UNOBTRUSIVE SOCIAL NETWORK DATA FROM EMAIL

MAJ Ian McCulloh,* MAJ Benjamin Ring
Network Science Center
United States Military Academy
West Point, New York 10996

Terrill L. Frantz, Professor Kathleen M. Carley
Center for Computational Analysis of Social and Organizational Systems
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

ABSTRACT

Email provides a rich source of longitudinal social network data that can be used for applications ranging from command and control, to military intelligence, to basic social science research. This project reviews several methods available to extract email network data and compares them in terms of data quality and convenience of collection. In general, it is preferable to obtain email data directly from the central SMTP email server. In situations where this is not possible, alternative approaches presented here can be useful. These techniques for analyzing email data have been automated in the Organizational Risk Analyzer (ORA) software, which is freely available to DoD and academia.

1. INTRODUCTION

Email has significantly changed how people communicate and interact. In many ways communication is easier and more reliable with email, however, there are many new challenges introduced. Over the past decade, many people have turned to email as the primary means to send information and to communicate (Ducheneaut, & Bellotti, 2001). It has enabled groups to work together, socialize and collaborate across any distances and outside of structured organizational boundaries. When organizational relationships do exist, email traffic among that group often mirrors this structure (Diesner, Frantz & Carley, 2005; Frantz & Carley, 2008; Tyler, Wilkinson, & Huberman, 2003). As a result, studying and analyzing communication patterns of email traffic can provide much insight into not only how an organization is structured, but also into how it actually operates (Carvalho, & Cohen, 2007). For example, a supervisor may typically send email to all his immediate subordinates and, likewise, those subordinates will respond. An increase in peer to peer collaboration may indicate that problems are being solved at a much lower level. Individual agents that connect disconnected groups might represent organizational vulnerabilities. Identifying these patterns from collected email data is extremely useful in identifying the underlying social network behavior of an organization.

We present two general methods for gathering and analyzing email data along with an analysis of each of these methods. During the course of this study, we gathered client-side email data over a seven month period to reveal the social network of a group of 24 mid-career Army officers. We also employed a centralized data collection procedure over a five month period directly from the central Simple Mail Transfer Protocol (SMTP) email server. The data collection schemes are compared in terms of data quality, ease of collection, and subject cooperation.

These email collection methods have been automated in a feature called CEMAP II contained in ORA (Carley, et al., 2008) --- a software package from the Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University (Frantz & Carley, 2008b). The ORA software program is freely available to people in the DoD and at academic institutions at www.casos.cs.cmu.edu.

2. BACKGROUND

Gathering email related data has shown to identify actual social and communal patterns among the email users (McCulloh et al, 2007; 2008). A collaborative group at Hewlett Packard Labs demonstrated that simply gathering the "TO" and "FROM" fields from a large collection of email messages can produce community structure when applied to a graph representation (Tyler, Wilkinson, & Huberman, 2008). This study focused on email data only at the organization's central mail server. In contrast, *Themail*, a visualization which shows an individual user's email exchange presents a visual network analysis of a user's email content simply by analyzing the archived mail on his or her personal computer (Viégas, Golder, & Donath, 2006). Users in this study were required to manually upload their entire Microsoft Outlook archive folder for analysis. Similar to this technique, Gloor and Zhao (2004) developed a software tool, TeCFlow, which gathers email data from a user's computer contained in various mailboxes and

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
DEC 2008	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Unobtrusive Social Network Data From Email	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Network Science Center United States Military Academy West Point, New York 10996**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002187. Proceedings of the Army Science Conference (26th) Held in Orlando, Florida on 1-4 December 2008

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**6**	

outlook archived files and stores that data into an SQL-database.

Communication via email can be divided into two types of relationships: the human-computer interaction; and the computer-computer interaction. People are usually most familiar with the human-computer interaction, where they sit at a computer, write an email, and push "send"; or they login to their email account and read messages contained in the "inbox". The computer-computer interaction, is actually an automated exchange between two computers, often with several other computers serving as intermediaries in the delivery process. A message sent from one computer is received by the target computer(s), in its electronic form, via a client software program that ultimately copies the email message from its host server. The email message is stored on a designated central server until the receiver "picks up" the message from the server. This process is the electronic version of picking up a package at the post office. The electronic email can be delivered to the post office repository for you to physically pick up, or directly to your personal mail box for you to pick up. Once a target computer picks up the message, the human-computer interaction allows the human to read, print, or store the electronic message via their email client software.

