Visualization of Communication Patterns in Collaborative Innovation Networks

Analysis of some W3C working groups

Peter A. Gloor[1,2], Rob Laubacher[1], Scott B.C. Dynes[2], Yan Zhao[3]

[1]MIT Center for Coordination Science, [2]Dartmouth Tuck Center for Digital Strategies, [3]Dartmouth College Dept. of Computer Science

Abstract

Collaborative Innovation Networks (COINs) are groups of self-motivated individuals from various parts of an organization or from multiple organizations, empowered by the Internet, who work together on a new idea, driven by a common vision. In this paper we report first results of a project that examines innovation networks by analyzing the e-mail archives of some W3C (WWW consortium) working groups. These groups exhibit ideal characteristics for our purpose, as they form truly global networks working together over the Internet to develop next generation technologies. We first describe the software tools we developed to visualize the temporal communication flow, which represent communication patterns as directed acyclic graphs. We then show initial results, which revealed significant variations between the communication patterns and network structures of the different groups. We were also able to identify distinctive communication patterns among group leaders, both those who were officially appointed and other who were assuming unofficial coordinating roles.

Keywords: collaborative innovation network, social network analysis, information visualization, knowledge management, collaborative applications.

1. Introduction

Collaborative Innovation Networks (COINs) are groups of self-motivated individuals from various parts of an organization or from multiple organizations, empowered by the Internet, who work together on a new idea, driven by a common vision. The COIN diagnostic project is a research effort co-located at the MIT Sloan Center for Coordination Science and the Dartmouth Tuck Center for Digital Strategies in collaboration with other research centers. The mission of the COIN diagnostic project is to understand the ways people join COINs, how COINs function and what COINs contribute to enterprises by:

1. Analyzing electronic interaction logs such as email to find COINs within organizations;

2. Identifying structural properties and parameters of successful COINs;

3. Finding the people that make a COIN successful by identifying the role profiles crucial for the success of COINs;

4. Defining a metric to measure the success of COINs;

5. Developing a framework and set of organizational guidelines that can help nurture and foster COINs within and across organizations and make them more effective.

2. Algorithm and System Architecture

Just as Google is very effective at finding pertinent documents based on linking patterns, we believe analysis of e-mail and other interaction logs of organizations will enable one to discern the structure of networks and identify core contributors. We propose a new methodology: mining computer logs such as

email archives to trace the emergence of COINs and their development over time. Our system computes and visualizes the structure of existing COINs by automatically generating a directed graph of communication flows.

of nodes is driven by the number of messages they have exchanged. The most active senders and receivers are depicted in the center of the graph.
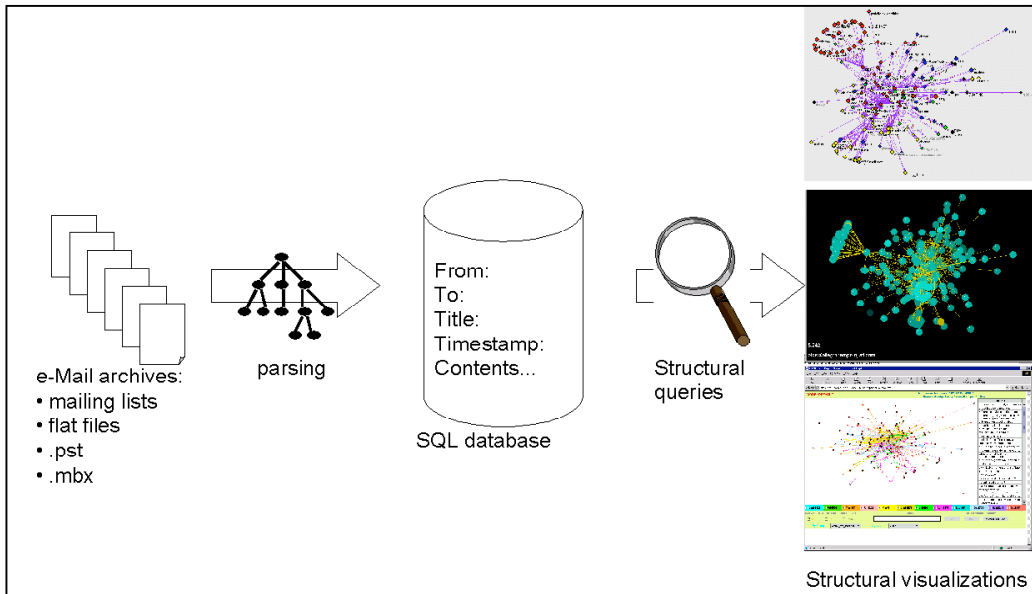

Figure 1. COIN e-mail analysis system architecture

We are implementing a flexible three-level architecture (figure 1): In the first step, the e-mail messages are parsed and stored in decomposed format in a SQL database. In the second step the database is queried to select messages sent and/or received by a group in a time period. In the third step the selected communication flows are visualized using SNA visualization tools such as Pajek (Batagelj & Mrvar, 1998) and ucinet (Borgatti et al., 1992) or our own communication flow visualization applets.

