DriftCatcher: Enhancing Social Networks Through Email

Presented at the International Sunbelt Social Networks Conference XXII: Understanding, Recognizing, and Enhancing Online Communication Networks

Andrea Lockerd, Ted Selker
MIT Media Lab
20 Ames St, E15-313
Cambridge, MA 02139 USA
<alockerd, selker>@media.mit.edu

ABSTRACT
Computer-mediated communication (CMC) opens up new possibilities in the maintenance of social networks. Examples include the ability to visualize networks in real time, monitor the flow of resources, and facilitate network maintenance. This paper discusses the design of an electronic mail system, *DriftCatcher*, which facilitates users awareness of their electronic social networks. This system models what different relationships and socioemotional qualities "look like" in terms of email characteristics. Built through statistical analysis of a hand labeled email corpus, this model enables the system to characterize and display new email in network terms. Some examples include: tie strength, symmetry, foci of activity, and kinds of social support. Visualizing various network features will help the user become aware of their network and add a social context to online communication that is not present in current email applications.

INTRODUCTION
This work deals with how people inherently manage and understand their social network and social cues of communication. These cues can get lost in CMC making it harder for the user to naturally interact in their social structure. The principle concern of this work is understanding how a user comprehends and manages their social network and how CMC can facilitate and augment this experience. *DriftCatcher* is an agent that classifies email in social terms and characterizes the various relationships in the user's personal communication network; this paper motivates and introduces the work.

The idea that motivates this project is the party scenario: If you walk into a meeting or a party or some physical place with a number of people, you scan the room and see who's there. You automatically make mental notes like "oh I haven't seen that person in a couple weeks", "I just saw this person", or "there's a friend talking to someone I haven't met". All of this helps you make an agenda of how you organize yourself to approach the event and the various people there, and is an example of how people automatically use social network analysis in face-to-face interactions. Many of the social cues necessary to do this are not inherently obvious in CMC, which therefore obfuscates the maintenance and utilization of ones' social network online. We submit that computers should perform automatic social network analysis in order to foster the user's coherent understanding of the people and resources of their communication network.

The question to address is whether or not a computer can recognize social resources in electronic communication. In order to address this I am carrying out this semester, for my masters thesis, the following method: build a classifier on a training data corpus of email, build a software agent that classifies email in social terms and characterizes the various relationships in the user's personal communication network. The system will then be fit to evaluate the degree to which the addition of social context information in the electronic communication interface augments a user's communication experience.

Background
A strong advocate for a system such as *DriftCatcher* is an ethnographic study done by Nardi, Whittaker and Heinrich [7]. They analyzed how people use and maintain their social networks at work, and how central this activity is to their success. They argue that netWORKing, (the process of building, maintaining, and activating your social network) is an absolute necessity in the modern work environment even more so than in the past since the work climate is much more transient and less defined by physical spaces. In this study it was noted that modern communication tools do not aptly support netWORKing, *DriftCatcher* would fall under the type of system they are promoting.

The motivation and interest in this work stems from theories and findings in the field of social networks concerning how people interact with each other both on and off line. Some of the attributes of social networks that *DriftCatcher* will be representing include: tie strength, symmetry, foci of activity, and types of social support. Some examples of the theories and studies that

support these are: social capital [6], the amount of support (of all forms) which can be called upon from the people in your social network; strength of weak ties [4], a group of studies which indicate that the people most important to you in terms of access to information and resources are on the outskirts of your social network; and studies of information flow [13], which describe the process by which information and influence disseminate across a network of individuals.

Additionally, the attempt to recognize social resources assumes that social and emotional resources are exhibited in electronic communication. The socioemotional content of email is addressed in [10] and [12], and Orlinkowsky [9] takes this further and defines communication in terms of media genres.

Prior Work

The systems such as [1], [11] and [8], deal with how to best visually represent the conversations in electronic communication. The Referral Web system [5], finds a path between two people in a social structure using a closeness metric based on web documents. Yenta [3] is a multi-agent system for matchmaking, based on subject matter of email messages to suggest matches between users. [2] is a speech-act application that tries to identify patterns of speech in an organization related to the sort of action that speech tends to induce.

All of the systems mentioned here show various aspects of social network analysis applied to computer applications. There are two main qualities that differentiate the work here. A personal network approach will be used; rather than take the point of view of a whole organization or community this work understands a social network from the point of view of a single user. Secondly, most of the current applications of social networks and online communication deal with information flow and task-oriented resources. *DriftCatcher* will attempt to recognize the social resources exchanged between people in the network that characterize various types of relationships.

APPROACH

The basic strategy starts with collecting a corpus of emails for training a classifier, and labeling each email with metadata pertaining to social intention and relationship between the sender and the receiver. Then various machine-learning techniques are applied to this corpus in order to find discriminating features and weigh the extent to which the classes depend on these features. After this phase, an email agent will be built to take advantage of this information and enhance the communication experience. The remainder of this section will go through my proposed methodology for completing this process.

Statistical Feature Analysis

This phase of the problem involves extracting features from email that have social significance in order to build a representation of the characteristics of the social connotation of email. The goal is to be able to put a particular piece of mail into a social context such as formal, informal, excited, urgent, calm, happy, sad, etc. These are examples of classes with which the data will be labeled.

There are a number of features that I believe will vary significantly depending on the social context and relationship between the sender and the receiver. Some obvious features are the information contained in the header: To, From, CC, Date, Subject. The frequency of communication, reply time, time of day the message was sent, and individual communication versus group mailing list communication, are a few examples of socially significant features that can be extracted from the header information. In addition to these, I think other features will prove to convey significant socioemotional information. For example: length of the mail, average length of the sentences, whether or not there are misspellings or non-dictionary words, characteristic opening and closing phrases, and whether or not the sender included a signature.

