TeCFlow – A Temporal Communication Flow Visualizer for Social Network Analysis

Peter A. Gloor
MIT Center for Coordination Science
pgloor@mit.edu

Yan Zhao
Dartmouth College
yan.zhao@dartmouth.edu

ABSTRACT

This paper introduces an approach for organizational redesign and optimization of communication flows based on temporal analysis of communication patterns in groups of people. Our Temporal Communication Flow Visualizer automatically generates interactive movies of communication flows among individuals by mining e-mail log files and other communication archives. Combining those movies with measures of social network analysis such as the change over time in group betweeness centrality (GBC) and group density leads to deep insights into organizational dynamics. In addition we have defined a contribution index, which measures the activity of an individual as a sender and receiver of messages relative to a team. Based on these findings we can make predictions on the productivity of teams and suggest interventions for improved performance.

DESIGN AND ARCHITECTURE OF TECFLOW

Our **Te**mporal **C**ommunication **Flow** Visualizer for the temporal analysis of social networks (TeCFlow) addresses three areas of related research: (1) visualization of social networks, (2) temporal analysis of social networks in animated visualizations, and (3) analysis of e-mail networks.

TeCFlow takes as input e-mail archives and automatically generates static and dynamic visualizations of the calculated communication networks. The static visualizations allow users to step through a chosen time period by looking at communication networks at subsequent time intervals. The dynamic visualization consists of an interactive movie showing the evolution over time of the communication network within the group. Active relationships are displayed in a sliding time window, with inactive relationships decaying over time. TeCFlow also calculates and plots the evolution of group betweeness centrality and density over time to discover interesting events in the lifetime of a virtual team. The interactive movie can be stopped anytime to drill-down into the messages that are currently exchanged between actors. Multiple e-mail addresses can be combined into an online personality, reflecting the fact that people frequently use different e-mail addresses.

We have implemented an open architecture. E-mail messages are processed locally in three steps (figure 1). In the first step, the e-mail messages and mailing lists are parsed and stored in decomposed format in a mysql 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 [6] and static and dynamic views [5].

This architecture provides a tested of high scalability and flexibility. The number of actors, ties, and messages to be analyzed is only limited by the size of the database and the amount of RAM available, and temporal queries can be run in an ad hoc way. We are also able to experiment with different visualizations of the retrieved structure and to easily add other social web applications.

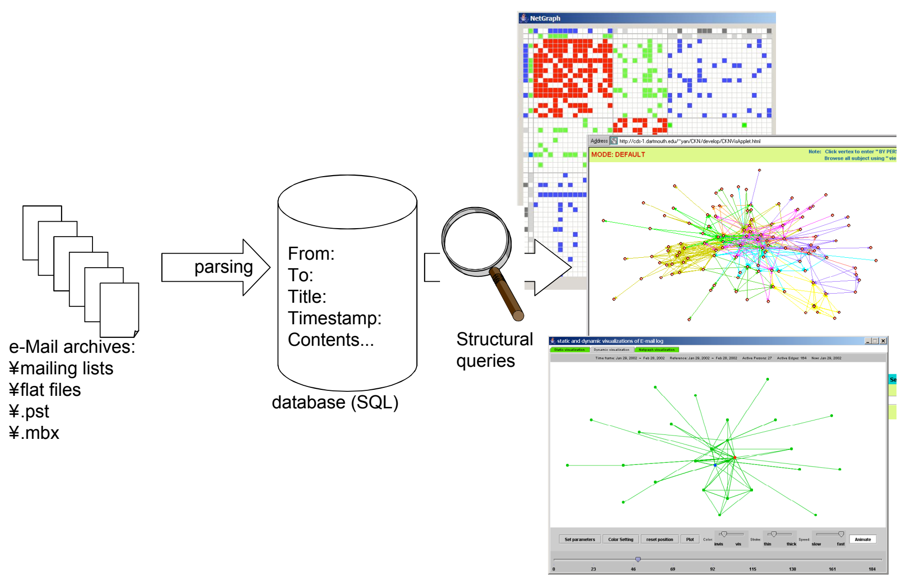

Figure 1. TeCFlow Architecture

We base our algorithm on the Fruchterman-Reingold graph drawing algorithm [4] 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.

