Temporal Visualization and Analysis of Social Networks

Peter A. Gloor*, Rob Laubacher
MIT
{pgloor,rjl}@mit.edu

Yan Zhao, Scott B.C. Dynes
*Dartmouth
{yan.zhao,sdynes}@dartmouth.edu

Abstract

   This paper describes a visual social browser for exploring the evolution of social networks over time. We consider the exchange of e-mails between actors as an approximation of social ties. Our system analyzes the dynamic progression of communication patterns of e-mail traffic within groups of individuals. It combines a discrete visualization view, a continuous visualization view, and an adjacency matrix view. The goal of our work is to develop a framework of visual temporal communication patterns of different types of collaborative knowledge networks. As a first application, our tool is used to analyze communication patterns and make recommendations for improved productivity in innovation communities in an emerging management consulting practice.

Contact:
Peter A. Gloor
Center for Coordination Science
MIT
Cambridge, MA 02142

Tel: 1-617-253-7018
Fax: 1-617-253-4424
Email: pgloor@mit.edu

Temporal Visualization and Analysis of Social Networks

Peter A. Gloor, Rob Laubacher, Yan Zhao, Scott Dynes

In this paper we introduce a visual browser for the visualization and analysis of social links (relationships). Our visual browser displays the progression of communication networks between individuals over time. Our goal is to come up with an environment for the analysis of the dynamics of communication in social spaces, while respecting individual privacy. While there has been substantial research dedicated to visualizing static social networks as directed graphs or adjacency matrices [Johnson 2000, Smith & Fiore 2001, Tyler, Wilkinson & Huberman 2002, Van Alstyne & Zhang 2003], little work has been done so far to visualize the evolution of social networks over time [Holme, Edling & Lijeros 2003.

Other researchers have analyzed email communication flow to study the community structure [Danah, Potter & Viegas 2003, Ebel, Mielsch & Bornholdt 2002, Girvan & Newman 2001], or automatically identifying communities, solely based on the frequency of the message exchange between individuals, by partitioning the total graph [Guimera et. al. 2002]. For visualization and analysis of those networks, there is a wealth of social network analysis tools available such as pajek [Batagelj & Mrvar, 1998], ucinet [Borgatti, Everett & Freeman, 1992], or AGNI [Varghese & Allen, 1993]. While pajek and ucinet also include the option to animate the graph, they do not allow temporal visualization similar to our own social browser.

Further research has used email data to map communication patterns from the perspective of the individual. This work typically creates representations of past messages that allow an individual to see the personal network implied by their prior e-mail traffic. Other studies have analyzed virtual communities by applying social network analysis methods and metrics [Ahuja & Carley, 1999,].

System Overview

We have implemented a flexible but scalable architecture. E-mail messages are processed locally in three steps. In the first step, the e-mail messages and mailing lists are parsed and stored in decomposed format in a database on the local machine. In the second step the database can be queried to select messages sent or received by a group in a given time period. In the third step the selected communication flows can be represented in our visual browser using our own netgraph and static and dynamic innomap views [Gloor et. al. 2003] or SNA visualization tools such as pajek and ucinet. Our own social browser is particularly optimized to add a temporal dimension to the visualization and manipulation of e-mail-based communication of a large number of actors. While it is straightforward to visualize static social networks as directed graphs or adjacency matrices, little work has been done so far to visualize the evolution of social networks over time.

For the visual placement of vertices and edges we use the Fruchterman-Reingold graph drawing algorithm [Fruchterman & Reingold 1991] for force-directed placement, which is commonly used to visualize social networks. This method compares a graph to a mechanical collection of electrically charged rings (the vertices) and connecting springs (the edges). Every two vertices reject each other with a repulsive force, and adjacent vertices (connected by an edge) are pulled together by an attractive force. Over a number of iterations, the forces modeled by the springs are calculated and the nodes are moved in a bid to minimize the forces felt.

Social network graphs attempt to represent the strength of social ties between parties. In our algorithm, we treat the exchanges of e-mail between actors as an approximation of social ties. In our visualization a communication initiated by actor A to actor B is represented as a directed edge from A to B. The more interaction between actors A and B occurs, the closer the two representing vertices will be placed. The most connected actors are placed in the center of the graph.

Our system currently includes three views: temporal visualization in a discrete and a continuous way as well as an adjacency matrix view. For the continuous temporal visualization we propose a new algorithm, called the sliding time frame algorithm described in the next section. While the three views have some common features, they work independently and give observers different information.