There are several different ways in which email can be delivered through the computer-computer interaction in the world-wide electronic email architecture. The message can be delivered to the equivalent of post office lobby-box, called an IMAP server. The email can be delivered to a personal mailbox, called a Post Office Protocol (POP) server. The email can also be routed through a Microsoft Exchange (MSEx) server. There are many technical differences between these email servers, but their purpose is the same. However, the principal difference between an IMAP and a POP email server is the storage feature of the server. An IMAP server will allow you (or your email client software) to persist, or store, your email physically on that server. A POP server only serves as a temporary holding station for a message that is removed once it has been retrieved by your email client software. An MSEx server is a Microsoft proprietary system that is widely used throughout the DoD. While it has some additional security features, it is more difficult to extract email network data from this system because of the propriety data format that Microsoft institutes. An IMAP server is designed to store the message even after the email has been initially retrieved. It should be noted that the POP protocol calls for an email to be removed from the incoming mail box once it has been retrieved, however some software extensions do allow for a read-only access to the POP inbox, resulting in the message remaining in the inbox when retrieved and is therefore managed by the client software level. The popular Yahoo mail service implements this feature for paying customers.

Once a target computer receives an email, the human-computer interaction involves the computer displaying the message using client software. Email messages at the computer-computer interaction level are most often formatted in a world-wide standard format called MBOX. MBOX allows for different email client software programs to access the email from the server without confusion. The MBOX format specifies two sections of the email, the header section and the body section. The header section includes the From:, To:, CC:, BCC:, Subject:, and Date:, information. The body section contains the message text and any attachments to the email.

The MSEx server does not store messages in the MBOX format. Microsoft's proprietary standards create technical and licensing hurdles in accessing email data directly from the server in any manner other than using Microsoft software. Unfortunately, the MSEx format is widely used throughout DoD, making email data extraction more difficult. There are three approaches that we have discovered for extracting email content from an MSEx format. One approach is a custom client-side visual basic patch (McCulloh, et. al., 2007). Another client-side approach involves using .NetMap, which is a plug-in for Microsoft Excel 2007 that extracts email data from a proprietary *.pst file into an Excel format. The data can then be manipulated or saved to other file formats. The third approach involves parsing header data from a server log file. These approaches will be discussed in more detail in this paper. Analysis of dyad counts will be used to compare the performance of a client-side data collection with a centralized data collection.

3. METHOD

This study involves monitoring the email traffic of 24 mid-career Army officers in a one-year graduate program administered jointly by Columbia University and the U.S. Military Academy (USMA). Each of the officers participating were asked to sign a consent form in accordance with the institutional review board (IRB), approved by the USMA Human Subjects Research Review Board allowing their data to be collected for research purposes.

As part of this study, the participants permitted us to place a custom developed program (McCulloh et al, 2007) that works in conjunction with their MSEx Outlook email accounts. This program allowed us to collect email data from the sent items folder found on participants' personal computers. The information included all of the header information associated with an email. We did not view or

include the body of the email in the study. We were also able to collect similar email header information directly from the log files maintained by the Directorate Of Information Management (DOIM). The data collected from the custom program is referred to as the *Client-Side Method*, while data collected from the DOIM log files is referred to as the *Centralized Method.* We did not investigate .NetMap as an approach as it has identical underlying email data-sourcing capabilites and functionality only with a different, al beit a more elegant, user interface. The email data collected from all methods was analyzed using a dynamic network analysis approach (Carley, 2003).

3.1 Client-Side Method

A client side Visual Basic for Applications (VBA) program was installed on the personal computers (PC) of all participants, in the session window of their Microsoft Outlook. Details of this data collection scheme to include the visual basic code are outlined in detail by McCulloh, et al. (2007). It is designed to overcome the difficulty in pulling information from a subject's sent mail folder in a proprietary Outlook Exchange system. This patch is easy to implement in Visual Basic and works harmoniously with Microsoft Outlook. The principal investigator could then compile the data from all participants into one master file and ensure anonymity of the names.

