAIL TRAFFIC FEATURED IN CHAPTER 2.

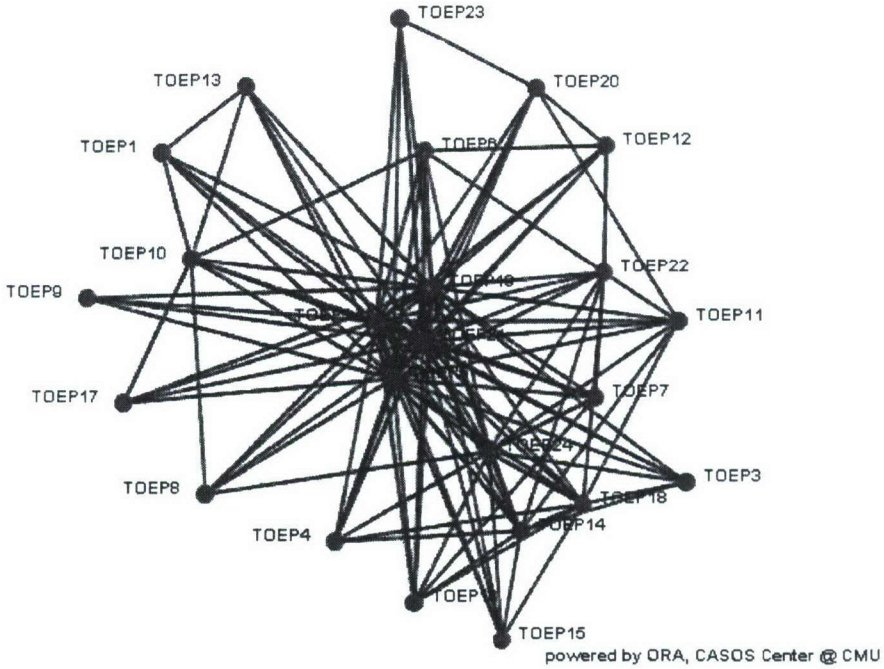
Academic network of 29th October 2006



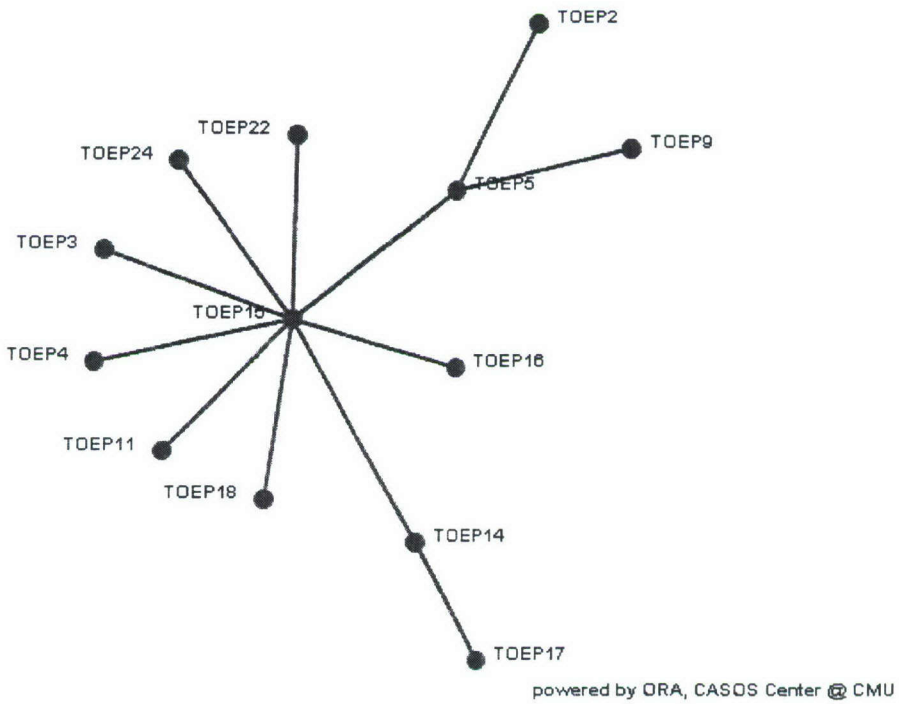
Academic network of 5th November 2006



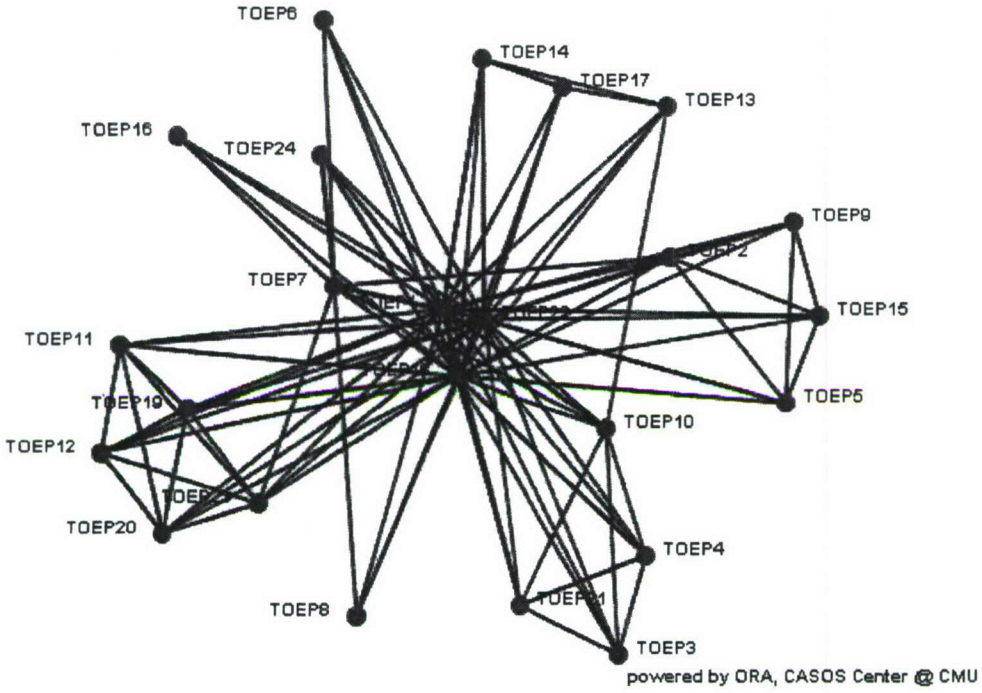
Academic network of 12th November 2006



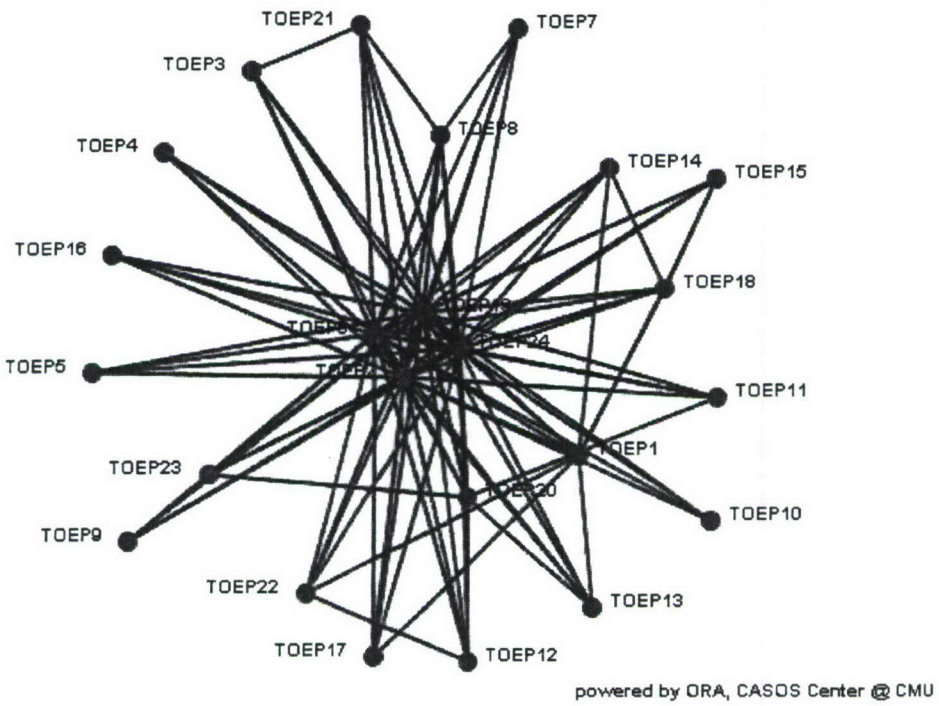
Academic network of 19th November 2006



Academic network of 26th November 2006



Academic network of 3rd December 2006



APPENDIX C: LIST OF ACADEMIC DEMANDS FOR THE WEEKS OF 29 0CT – 3 DEC 2006

Week	Academic Requirements	Academic Demands
Oct. 29, 2006	Understanding behavioral research Preparation for coaching Cross culture leadership	Quizzes, Group Work/Homework, Surveys Group Work Group Work
Nov. 05, 2006	Human research management exam Prepare for coaching and research Group dynamics Army-Air Force football game	Group Studying, Project Quizzes, Group Work, Journals Group Work Tailgates
Nov. 12, 2006	Group dynamics paper Prepare for coaching and research Cross culture leadership Group dynamics conference	Group Work Group Work
Nov. 19, 2006	Prepare for coaching and research Thanksgiving leave	
Nov. 26, 2006	Group dynamics paper is due Prepare for coaching and research	Group Work, Presentation
Dec. 03, 2006	Human research management Understanding behavioral research Prepare for coaching and research Army-Navy football game	