Modeling Face-to-Face Communication using the Sociometer

Tanzeem Choudhury and Alex Pentland
MIT Media Laboratory
20 Ames Street
Cambridge, MA 02139 USA
+1 617 253 0370
tanzeem@media.mit.edu

ABSTRACT
Knowledge of how people interact is important in many disciplines, e.g. organizational behavior, social network analysis, information diffusion and knowledge management applications. We are developing methods to automatically and unobtrusively learn the social network structures that arise within human groups based on wearable sensors. At present researchers mainly have to rely on questionnaires, surveys or diaries in order to obtain data on physical interactions between people. In this paper, we show how sensor measurements from the sociometer can be used to build computational models of group interactions. We present results on how we can learn the structure of face-to-face interactions within groups, detect when members are in face-to-face proximity and also when they are having conversations.

Keywords
Organizational behavior, social network analysis, expertise networks, wearable computing, Bayesian networks.

INTRODUCTION
In almost any social and work situation our decision-making is influenced by the actions of others around us. Who are the people we talk to? How often do we talk to them and how long do the conversations last? How actively do we participate in those conversations? Answers to these questions have been used to understand the success and effectiveness of a work group or an organization as a whole. Can we identify the differences between people's interactions? Can we identify the individuals who talk to a large fraction of the group or community members? Such individuals, often referred to the *connectors*, have an important role in information diffusion [1]. Thus, learning the connection structure and nature of communication among people are important in trying to understand the following phenomena: (i) diffusion of information (ii) group problem solving (iii) consensus building (iv) coalition formation etc. Although people heavily rely on email, telephone and other virtual means of communication, research shows that high complexity information is mostly exchanged through face-to-face interactions [2]. Informal networks of collaboration within organizations coexist with the formal structure of the institution and can enhance the productivity of the formal organization [3]. Furthermore, the physical structure of an institution can either hinder or encourage communication. Usually the probability that two people communicate declines rapidly with the distance between their work locations [2, 4].

We believe the best way to learn informal networks is through observations. We then need to have a mechanism to understand how individuals interact with each other from these observations. Data-driven approach can augment and complement existing manual techniques for data collection and analysis. The goal of our research is twofold – (i) build systems and sensors that can play the role of a mythical "familiar" that sits perched on a user's shoulder, seeing what he sees, with the opportunity to learn what he learns (ii) build an algorithmic pipeline that can take these sensors data and model the dynamics and interconnections between different players in the community. We hope to lay the groundwork for being able to automatically study how different groups within social or business institutions connect. This will help in understanding how information propagates between groups. The knowledge of people's communication networks can also be used in improving context-aware computing environments and coordinating collaboration between group and community members.

SENSING AND MODELING FACE-TO-FACE NETWORKS
As far as we know, there has been no previous work on modeling face-to-face interactions within a community. This absence is probably due to the difficulty in obtaining reliable measurements from real-world interactions. One has to overcome the uncertainty in sensor measurements, this is in contrast to modeling virtual communities where we can get unambiguous measurements about how people interact – the duration and frequency (available from chat and email logs) and sometime even detailed transcription of interactions [5, 6].

We believe sensing and modeling physical interactions among people is an untapped resource. In this paper, we present statistical learning methods that use wearable sensor

data to make reliable estimates about a user's interaction state (e.g. who is she talking to, how long did the conversation last, etc.). We use these results to infer the structure/connections that exists in groups of people. This can be much cheaper and more reliable than human-delivered questionnaires. Discovering face-to-face communication networks automatically will also allow researchers to gather interaction data from larger groups of people. This can potentially remove one of the current bottlenecks in the analysis of human networks: the number of people that can be surveyed using manual techniques. Sensor-based approach is free from recall failures and personal interpretation bias of surveys.

Measuring Interactions using the Sociometer

In this section we describe how we use wearable sensors to measure interactions. The first step towards reliably measuring communication is to have sensors that can capture interaction features. For example, in order to measure face-to-face interactions we need to know who is talking to whom, the frequency and duration of conversations.

We have conducted an experiment at the MIT Media lab where a group of people agreed to wear the sociometer. The sociometer is wearable sensor package that measures people's interactions. It is an adaptation of the hoarder board, a wearable data acquisition board, designed by the electronic publishing and the wearable computing group at the Media lab, for details on the hardware design please refer to [7, 8]. While designing the sociometer, we put special emphasis on the following issues: comfort of the wearer, aesthetics, and placement of the sensors. We believe these are important points when it comes to greater user acceptance and reliable sensor measurements [9]. The design of the device follows closely the wearability criterion specified in [10], which explores the interaction between the human body and a wearable and provides a guideline on shape and placement of wearables that are unobtrusive and do not interfere with the natural movement of the body.