One of the chief advantages in managing a client-side patch is the low-level control in gathering data. A researcher does not have to obtain permissions from a network administrator to collect email data. They merely need the consent of the monitored individuals, who must login to their Outlook for the client-side patch to be installed. Furthermore, the program designers can pick and choose which data to import from the local client. If, for example, we wanted to include message content, then that could have been an option. We could have also just as easily gathered incoming email traffic, as opposed to only monitoring outgoing mail. This could provide further insight into areas such as whether a user classifies email as junk mail, whether they delete an incoming message, or even if they flag a particular message as important.

Managing the data collection from the individual participant required minimal effort. Once fully developed and installed, the Visual Basic patch is little to no overhead on the part of the user to manage. Furthermore, these participants felt more comfortable knowing that they have some degree of control in how the data is collected. While this could impede the data collection process, the subjects felt more comfortable knowing what was actually monitoring their email. Initially, most of the participants' email were sent to other students or people affiliated with their graduate program. Within two to three weeks, the participants began to email family members and friends. We suppose that this represents an increased level of trust. In the beginning, participants felt that their email needed to appear strictly business related. Gradually, as they incrementally sent personal email messages while they were "at work" without any negative consequences, they began to feel comfortable and appear to have returned into a normal cadence of email communication. Most of the participants knew how to remove the patch when their participation in the project ended. Several participants said they felt more comfortable knowing that the software sending the principal investigator information was on their computer, and that "Big Brother" was not pulling their information from somewhere else.

3.2 Centralized Method

As an alternative method, we developed a software application which analyzes email data gathered directly from a centralized email exchange server. This software gathered data over a five month period and extracted those email messages which were sent and received from the participants in this study. The server log files contain the email header information. This information was parsed into the same format as the client-side method.

With this method of data collection, the participants were not aware of the precise time that the collection process started. They did provide consent in accordance with the IRB, however, we were not required to inform them of the exact date when collection would begin. There was no significant observable change in the participants' pattern of communication. The centralized method was completely unobtrusive.

3.3 Dyad Analysis

It was not clear at the beginning of this investigation whether email 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. To investigate the structure of the network, we computed the dyad count. The dyad count, defined as the communication between two nodes (Wasserman, & Faust, 1994) distinguishes three different types of communication: asymmetric, mutual, and null. In an asymmetric dyad, one node talks to another, but does not receive a response. This type of communication could be an example of a group that has members who are sending out information. A mutual dyad signifies two nodes communicating with each other. 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 occurs when two nodes which are part of the network do not have any communication activity. In a dyad count, we conduct a

census and tabulate the number of null, mutual, and asymmetric dyads. With 24 members in our study comprising a network, there exists 276 combinations of possible undirected pairs. Each of the 276 dyads could be either null, mutual, or asymmetric. The dyad counts are compared for data collected with the client-side method, centralized method, and with a calendar of significant events.

4. RESULTS

There were significant differences in the client-side and centralized methods of data collection. The data from both methods was coded as a meta-network (Carley, 2002). Considering that the participants are a random sample of mid-career Army officers that all fulfill the same role of student in the organization, we might hypothesize that the email relationships formed in the network are random. Given that there are 24 nodes in the network, there exist 24 x 23 = 552 possible dyads. We can test the hypothesis:

$$H_0: \text{Graph} \sim \text{Binomial}(552, 0.5)$$
$$H_A: \text{Graph} \neq \text{Binomial}(552, 0.5),$$

using the test statistic $z = (l - E(l)) / \text{Sqrt}(V(l))$, where l is the number of directed links in the graph. This reduces to $z = (l - 276) / 11.75$, where l is the sum of the mutual and asymmetric dyad counts. Under the null hypothesis, this number follows a standard normal distribution. The p-value was significant at the 0.05 level for most weeks, providing evidence to reject the hypothesis that email communication patterns are random binomial with a probability parameter of 0.5. A week with a corresponding p-value that was not significant at the 0.05 level can be identified in Table 1 by the 95% confidence interval on the Binomial parameter p that includes 0.5.