Continuous Temporal View

To visually analyze the evolution of communication patterns over time, we developed a dynamic visualization algorithm where the layout of the graph is automatically recalculated every day, resulting in an interactive movie.

The simplistic approach would be, for any given day, to base the graph structure on the communications that occurred during this day. However, this approach does not take into account communications that happened before

this day or after this day in a specific time-frame. For our dynamic visualization, we propose a new algorithm based on the FR algorithm: the sliding time frame algorithm.
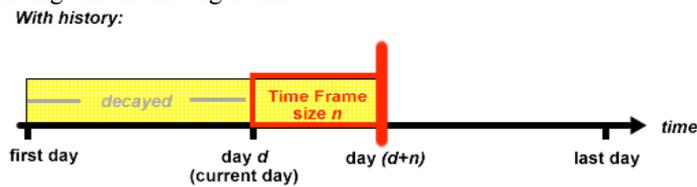

Figure 1. Sliding time frame algorithm in "with history" mode

The basic idea of the sliding time frame algorithm is to display active ties between actors in a sliding time frame covering a flexibly selected interval of n days starting from the specific day the visualization is showing. The window frame moves forward day by day, and new ties are subsequently added to the graph each day until the desired width n of the sliding time frame is reached. This time frame window allows users to foresee the activities happening inside the time frame after the current day. By default, all the old communication activities before the current time frame window are included in the layout of the graph.

In figure 1 the time frame moves to day d as the animation progresses. Thus, day d is the current day that the visualization is showing and the current time frame is [d, d+n]. All communications through day (d+n) are calculated and displayed, and if a communication takes place before or on day d, it is active. In the "No history" mode (figure 2), only the edges in the current window are included. The time frame moves to day d as the animation goes. Thus, day d is the current day that the visualization is showing and the current time frame is [d, d+n]. Only communications inside the current time frame are calculated and displayed, and only communications on day d are considered active.
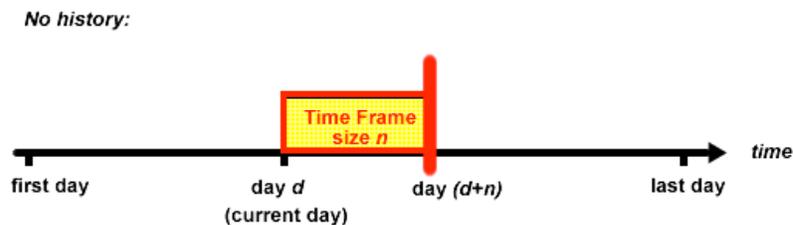

Figure 2. Sliding time frame algorithm in "no history" mode

To define the amount of "animated action" and animation speed we are using "keyframes" adjustable by the user based on the number of new edges appearing in the visualization. The animation of the changing layout is interpolated between those keyframes.

Application: Correlating Temporal Communication Patterns With Innovation

We are aiming to distinguish temporal communication patterns typical of different types of innovation networks. Our hypothesis is that innovation networks are core/periphery structures [Borgatti & Everett, 1999] with small world properties. They consist of a central cluster of people, the core team, forming a network with low centrality, but high density. The external part consists of a network forming a ring around the core team. It has comparatively low density, but high centrality, thanks to the central team. The actors in the outer ring have low betweenness centrality, as they are only connected to core team members, but not among themselves. In the experiment outlined below we investigated this hypothesis and made initial correlations between these patterns and success of the individual efforts, given our knowledge of the outcome of these endeavors and their communication patterns.

Our dataset consists of a one-year e-mail archive of a 200 people global consulting practice. As an approximation of the ego network of the practice leader we are using his mailbox, similarly we obtained the mailbox of the practice coordinator as an estimate of his ego network. We are taking these two mailboxes as an approximation of the organizational memory of the consulting practice. We distinguish 15 communities through messages grouped by the practice coordinator and the practice leader into separate mailfolders. The 15 communities consist of 8 innovation teams developing new consulting service offerings, of the sales and marketing activities, of a weekly brownbag that was also used to coordinate global activities of the practice, of the organization of a global Webinar, of the development of the practice Web site, and of the team handling the collaboration with software