During the data collection phase, the users had the device on them for six hours a day (11AM –5PM) while they are on the MIT campus. We performed the experiment in two stages – (i) single group stage where 8 subjects from the same research group wore the sociometer for 10 days (60 hours of data per subject) and (ii) multi-group stage where 23 subjects from 4 different research group wore the sociometer for 11 days (over two full work weeks and 66 hours of data per subject). The subjects were a representative sample of the community, including students, faculty and administrative staff.

The sociometer has an IR transceiver, a microphone, two accelerometers, on-board storage, and power supply. The wearable stores the data locally on a 256MB compact flash card and is powered by four AAA batteries. A set of four AAA batteries is enough to power the device for 24 hours. Everything is packaged into a shoulder mount so that it can be worn all day without any discomfort.

The sociometer stores the following information for each individual:

1. Information about people nearby (sampling rate 17Hz – sensor IR)
2. Speech information (8KHz - microphone)
3. Motion information (50Hz - accelerometer)

Other sensors (e.g. light sensors, GPS etc.) can also be added in the future using the extension board. For this paper we do not use the data obtained from the accelerometer.

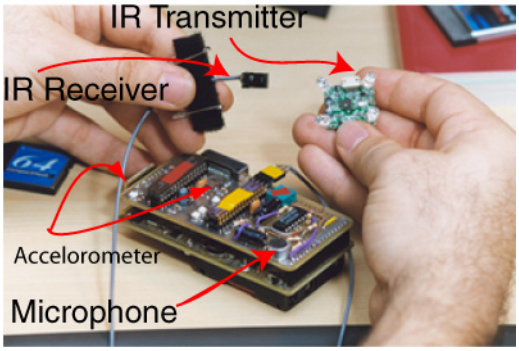

Figure 1 - The wearable sensor board

The success of IR detection depends on the line-of-sight between the transmitter-receiver pair. The sociometer has four low powered IR transmitters. The use of low powered IR transmitters is optimal because (i) we only detect people in close proximity as opposed to far apart in a room (as with high-powered IR) and (ii) we detect people who are facing us and not people all around us (as with RF transmitter). The IR transmitters in the sociometer create a cone shaped region in front of the user where other sociometers can pick up the signal. The range of detection is approximately six feet, which is adequate for picking up face-to-face communication. The design and mounting of the sociometer places the microphone six inches below the wearer's mouth, which enables us to get good audio without a headset. The shoulder mounting also prevents clothing and movement noise that one often gets from clip-on microphones. Most of the users were very satisfied with the comfortable and aesthetic design of the device. The majority made no complaints about any inconvenience or discomfort from wearing the device for six hours everyday.

Despite the comfort and convenience of wearing a sociometer, we are aware that subject's privacy is a concern for any study of human interactions. Most people are wary about how this information will be used. To protect the user's privacy we agree only to extract speech features, e.g. energy, and spectral features from the stored audio and never to process the content of the speech. But, to obtain ground truth we need to label the data somehow, i.e. where

do the conversations occur in the data and who are the participants in the conversations. Our proposed solution is to use garbled audio instead of the real audio for labeling. Garbling the audio by swapping 100ms of consecutive audio segments makes the audio content unintelligible but maintains the identity and pitch of the speaker [11]. In future versions of the sociometer we will store encrypted audio instead of the audio, which can also prevent unauthorized access to the data.

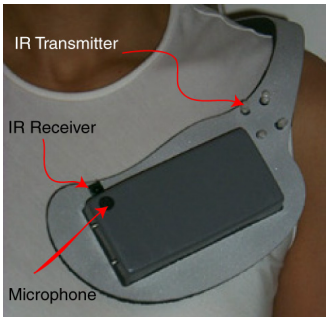

Figure 2 - The shoulder mounted sociometer

Figure 3 Subjects wearing sociometers during their daily interactions.