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, i.e. a message sent from A to B is depicted as an arc. The more interactions between actors A and B occur, the closer the two representing vertices will be placed. The most connected actors are positioned in the center of the graph. This means that the actors who send and receive the largest number of e-mail messages in a given time frame are placed in the center of the graph. Similarly, the more messages A and B exchange, the shorter their connecting arc becomes.

To display 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 have been, 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 interactions that happened before this particular day, and would result in a jerky animation of low quality. For our dynamic visualization, we therefore propose a new algorithm: the **sliding time frame algorithm**, where we are always looking at a time interval consisting of a flexibly chosen number of days.

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 current day d the visualization is showing. The window frame moves forward day by day, and new ties (i.e. e-mail messages exchanged) are subsequently added to the graph each day until the desired width of n days of the sliding time frame is reached. This time frame window allows users to see all activities happening inside the time frame after the current day. By default, old communication activities before the current time frame window are included in the layout of the graph. This reflects persistent ties that stay active once they have been established for the remainder of the lifetime of the team. Once an e-mail message has been sent, it will influence positioning of the actors in the graph for the rest of the animation, meaning that a link does not decay. Rather, after it moves out of the n-days wide time frame, it is displayed in the visualization as a dimmed out arc.

TREATMENT OF INDIVIDUAL ACTORS

In addition, TeCFlow also allows the user to define personalities consisting of multiple e-mail addresses, and identify groups consisting of multiple personalities.

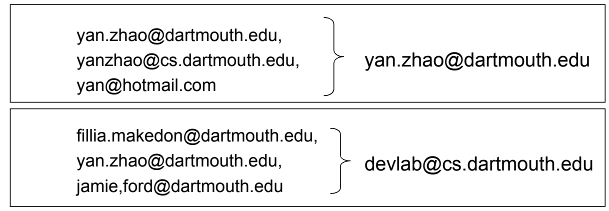

Figure 2. Merging multiple e-mail addresses of persons and groups

In figure 2, actor yan.zhao@dartmouth.edu consists of three e-mail addresses. The group devlab@cs.dartmouth.edu is composed of three actors. Maintaining the organization and domain parts of the e-mail addresses permits automatic analysis on the domain and organizational level.

We also looked at the frequency with which individuals send and receive messages. We have defined a measure, which we call the "contribution index" [5].

The contribution index is +1, if somebody only sends messages and does not receive any message. The contribution index is –1, if somebody only receives messages, and never sends any message. The contribution index is 0, if somebody has a totally balanced communication behavior, sending and receiving the same number of messages:

$$CI = \frac{messages\_sent - messages\_received}{messages\_sent + messages\_received}$$

We then plotted the contribution index against the total number of messages sent and received of each participant.

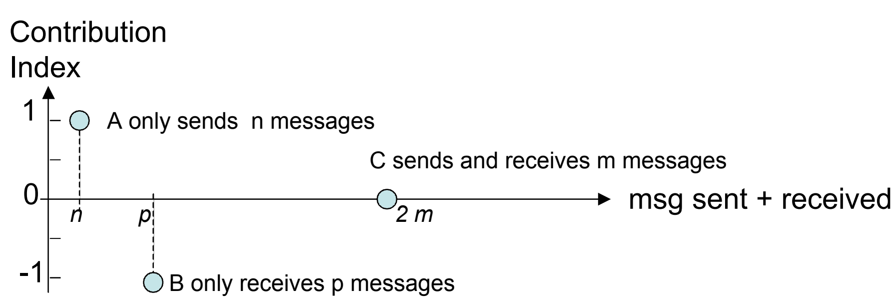

Figure 3. Contribution Index

In figure 3, actor A only sent n messages, never receiving any, actor B only received p messages, never sending any, while actor C sent and received m messages (C is located on the x-axis, because C sent and received the same number of messages).