A confidence interval on the probability of communication can be constructed for each week according to the expression given by,

$$\hat{P} \pm z_{\alpha/2}\sqrt{\hat{P}(1-\hat{P})/552}$$

where $\hat{P}$ is the maximum likelihood estimate of the unknown parameter p in the assumed binomial distribution and equal to l / 552. Table 1 shows the mutual, asymmetric, and null dyad counts recorded using the client-side and centralized methods. The right most column of Table 1 shows the 95% confidence interval on the random probability of communication. A confidence interval that spans 0.5 will correspond to a significant p-value in the random binomial hypothesis test above. For each week in Table 1, two values are shown for each of the dyad counts: Mutual, Asymmetric, and Null. The numbers in the top of each cell in Table 1 correspond to the client-side data collection method. The numbers in the bottom of each cell in Table 1 correspond to the centralized data collection method. The data presented in Table 1 corresponds to the time period beginning with the first week of the Spring semester and ending with the week before Spring break. The students took their comprehensive exam following Spring break and then began to transition to their military duties at West Point. Therefore, this data represents a reasonable time period for comparison of the client-side and centralized methods of data collection.

Table 1. Recorded directed links using client-side and central methods.

Week	Mutual	Asymmetric	Null	Confidence
13 Jan 2008	0	44	232	(0.06,0.10)
	54	89	133	(0.22,0.30)
20 Jan 2008	6	88	182	(0.14,0.20)
	218	83	0	(0.50,0.59)
27 Jan 2008	0	78	198	(0.11,0.17)
	118	92	66	(0.34,0.42)
3 Feb 2008	8	162	106	(0.27,0.35)
	202	81	0	(0.47,0.55)
10 Feb 2008	0	148	128	(0.23,0.31)
	112	100	64	(0.34,0.42)
17 Feb 2008	6	114	156	(0.18,0.25)
	230	79	0	(0.52,0.60)
24 Feb 2008	26	108	142	(0.21,0.28)
	204	92	0	(0.49,0.58)
2 Mar 2008	84	192	0	(0.46,0.54)
	320	51	0	(0.58,0.66)
9 Mar 2008	26	143	107	(0.27,0.34)
	204	73	0	(0.46,0.54)

* Client-side dyad counts are above central dyad counts.

It can be seen in the Confidence column of Table 1 that there is a statistically significant difference in the probability of communication between the client-side and central data collection methods for all weeks, by observing that the 95% confidence intervals do not overlap. In all cases, the client-side method underestimates the probability of communication in the network. The general pattern of the probability parameter is correlated at a value of 0.69, which is low considering they are estimates on the same group of individuals during the same week. The client-side data collection method is therefore biased.

The dyad count analysis can provide additional insight into the organizational dynamics of the participants by comparing their probability of interaction to significant events on their academic calendar. We restrict our investigation to data collected using the centralized method since it is complete. The centralized method captures all data sent or received through the central server. The maximum likelihood estimate of the parameter, p, in the binomial distribution of dyads is plotted over time and displayed in Figure 1.

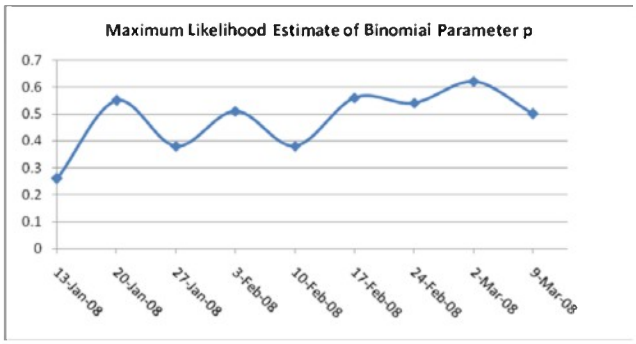

Figure 1. MLE of parameter *p* using centralized method.

The lowest MLE of *p* is shown in the first week of the semester, when the participants were just returning from Christmas leave. This was followed by an increase in communication as the group begins to plan for group academic assignments, carpooling, and other administrative issues. The low points in the MLE of *p* occur during the weeks of 27 January and 10 February when major group academic projects or presentations were due. This is consistent with the findings of McCulloh, et. al. (2007) who observed a similar decrease in email communication during times of group activity. They hypothesized that during times of increased face-to-face communication, people communicate verbally and have less time and need for email communication. Furthermore, during these times of increased subgroup activity, people have less time to write and respond to emails from individuals outside of their immediate subgroup. Following the group assignments due during the week of 10 February, the next major academic event was the comprehensive exam following Spring break.