DATA ANALYSIS METHODS

The first step in the data analysis process is to find out when people are in close proximity. We use the data from the IR receiver to detect proximity of other IR transmitters. The receiver measurements are noisy – the transmitted ID numbers that the IR receivers pick up are not continuous and are often bursty and sporadic. The reason for this bursty signal is that people move around quite a lot when they are talking, so one person's transmitter will not always be within the range of another person's receiver. Consequently, the receiver will not receive the ID number continuously at 17Hz. Also, each receiver will sometimes receive its self-ID number. We pre-process the IR receiver data by filtering out detection of self ID number as well as propagating one IR receiver information to other nearby receivers (if receiver #1 detects the presence of tag id #2, receiver #2 should also receive tag id #1). This pre-processing ensures that we maintain consistency between different information channels. However, we still need to identify continuous chunks of time(an episode) when people are in proximity from the bursty receiver measurements. Two episodes are separated by contiguous time chunk in between where no ID is detected. A hidden Markov model (HMM) [12] is trained to learn the pattern of IR signal received over time. Typically an HMM takes noisy observation data (the IR receiver data) and learns the temporal dynamics of the underlying hidden node and its relationship to the observation data. The hidden node in our case has binary state - 1 when the IDs received come from the same episode and 0 when they are from different episodes. We hand-label the hidden states by labeling 6 hours of data. The HMM uses the observation and hidden node labels to learn its parameters. We can now use the trained HMM to assign the most likely hidden states for new observations. From the state labels we can estimate the frequency and the duration that two people are within face-to-face proximity. Figure 4 shows five days of one person's proximity information. Each shade of gray in the sub-image identifies a person to whom the wearer is in close proximity of and the width is the duration contact. Note that we are also able to detect when multiple people are in close proximity at the same time.

Figure 4 - Proximity information for person 1. Each sub-image shows one day's information. Each row within the sub-image corresponds to a different person. HMM which groups the data into contiguous time chunks.

The IR tag can provide information about when people are in close face-to-face proximity. But it provides no information about whether two people are actually having a conversation. They may just have been sitting face-to-face during a meeting. In order to identify if two people are actually having a conversation we first need to segment out the speaker from all other ambient noise and other people speaking in the environment. Because of the close placement of the microphone with respect to the speaker's mouth we can use simple energy threshold to segment the

speech from most of the other speech and ambient sounds. It is been shown that one can segment speech using voiced regions (speech regions that have pitch) alone [13]. In voiced regions energy is biased towards low-frequency range and hence we use low-energy threshold (2KHz cut off) instead of total energy. The output of the low-frequency energy threshold is passed to another HMM as observation, which segments speech regions from non-speech regions. The two states of the hidden node correspond to the speech chunks labels (1 = a speech region and 0 = non-speech region). We train our HMM on 10 minutes of speech where the hidden nodes are again hand labeled.

Figure 5 shows the segmentation results for a 35 second audio chunk. In this example two people wearing sociometers are talking to each other and are interrupted by a third person (between t=20s and t=30s). The output of low frequency energy threshold for each sociometer is fed into the speech HMM which segments the speech of the wearer. The shaded boxes overlaid on top of the speech signal show the segmentation boundaries for the two speakers. Also note that the third speaker's speech in the 20s-30s region is correctly rejected, as indicated by the grayed region in the figure.

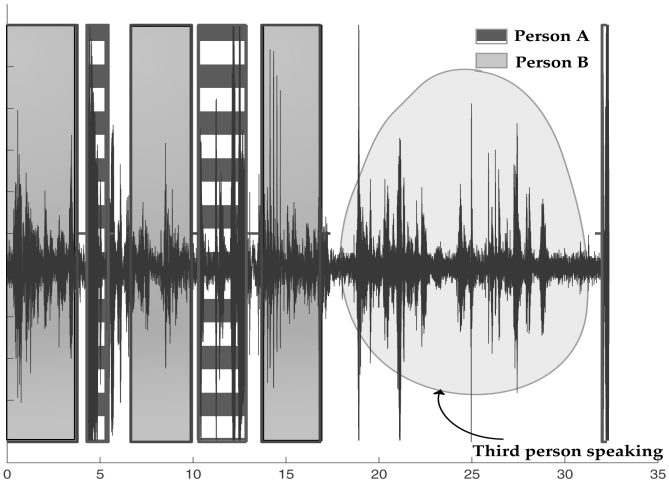

Figure 5 - Speech segmentation for the two subjects wearing the sociometer.

Purely energy-based approach to speaker segmentation is potentially very susceptible to the noise level of the environment and sound from the user's regular activity. In order to overcome this problem we have incorporated robust speech features (non-initial maximum of the auto-correlation, the number of auto-correlation peaks and the normalized spectral entropy) proposed in [13]. An HMM trained to detect voiced/unvoiced regions using these features is very reliable even in noisy environment with less than 2% error at 10dB SSNR. However, the downside of this is any speech and not just the user's speech is detected. So we use a second stage HMM model on the derived features based on energy to segment out only the user's speech and discard all the rest.