APPLYING TECFLOW STEP BY STEP

The TeCFlow tool is used in three steps:

(1) Watch social interaction pattern movies to find dense clusters indicating potential emergence of collaborating teams.

(2) Look for peaks and troughs in the temporal evolution of group betweeness centrality and density to find the most "interesting" phases of collaboration in the lifetime of the team.

(3) Look at the contribution index to better understand the roles of the individuals in teams.

Connecting the dots by combining steps (1) to (3) leads to new insights not easily obtainable through other means by giving a thorough understanding of the team dynamics during the chosen time interval. In this section we illustrate the use of TeCFlow by analyzing a globally active research and development community of a global management consulting firm.

Our sample data set consists of an e-mail archive of a virtual consulting practice with 200 members of a global consulting firm covering the time period from mid-2000 to early 2002. It is composed of the ego-networks of the practice leader and the practice coordinator (i.e. their e-mailboxes). Those e-mailboxes are taken as an approximation of the organizational memory of the consulting practice, as the practice leader and the coordinator were informed of all major events in the practice. The mailboxes were partitioned manually into mail folders by subject areas. Mail folders included one of eight service offerings, a folder for each project currently active, sales efforts, marketing activities, and the organization of two practice-wide seminars conducted over the Web ("Webinars"). The major advantage of this data set is that one of us was intimately involved in the consulting practice. The disadvantage is that the mailboxes of the practice coordinator and the practice leader do not include the direct one-to-one communication among the practice members bypassing the practice coordinator or the practice leader.

The three step-process of using the TeCFlow tool as outlined above is now applied to this data set. To gain on overview of communication activities in the practice, we start with an analysis of all messages of the practice.

Step 1 – Watch movies to find communities

By watching a movie containing the entire e-mail traffic of the consulting practice we are able to discover core/periphery structures [1,2,3]. This is a strong indicator for the emergence of an collaborative innovation team (top and bottom windows of figure 4).

Step 2 – Find interesting time periods.

In the next step we look at the progression of group betweeness centrality (GBC) and group density over time. A rapid change in slope in the graph, i.e. a spike or a trough, indicates an interesting event. In those cases, we go back to the movie, and drill down into the graph by clicking on interesting actors, and looking at the contents of the e-mail messages exchanged.

The two troughs in the GBC graph in the center left window of figure 4 correspond with periods of high activity of the community.

Step 3 – identify most active actors

In the third step, plotting the contribution index identifies the most active actors. In figure 4 the contribution index pattern of the team (center right window of figure 4) is almost consistent with our earlier results. In [5] we found that the coordinator sends the most messages, sending more than he receives, while the leader sends fewer messages, and receives significantly more than he sends.

In figure 4, the practice leader receives vastly more messages than he sends, and the practice coordinator is the most active sender of messages. Usually the practice coordinator would also be the most active contributor. Surprisingly though, a practice member is the most active participant, making herself the leader at the core of a new innovation community.

Combining the steps – analyzing the birth of a new Community

The analysis with TeCFlow allows us an intimate look into the emergence of new teams and online communities not possible by other means.

The creation of the new community as well as the emergent role of the leader of the new community would have remained completely hidden, had we not combined the contribution index plot with the dynamic movie. The group betweeness centrality view allows us to quickly zoom into time periods of particular interest, where we then can use the drill-down features of the dynamic view to look at the contents of the e-mail exchange.

Surprising results of this analysis are:

- The emergence of a new innovation team, coming up with a new and creative consulting service offering.
- The central role of a non-executive member of the consulting practice in creating this new service offering.
- The easy identification of the time period when the new innovation team was most active.
- Easy identification of the core team members of the new innovation team.