A similar dyad analysis for the client-side method is shown in Figure 2. The characteristic dip in email communication corresponding to group activity is not clear. A careful review of the participants' academic calendar does not reveal any activities or events that would explain the behavior of the plot in Figure 2. This further suggests the importance of centralized email data collection.

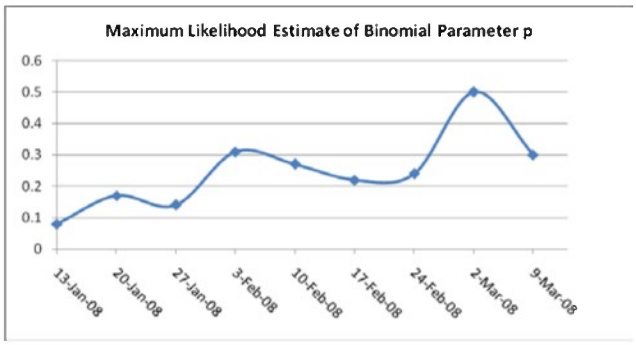

Figure 2. MLE of parameter *p* using client-side method.

The client-side method of data collection is not completely without merit. It can still be seen in Figure 2 that the first week has the lowest MLE of *p*. There is also a dip in the plot for the group assignment for the week of 27 January. The identification of the week of 10 February is missed however. This suggests that even client-side data can provide some insight into group behavior. This may be an appropriate method to use when complete centralized data is unavailable. Centralized data may be unavailable for reasons of security, privacy, damage, or other technical difficulties. In these situations, the client-side method may still provide valuable information on social network behavior.

5. CONCLUSION

We found that the primary advantage to utilizing a server-side method to gather data is the improved data integrity. Every user with an email account must both send and receive data from that account's associated mail server. Therefore, to ensure that all data is gathered it must be collected at its source. All data contained within the centralized server is available for collection, such as from, to, cc, bcc, subject, time of receipt at the mail server, etc. Copying data directly from the server allows the social network analyst to accurately study all email communications within a study group for those utilizing their given email address.

Implementing a server based application also precludes the subjects involved in the study from corrupting and inserting bias into the data. With a client-side application, users had the ability to turn off, remove or disrupt the execution of the program used to monitor email. With a server-side collection technique, the clients are completely unaware or knowledgeable about when or what is collected. We found that while it takes more overhead to initiate the retrieval of email traffic from a mail server, there is surprisingly little overhead on the part of a server administrator to actually assist the research effort in gathering data. Since log files are typically stored in a common location on the server, the administrator need only make these files available. When operated across a network, he/she can easily copy these log files to a common location from where the server-based data collection program can import the data.

By presenting two methods for gathering and analyzing email data, we have shown both advantages and disadvantages for the social network analyst. These strengths and limitations must be considered by any social network analyst when studying email traffic. Even though gathering data at its source does provide better data integrity, such data collection means are not always feasible. In these cases, email data collected in a decentralized manner can still provide insightful analysis of the underlying social network.

We advise a practitioner to be highly sensitive to the privacy implications of this process, especially in the public and private sectors. People within the military typically do not maintain the expectation of email and internet privacy. This may not be true in other populations. Care must also be exercised with interpreting the results of these types of social networks. It is important that trained social network analysts provide proper interpretation of the organizational behavior, while respecting the privacy of individual identities. Revealing the position an individual maintains in the social network of an organization may lead to an overall decrease in trust and adversely affect the leadership climate within the organization. When used properly, however, social network analysis can provide a wealth of valuable information to the organization. Several commands within the Army have already implemented social network data collection from email. These methods have been automated in the software package ORA, which is maintained by CASOS at Carnegie Mellon University and can be freely downloaded by the military and academia.