We now have information about when people are in close proximity and when they are talking. When two people are nearby and talking, it is highly likely that they are talking to each other, but we cannot say this with certainty. Results presented by Basu in [13] demonstrate that we can detect whether two people are in a conversation by relying on the fact that the speech of two people in a conversation is tightly synchronized. We reliably detect when two people are talking to each other by calculating the mutual information of the two voicing streams, which peaks sharply when they are in a conversation as opposed to talking to someone else. The conversation mutual information measure is as follow**:**

$$a[k] = I(v_1[t], v_2[t-k])$$

$$= \sum_{i,j} p(v_1[t]=i, v_2[t-k]=j) \log \frac{p(v_1[t]=i, v_2[t-k]=j)}{p(v_1[t]=i)\,p(v_2[t-k]=j)}$$

where v_1 and v_2 are two voicing streams and i and j range over 0 and 1 for voiced and unvoiced frames. The performance accuracy in detecting conversations was 63.5% overall and 87.5% for conversations greater or equal to one minute. These accuracy numbers were estimated from hand labeled data from four subjects, each of them labeled two days of their data (12 hours each). During the data collection stage we asked the subjects to fill out a daily survey providing a list of their interactions with other members. The survey data had 54% agreement between subjects (where both subjects acknowledged having the conversation) and only 29% agreement in the number of conversations.

LEARNING THE SOCIAL NETWORK

Once we detect the pair-wise conversation chunks we can analyze the actual communication patterns that occur within the community. Figure 6 shows the network interaction matrix based on conversations for the larger group. Each row shows the interactions of one person with other people in the network. The value of each entry (row *i* column *j*) is equal to person i's interaction with person j as a fraction of person i's total interaction. Subject IDs 2-8 belong to group 1, IDs 9,10,12-14, and 18 belong to group 2, IDs 15-17 and 19 to group 3 and 21-24 to group 4, IDs 20 and 25 were physically co-located with groups 1&2 (no one was assigned ID# 1 or 11). ID 1 and ID 11 were not assigned. The microphone of ID 15 failed most of the days we conducted the experiment and ID 23 did not have the sociometer on most of the time. So we have excluded their data in future analysis.

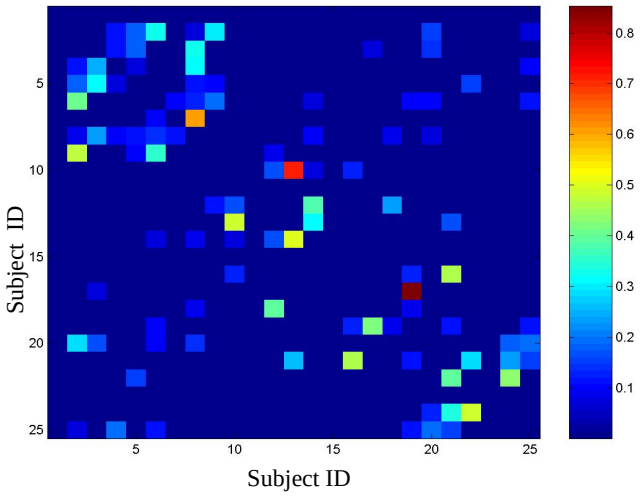

Figure 6: Interaction matrix. Each row corresponds to a different person. The values are proportional to fraction of total interaction.

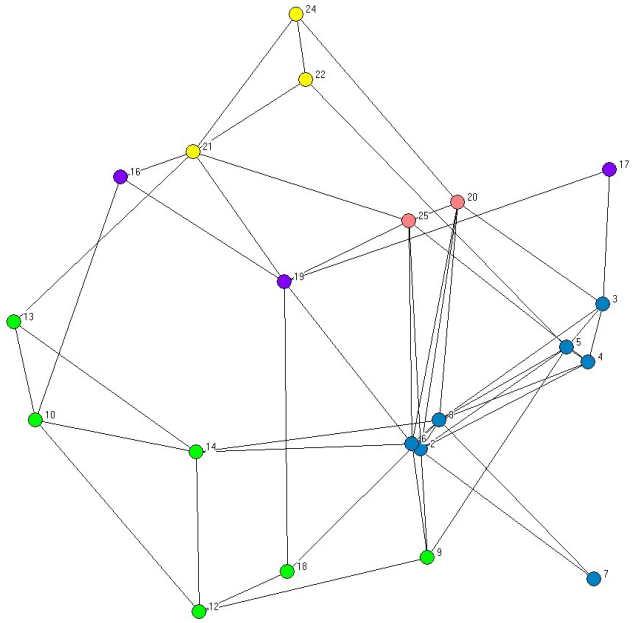

Figure 7: Network diagram based on multi-dimensional scaling of geodesic distances. Node numbers represent the subject IDs.