vendors. As a measure of the performance of the communities we are taking the quality of the community output as judged by the practice leader.
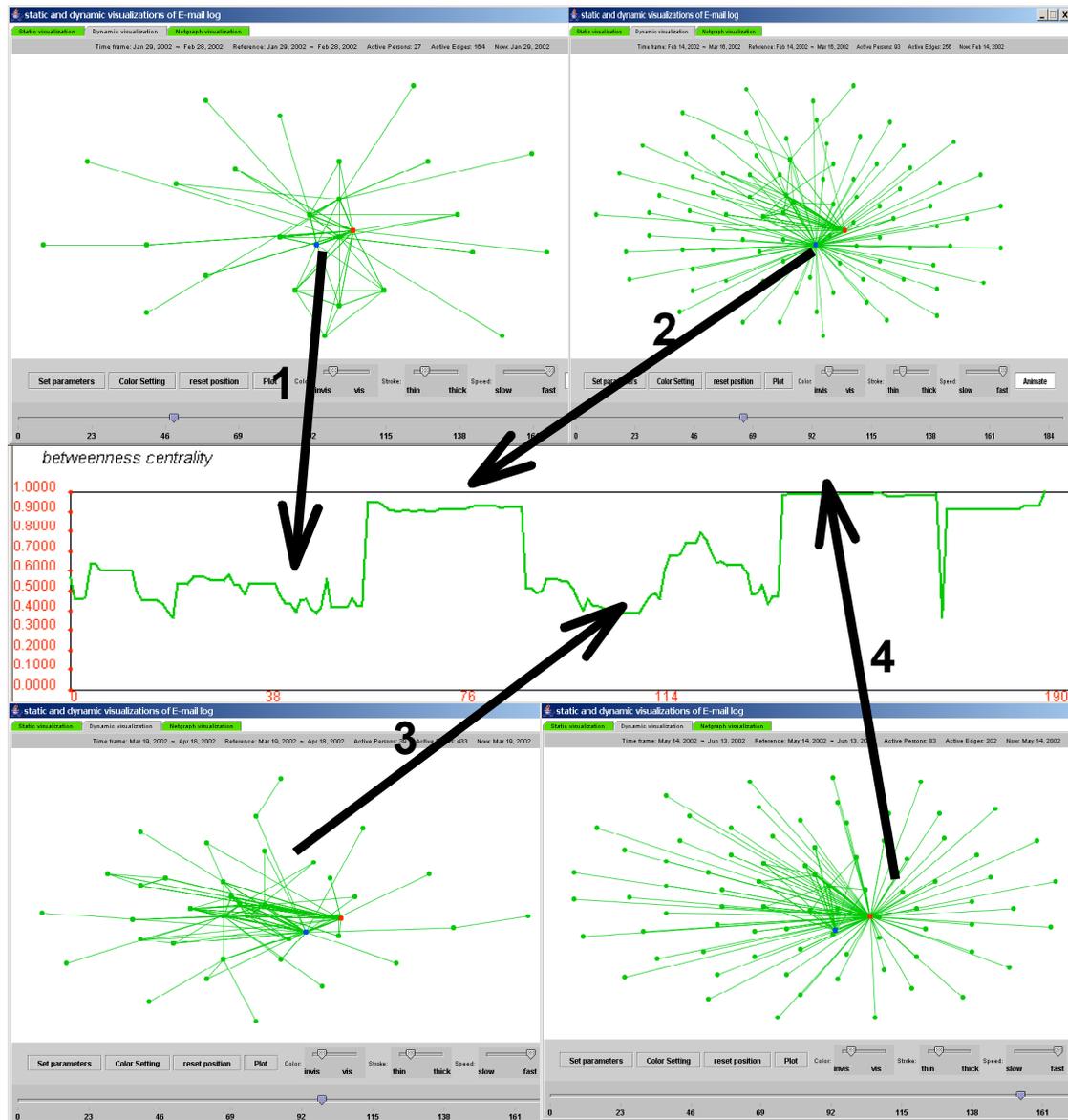

Figure 3. Four Screen shots of movie of Webinar visualization (sliding time frame 30 days) and evolution of group betweeness centrality

Figure 3 illustrates the progress over time of the communication activities around preparing and conducting a Webinar, i.e. a Web based conference. The picture in the top left of figure 3 shows the structure of the team preparing the Webinar. This team has high density and relatively low group betweeness centrality, as shown by the first arrow in the center of the graph. The picture in the top right of figure 3 shows a screen shot of the communication pattern during the first time the Webinar was conducted. The practice leader (blue dot) is sending and receiving information in a star structure, the graph in the center as pointed out by the second arrow displays now relatively high centrality. The third picture at the lower left displays the team preparing a rerun of the Webinar, again communicating with relatively low centrality (third arrow). The final screen shot in the lower right of figure 3 displays the practice coordinator (red dot) rerunning the Webinar, communicating in a star structure with relatively high centrality (fourth arrow) with his audience.