The hypothesis is that there are subsets of features that discriminate between the various classes of social context. A first goal is to cluster the data in terms of these features and classes. Once a subset of discriminating features is found, it is necessary to estimate the weight of each feature. For example, time of day and frequency of contact may be the discriminating features between the formal and informal classes of email, but they might not be equal contributors. Additionally, it seems intuitively plausible that certain relationship information may only be apparent in the analysis of email through time.

Application

Applications of social network analysis and email already exist. Some of these require the user to input their network and relationships by hand. Others have various schemes for calculating the closeness or strength of relationship between two people. These schemes are sub-optimal, for most of them tend to deal with relationships in the context of subject matter without regard for emotional/social content. The proposed *DriftCatcher* agent will have a more realistic representation of a social network. This agent generates a network by looking at header info of emails to come up with the structure (who is connected to who), which is then informed about the types of social resources exchanged in a relationship from the social context of the mail between two people in the network using the classifier described above. Thus letting the agent catch the drift of what's going on with the user's various relationships.

With this model of a user's social network, *DriftCatcher* will be in a position to organize and visualize mail according to social information. This will make it easier for the user to see what is happening in the various groups of their network, and allow them to deal with communication in a social context rather than the current temporal context of mail browsers.

User Scenario

Once implemented we expect the email application, enhanced with social network analysis, to have the following features:

- It allows the user to bring up a graph of their social network and see various statistics like symmetry of relationships and frequency of contact.
- At any point the user can switch to a view that lets them see where the most activity has been going on in their communication network using varying intervals (this hour, this day, this week).
- A new piece of mail is prioritized based on when it came into the system, who the sender is and how frequently the user communicates with this person, and whether or not the system detects that the mail is of a time sensitive matter.
- Social resources of a particular piece of mail are revealed in the stationary the system chooses for display. Social mail has a flower by the header; work mail has a computer, etc.

In an email application with these features we anticipate the following typical user scenario. Sara sits down to check her email at the end of the day, and her inbox has thirty new messages. She starts reading through them and replying with confidence that the system has brought the most important mail to the top of the queue. While she's reading through her mail a new piece of mail comes in and is brought to her attention because it is an inquiry from a person with whom she is in frequent communication and they need a response as soon as possible. Consequently, two other pieces of mail came in at the same time, but were not brought to Sara's attention because one was from someone she's never talked to before and the other was not recognized as a time sensitive matter.

CONCLUSION

This work uses automatic social network analysis to foster the user's coherent understanding of the people and resources of their communication network, and experiments with ways in which CMC can facilitate and augment this experience. Many of the social cues that allow us to understand and manage our social networks get lost in CMC making it harder for the user to naturally interact in their social structure. *DriftCatcher* is an agent that classifies email in social terms and characterizes the various relationships in the user's personal communication network, exploring the extent to which a computer can recognize and utilize social resources in electronic communication.

REFERENCES

1. Donath, J., Lee, H., Boyd, D., & Goler, J. Loom2: Intuitively visualizing usenet, CSCW 2001 (workshop) http://smg.media.mit.edu/projects/loom2

2. Flores, F., Graves, M., Hartfield B. & Winograd, T. Computer System and the Design of Organizational Interaction, *ACM Trans. on Information Systems*, Vol. 6. No. 2, 1988, (153-172).

3. Foner, L. Yenta: Multi-Agent Referral System for Matchmaking, *First International Conference on Autonomous Agents*, 1997.

4. Granovetter, M. The Strength of Weak Ties: A Network Theory Revisited, Social Structure and Network Analysis, edited by Marsden, P. and Lin, N. Beverly Hills: Sage, 1982, (105-130).

5. Kautz, H., Selman, B., Shah, M. Referral Web: Combining Social Networks and Collaborative Filtering, *Communications of the ACM*, Volume 40, No. 3, March 1997, 63-65.

6. Lin, N. Building a Network Theory of Social Capital, Social Capital: Theory and Research, edited by Lin, N., Cook, K., and Burt, R.S., New York: Aldine de Gruyter, 2001, (3-30).

7. Nardi, B., Whittaker, S., Heinrich, S. It's Not What You Know, It's Who You Know: Work in the Information Age, *First Monday*, May 2000, http://www.firstmonday.org/issuse/issue5_5/nardi/index.html.

8. Newman, P.S. Treetables and Other Visualizations for Email Threads, Technical Report in the Human-Document Interaction Group at Xerox PARC, http://www.parc.xerox.com/istl/groups/hdi/papers/psn_emailvis01.pdf

9. Orlikowski, W. J. and Yates, J. Genre repertoire: The structuring of communicative practices in organizations. *Administrative Sciences Quarterly*, 33, 1994, (541-574).

10. Rice, R. and G. Love (1987, February). Electronic emotion: Socioemotional content in a computer-mediated communication. *Communication Research* 14 (1), (85-108).

11. Sac, W. Conversation Map: a Content-based Usenet Newsgroup Browser, *Proc. of IUI 2000*, (233-240).

12. Walther, J. B. (1992, February). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication Research* 19 (1), (52-90).

13. Weimann, G. On the Importance of Marginality: One More Step into the Two-Step Flow of Communication, *American Sociological Review*, Volume 47, Issue 6, Dec., 1982, (764-773).