To get a more intuitive picture of the interaction pattern within the group of people who were equipped with sociometer, we visualize the network diagram by performing MDS (multi-dimensional scaling) on the geodesic distances. This type of visualization is commonly used in social network analysis [14]. The link structure for the nodes is calculated by thresholding or binarizing the interaction matrix, and the distances between a pair of nodes is the length of the shortest path connecting the two nodes. Multi-dimensional scaling provides a visual representation of the pattern of proximities among the set of people based on some distance measure [15], we use the geodesic distance as our distance metric. MDS method projects points from a higher dimensional space to a lower dimensional space (2D in our case) such that the distances in the projected space is as close as possible to distances in the original space. Figure 7 shows the network visualization obtained via MDS. The nodes are colored according to physical closeness of office location. People whose offices are in the same general space seem to be close in the communication space as well. In the next subsection we show the effect of distance on the overall communication pattern.

Effects of Distance on Face-to-Face Interaction

The architecture or the structural layout is known to affect the communication within an organization or community [2, 16,17]. We measured how the probability of communication changes as the physical distance between the subject increase. Figure 8 shows the probability of communication as a function distance between offices. We grouped distances into six different categories – (i) office mates (x-axis 0) (ii) 1-2 offices away (x-axis 1) (iii) 3-5 offices away (x-axis 2) (iv) office on the same floor (x-axis 3) (v) offices separated by a floor (x-axis 4) (vi) office separated by two floor (x-axis 5).

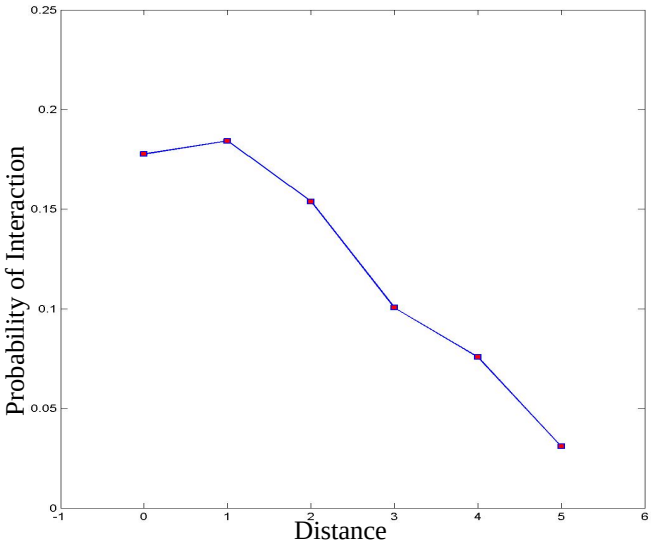

Figure 8: Probability of communication as a function of office distance. The distance is grouped into six different categories.

Changes Speech Activity throughout the day

We have also calculated the average talking pattern throughout the day based on the fraction of time that speech was detected from a wearer's device (for every one-minute unit of time). Figure 9 shows the daily pattern averaged over all the subjects and over nine days. This result is quite intuitive, as talking peaks during lunch time and also in the late afternoon when students often take breaks (the weekly Media Lab student tea is also held in the afternoon). These types of analysis of network behavior are much harder to do

using surveys or self-reports, whereas they are easily extracted from the analysis of the sensor data.