Future research in this area will likely explore the impact of cellular phone communication and blackberries on social networks within the military. This line of research will further support the efficacy of Netcentric Operations within the Army. Focused research into the usage of cell phones, blackberries, e-mail, and face-to-face communication during major group activities will provide greater insight into social network data collection. Understanding the desired channels of communication for military leaders, may significantly contribute to shaping the communication technologies that the DoD invests in. This line of research may also provide data for real-time monitoring of organizational change. It will certainly be valuable in enhancing command and control systems used by the military.

ACKNOWLEDGEMENTS

This research is part of the IkeNet project in the U.S. Military Academy Network Science Center and the Dynamics Networks project in CASOS (Center for Computational Analysis of Social and Organizational Systems, http://www.casos.cs.cmu.edu) at Carnegie Mellon University. This work was supported in part by:

- The Army Research Institute for the Behavioral and Social Sciences, Army Project No. 611102B74F/ Grant No. CMU - W91WAW07C0063
- The Army Research Labs Grant No. DAAD19-01-2-0009
- The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the National Science Foundation or the U.S. government.

REFERENCES

Carley, K.M. (2003). Dynamic Network Analysis. In *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, Ronald Breiger, Kathleen Carley, and Philippa Pattison, (Eds.) Committee on Human Factors, National Research Council*, National Research Council. Pp. 133-145.

Carley, K.M. (2002) Smart Agents and Organizations of the Future. *The Handbook of New Media*. Edited by Leah Lievrouw and Sonia Livingstone, Ch. 12, pp. 206-220, Thousand Oaks, CA, Sage.

Carley, K.M., Columbus, D., DeReno, M., Reminga, J. and Moon, I. (2008). ORA User's Guide 2008. *Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report, CMU-ISR-08-125.*

Carvalho, V.R., , Wu, W. & Cohen W. W., (2007). Discovering Leadership Roles in Email Workgroups. In *Proceedings, CEAS 2007 -- Fourth Conference on Email and Anti-Spam,* Aug 2 - 3, 2007 Mountain View, CA, www.ceas.cc.

Diesner, J., Frantz, T., Carley, K. M. (2005) Communication Networks from the Enron Email Corpus It's Always About the People. Enron is no Different. *Computational and Mathematical Organization Theory*, 11, 201 – 228.

Ducheneaut, N.; Bellotti, V. (2001). Email as habitat: An exploration of embedded personal information management. *ACM Interactions*. Sep-Oct, 30-38.

Frantz, T., Carley, K.M. (2008a) Transforming raw-email data into social-network information. In Christopher C. Yang, Hsinchun Chen, Michael Chau, Kuiyu Chang, Sheau-Dong Lang, Patrick S. Chen, Raymond Hsieh, Daniel Zeng, Fei-Yue Wang, Kathleen Carley, Wenji Mao, and Justin Zhan (Eds.). '*Intelligence and Security Informatics Workshops, PAISI, PACCF and SOCO 2008' Springer, Lecture Notes in Computer Science, No. 5075*. Pacific Asia Workshop on Intelligence and Security Informatics (PAISI 2008).

Frantz, T. & Carley, K.M. (2008b). *CEMAP II: An Architecture and Specifications to Facilitate the Importing of Real-World Data into the CASOS Software Suite.* Carnegie Mellon University Technical Report (ISR-08-130)

Gloor, P. & Zhao, Y. (2004). TeCFlow - A Temporal Communication Flow Visualizer for Social Networks Analysis. In *Proceedings, ACM CSCW Workshop on Social Networks*, Nov 6, Chicago, IL,

McCulloh, I.A., Carley, K.M. (2008). *Social Network Change Detection.* Carnegie Mellon University, Technical Report, CMU-CS-08-116.

McCulloh, I.A., Garcia, G., MacGibbon, J., Tardieu, K., Dye, H., Moores, K., Graham, J. (2007). *IkeNet: Social Network Analysis of e-mail Traffic in the Eisenhower Leadership Development Program.* Army Research Institute Technical Report 1218.

Tyler, J. R., Wilkinson, D. M., Huberman, B. A. (2003). Email as Spectroscopy: Automated Discovery of Community Structure within Organizations. *Communities and Technologies*, 81-96.

Viégas, F.B., Golder, S., & Donath, J. (2006). Visualizing email content: portraying relationships from conversational histories. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. Montréal, Québec, Canada, 979-988.