This architecture provides a testbed with high scalability and flexibility: the number of messages to be analyzed is only limited by the size of the database, and temporal queries can be run in an ad hoc way. We are also able to experiment with different visualizations of the retrieved structure.

Figure 2 contains our community visualization applet in the generic mode, where all relationships are visualized. Each node on the screen represents a person, and e-mails exchanged are visualized as arcs. The proximity
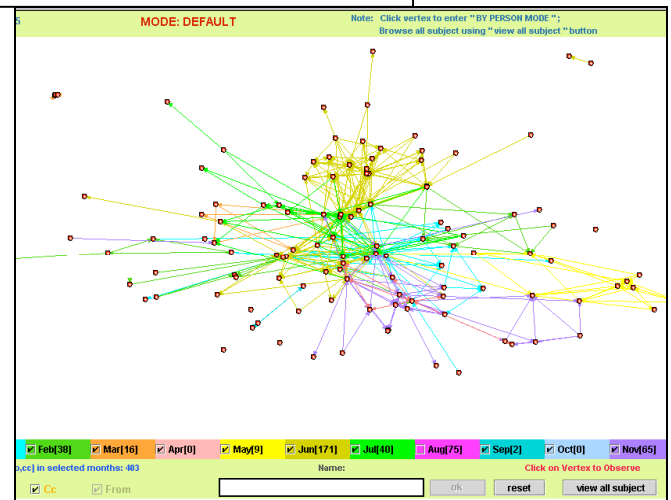

Figure 2. COIN visualization applet

The visualizations in figures 2 employ the Fruchterman-Reingold graph drawing algorithm (Fruchterman & Reingold, 1991) 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 by 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 to minimize the forces felt. In our visualizations, the attraction between connected nodes is scaled to the number of messages exchanged.
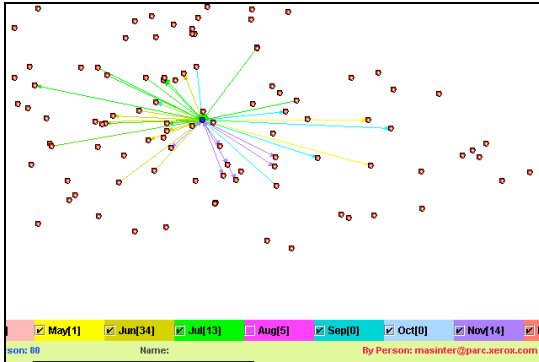

Figure 3. COIN visualization applet in personalized mode

Figure 3 illustrates the COIN visualization applet in personalized mode. Clicking on a node displays all the links going to or originating from a particular person. "To", "Cc", and "From" links can be filtered separately.
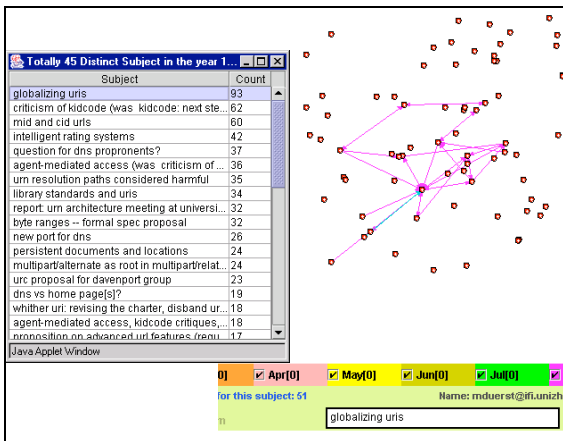