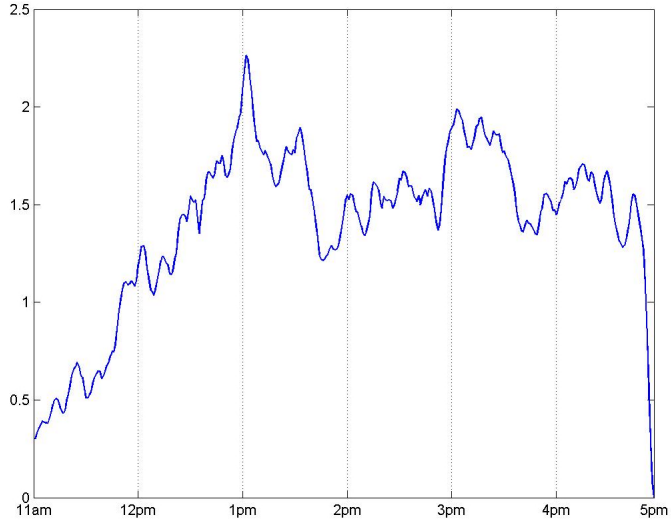

Figure 9: Speech activity over the course of the day averaged over all subjects

CONCLUSIONS

In many studies it has been shown that topology of people's connectivity is the most important feature and the actual interaction content is not as crucial in understanding a person's role within the community[1,18-20]. In this paper, we present a method for analyzing the connectivity of interacting groups using data gathered from wearable sensors. We have presented results from our efforts in sensor-based modeling of human communication networks. We show that we can automatically and reliably estimate when people are in close proximity and when they are talking. We demonstrate the advantages of continuous sensing of interactions that allows us to automatically measure the structure of communication networks and other network properties. Such analysis is very difficult and costly to do using manual approaches.

ACKNOWLEDGMENTS

This work has been partially supported by Center for Bits and Atoms NSF research grant NSF CCR-0122419. We want thank Brian Clarkson for helping with design of the sociometer, and Sumit Basu, whose work on Conversational Scene Analysis has guided our work on audio processing.

REFERENCES

1. Gladwell, M., *The Tipping Point: How little things make can make a big difference*. 2000, New York: Little Brown.
2. Allen, T., *Architecture and Communication Among Product Development Engineers*, 1997, Sloan School of Management, MIT, WP Number 165-97
3. Huberman, B. and Hogg, T., *Communities of Practice: Performance and Evolution.* Computational and Mathematical Organization Theory, 1995. **1**: p. 73-95.
4. Allen, T., Organizational Structure for Product Development, 2000, Sloan School of Management, MIT, WP Number 166-97
5. Gibson, D., Kleinberg, J., and Raghavan, P. *Inferring Web Communities from Link Topology*. In *9th ACM Conference on Hypertext and Hypermedia*. 1998.
6. Lukose, R., Adar, E., Tyler, J., and Sengupta, C. *SHOCK: Communicating with Computational Messages and Automatic Private Profiles*. In *Proceedings of the Twelfth International World Wide Web Conference*. 2003.
7. Gerasimov, V., Selker, T., and Bender, W., *Sensing and Effecting Environment with Extremity Computing Devices.* Motorola Offspring, 2002. **1**(1).
8. DeVaul, R. and Weaver, J., *MIT Wearable Computing Group*. 2002. http://www.media.mit.edu/wearables/.
9. Choudhury, T. and Clarkson, B., *Reference Design for A Social Interaction Sensing Platform,*, M.L.I.D. MIT. 2002: Cambridge.
10. Gemperle, F., Kasabach, C., Stivoric, J., Bauer, M., and Martin, R., *Design for Wearability*. 1998, Institute for Complex Engineered Systems, CMU. http://www.ices.cmu.edu/design/wearability/files/Wearability.pdf.
11. Marti, S., Sawhney, N., Jacknis, M., and Schmandt, C., *Garble Phone: Auditory Lurking*. 2001. http://www.media.mit.edu/speech/projects/garblephone.html
12. Jordan, M. and Bishop, C., *An Introduction to Graphical Models*. In press: MIT Press.
13. Basu, S., *Conversation Scene Analysis*, in *Dept. of Electrical Engineering and Computer science*. Doctoral. 2002, MIT. p. 1-109.
14. Wasserman, S. and Faust, K., *Social Network Analysis Methods and Applications*. 1994: Cambridge University Press.
15. Kruskal, J. and Wish, M., *Multidimensional Sclaing*. 1978: Sage.
16. Bertodo, R., *Evolution of an engineering organization*. International Journal of Technology Management, 1990. 3(6): p. 693-710.
17. Menzel, H., *Review of studies in the flow of information among scientists.* Columbia University Bureau of Applied Social Research, 1960.19.
18. Tyler, J., D. Wilkinson, and B. Huberman. *Email as spectroscopy: Automated discovery of community structure within organizations*. in *International Conference on Communities and Technologies*. 2003. Amsterdam, The Netherlands.
19. Granovetter, M., *The strength of weak ties.* American Journal of Sociology, 1973. **78**(6): p. 1360-1380.
20. Watts, D., *Six Degress: The Science of a Connected Age*. 2003: W. W. Norton & Company.