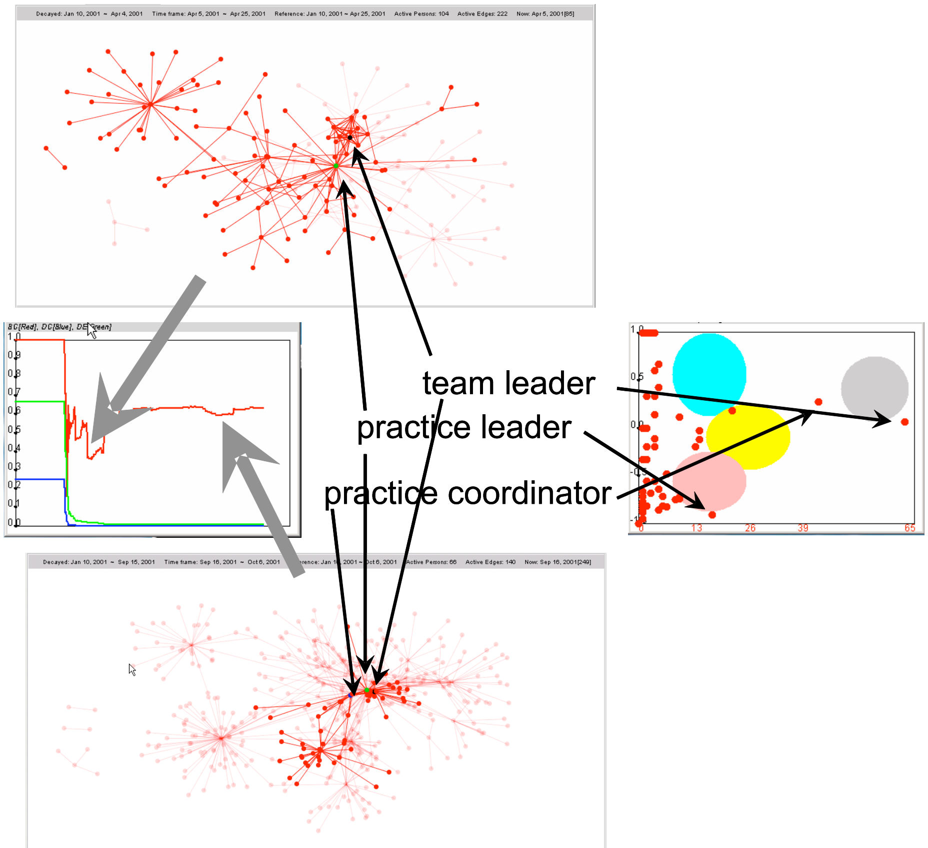

Figure 4. 3-Step Analysis with TecFlow, combining movie, GBC, and Contribution Index

This analysis was done two years after the data had been collected. Had this tool been available to monitor virtual interaction in real time, senior management would have been much better able to adequately support these activities, thereby reducing time to market while also increasing awareness of the team's output within the consulting firm.

FUTURE WORK AND CONCLUSIONS

We are currently working to develop a multiuser version of our system, where users can upload anonymized communication data sets to a "Global Social Web" under strict privacy and anonymity. We hope that this will encourage users to share their communication data such that we can get a much broader view on social interactions than is possible until now. We also plan to use our tools for other types of communication activities. Because TeCFlow runs on top of a database, it is straightforward to import, for example, phone logs, instant messaging logs, and blog transcripts into the database instead of e-mail archives.

Our continuing goals are to gain deeper insights into the evolution of online group dynamics and developing a theory of member roles in virtual communities using more detailed communication pattern analysis.

REFERENCES

1. Borgatti, S.P. & Everett, M.G. (1999). Models of Core/Periphery Structures. Social Networks 21: 375-395.
2. Bulkley, N. & Van Alstyne, M. (2004), Why Information Should Influence Productivity. (forthcoming) In The Network Society: A Cross-Cultural Perspective. Manuel Castells, ed. Edward Elgar Publishing.
3. Cross, R. & Cummings, J. (2003). Relational and Structural Network Correlates of Performance in Knowledge Intensive Work. Academy of Management, Seattle, WA. Paper published in Proceedings.
4. Fruchterman, T.M.J & Reingold, E.M. (1991), Graph drawing by force directed placement. Software: Practice and Experience, 21(11), 1991.
5. 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.
6. 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.