Figure 4. COIN visualization applet in subject mode

Figure 4 illustrates the COIN visualization applet in subject mode, where the message flow on a certain subject (as indicated in the "Subject" line

of the message) can be visualized. Messages can be further sub-selected by person and by month.
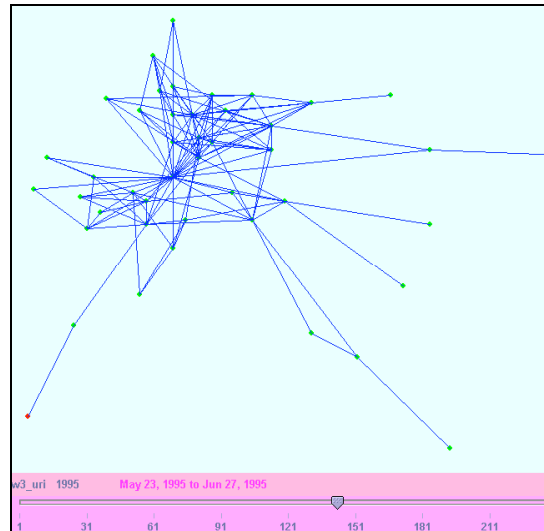

Figure 5. COIN animation applet

Figure 5 illustrates the COIN animation applet view. This applet displays a movie of the communication flow over time, where all active links in an adjustable time-window, ranging from 5 to 100 days, are shown. The user can single-step through the time-windows using a slider, or play back the communication activity of the whole year as a movie.

3. Selection of the Communities

We compared three similar communities, exhibiting strong COIN characteristics. They share the goal of furthering the development of the Web under the auspices of the WWW consortium (W3C, www.w3.org). W3C activities are generally organized into working groups. These groups, made up of representatives from W3C member organizations, the W3C core team, and invited experts, produce the bulk of W3C's results: technical reports, open source software, and services such as validation services. These groups also ensure coordination with other standards bodies and technical communities. There are currently over thirty W3C working groups. W3C is keeping a public archive of the e-mail communication of the ongoing discussions of the working groups at http://lists.w3.org/Archives/Public/. In our work

we analyzed the mailing list archives of three W3C working groups:

Group A

Group A is made up of Web enthusiasts, who contribute to the group because they are interested into the further development of the Web. There is little encouragement by large companies for their employees to participate in this group, so it is made up of academics and company researchers working on their own time and representing their own opinion. The motivation to participate is the recognition of their work by peers. We were able to analyze data from Group A for 1999 to 2003.

Group B

The goal of Group B is to develop a Web standard with high commercial value to large software companies. There are two sets of people active in this group. The first consists of software company representatives, who are usually appointed by their firm and have to represent the viewpoint of their employer. This means that firm management has a close eye on the ongoing communication of this group, as there are diverging interests of the companies sending their representatives. The second set of people in this group are consultants and academics, who participate out of personal interest or to create consulting opportunities by making a name for themselves. The standard covered by this group is fairly new, so we could only obtain data for Group B from 2002 to 2003.

Group C

Group C is also composed of two sets of people. One set consists of a core group selected by the W3C and its member firms to to discuss and further develop the technical architecture of the Web. The second set are academics and consultants who participate either because they are genuinely interested or because they want to build a reputation. The goal of the group is of somewhat lesser commercial impact than that of Group B. This group has been active from 2002 to 2003.

4. Analysis of Networks

One initial finding is that each group had between 150 and 200 members when fully active. This finding corresponds to results from past work by ethnologists that 150-200 is the largest sized group that can work together productively before fragmenting (Gladwell 2002).

We used several metrics developed by social network analysts to compare the three W3C working groups: density, betweenness centrality, and group degree centrality.

Density (Wassermann & Faust, 1994) is the proportion of potential arcs in the graph that are actually connected, a measure which can range between 0 and 1.

Centrality of an actor is a measure for its importance. The simplest measure of centrality is the number of arcs one node has to other nodes (known as the node's degree).

Group degree centrality measures the similarity in the communication pattern among different group members. It is 1 for a group where one actor communicates with all others in a star configuration, and 0 for a circle where everybody communicates with everybody.

Betweenness of actors is a measure for the interpersonal influence they have on others, by being "between" other actors. Freeman (1979) defines group *betweenness centrality* as a measure for the homogeneity of betweenness of different actors. Betweenness centrality is 1 in a star configuration, and 0 if all actors have the same degree of betweenness. Lower betweeness centrality means the communication behavior of the group members is more egalitarian.

For our three groups, we obtained the following results for 2002:

2002	Density	Group Degree Centrality	Group Between-ness Centrality
Group A	0.06	0.29	0.13
Group B	0.12	0.41	0.20
Group C	0.10	0.46	0.18

We notice that Group A exhibits significantly lower betweenness centrality and group degree centrality than the other two groups.

We can speculate that these differences may be due to the nature of the tasks being carried out by the different groups. The more focused tasks of Group B and C, for example development and gaining agreement on standards, could mean more hierarchical group interactions. Another possible reason for different group communication patterns might be the characteristics of the groups' members. The corporate researchers in Groups B and C may be more accustomed to hierarchical interactions than the university researchers in Group A, who are more used to interacting with other university researchers as peers.