For our test dataset our temporal analysis conveys new insights which would have been much more expensive to obtain with traditional means. Our tool offers a fast way to find periods of low and high centrality, and to identify periods of high productivity and information dissemination. Nevertheless we needed other contextual cues to obtain a full understanding of the activities, such as interviews with community members and a content analysis of the e-mail messages.

Our continuing goals are to gain deeper insights into the correlation of the evolution of online group dynamics with innovation, and developing a theory of member roles in innovation communities.

References

[Ahuja & Carley, 1999] Ahuja, M. & Carley, K. 1999, Network Structure in Virtual Organizations. Organization Science, Vol. 10, No 6. 741-757.

[Batagelj & Mrvar, 1998] Batagelj, V. & Mrvar, A. 1998, Pajek - Program for Large Network Analysis. Connections 21, 2, 47-57

[Borgatti, Everett & Freeman, 1992] Borgatti, S. Everett, M. & Freeman, L.C. 1992, UCINET IV, Version 1.0, Columbia: Analytic Technologies

[Borgatti & Everett, 1999] Borgatti, S. & Everett, M. 1999, Models of core/periphery structures, Social Networks 21 375-395

[Cummings & Cross, 2003] Cummings, J. & Cross, R. 2003, Structural properties of work groups and their consequences for performance, Social Networks

[Danah, Potter & Viegas, 2003] Danah, B. Potter, J. & Viegas, F. 2003, Fragmentation of identity through structural holes in email contacts. Proc. Sunbelt XII.

[Ebel, Mielsch & Bornholdt, 2002] Ebel, H. Mielsch, L. & Bornholdt, S. 2002, Scale-free topology of e-mail networks. arXiv:cond-mat/0201476v2 12 Feb

[Fruchterman & Reingold, 1991] Fruchterman, T.M.J & Reingold, E.M. 1991, Graph drawing by force directed placement. Software: Practice and Experience, 21(11)

[Girvan & Newman, 2001] Girvan, M. & Newman, M.E.J.2001, Community structure in social and biological networks. arXiv:cond-mat/0112110v1, 7 Dec.

[Gloor et. al., 2003] Gloor, P. Laubacher, R. Dynes, S. & Zhao, Y. 2003, Visualization of Communication Patterns in Collaborative Innovation Networks: Analysis of some W3C working groups. Proc. ACM CIKM 2003, New Orleans, Nov. 5-6, 2003

[Guimera et. al., 2002] Guimera, R., Danon, L., Diaz-Guilera, A. Giralt, F. & Arenas, A. 2002, Self-similar community structure in organizations. ArXiv:cond-mat/0211498 v1, 22 Nov.

[Holme, Edling & Lijeros, 2003] Holme, P. Edling, C.R. & Lijeros, F. 2003, Structure and Time-Evolution of an Internet Dating Community, arXiv cond-math.0210514, 15 Oct.

[Johnson 2000] Johnson, S. 2003, Who Loves Ya, Baby? Discover, Vol. 24 No. 4 (April) http://www.discover.com/apr_03/feattech.html

[Smith & Fiore, 2001] Smith, M. & Fiore, A. 2001, Visualization components for persistent conversations, ACM CHI.

[Tyler, Wilkinson & Huberman, 2002] Tyler, J. Wilkinson, D. & Huberman, B. 2002, "Email as Spectroscopy: Automated Discovery of Community Structure within Organizations" http://www.hpl.hp.com/shl/papers/email/index.html

[Van Alstyne & Zhang, 2003] Van Alstyne, M. & Zhang, J. 2003, EmailNet: A System for Mining Social Influence & Network Topology in Communication. University of Michigan working paper

[Varghese & Allen, 1993] Varghese, G. & Allen, T. 1993, Relational Data in Organizational Settings: An Introductory Note for Using AGNI and Netgraphs to Analyze Nodes, Relationships, Partitions and Boundaries. Connections, Volume XVI, Number 1 & 2, Spring