Our animations suggested further differences between the groups' communication patterns. For example, the animation for one group showed the emergence of a core that communicated with high frequency among themselves, while outlying members contributed only sporadically. It appeared that a large group of peripheral members were only weakly connected to the core group. Another group exhibited more egalitarian behavior, where "everybody was talking to everybody." We plan to undertake future work which will allow us to analyze such differences in greater detail.

5. Analysis of Individual Communication Behavior

From the communication patterns, it was possible to identify a group of leaders in the COINs we analyzed. We were also able to identify contributors who had assumed leadership roles without having been officially appointed.

For Group C we looked at the message pattern of the 9 appointed leaders. We only analyzed threads that contain more than five messages. This allowed us to filter out requests for information and other insignificant exchanges. As was expected, the leaders showed up in the center of the graph (indicated as large nodes), together with other significant contributors (figure 6).

We also noticed a difference within the leadership group along two dimensions: contribution frequency (measured in the numbers of messages sent), and the extent to which their

communication was balanced between sending and receiving messages, which we measured via a simple contribution index:

$$\frac{\text{messages sent} - \text{messages received}}{\text{total of messages sent and received}}$$

This index is −1 for somebody who only receives messages, 0 for somebody who sends and receives the same number of messages, and +1 for somebody who only sends messages.
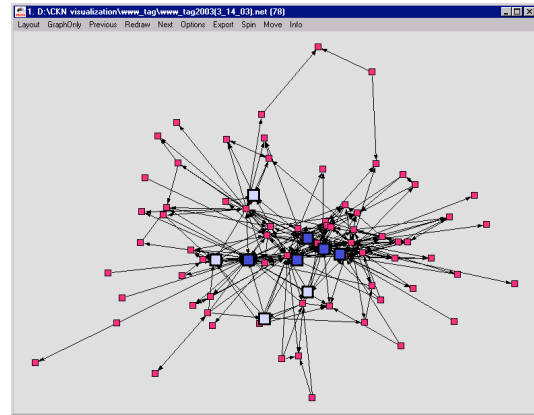

Figure 6. *Group C in 2003 with 9 leaders (leaders indicated with large squares)*

Figure 7 illustrates communication patterns for leaders in Group C. The dark rectangles in the center are the W3C leaders. The large company representatives are the lightly shaded rectangles. The leaders appointed by the W3C were among the most active contributors and also exhibited balanced communication behavior, with a contribution index close to 0. The representatives of the large IT companies participated much less in absolute numbers, and some of them sent substantially more messages than they received. There were also some participants who were even more active than the officially appointed leaders. These members appeared to be assuming a self-appointed leadership role. Preliminary analysis indicated that such unofficial leaders also emerged in other groups. We intend to undertake further analysis to identify how these unofficial leaders come to assume this role.
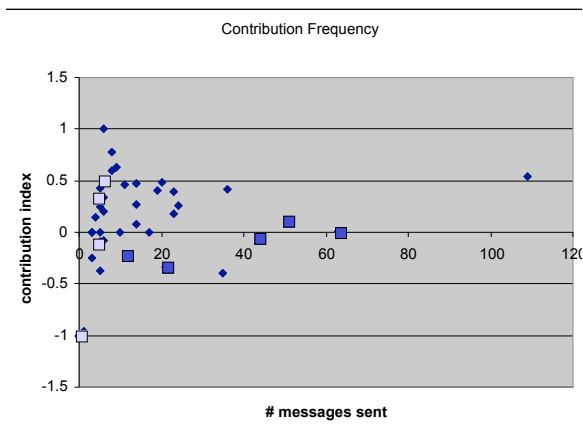
Contribution Frequency

Figure 7. *Contribution Frequency of Group C members in 2003*

.

6. Related Work

Other researchers have analyzed email communication flow to study network structure. In an early study, undertaken well before usage of the Internet became widespread, Freeman (1979) looked at how the use of email assisted in the development of acquaintanceships and friendships between thirty-two SNA researchers.

Guimera et al. (2002) mapped the structure of a university network by automatically analyzing the e-mail log and found that the network exhibited a self-similar structure. Ebel et al. (2002) also analyzed the e-mail logs of Kiel Unversity and found it to exhibit a scale-free topology.

Tyler, Wilkinson, et al (2002) describe a project which mined the e-mail archive of 485 employees at HP Labs. By partitioning the total graph with a clustering algorithm proposed by Girvan and Newman (2001), the HP Labs researchers were able to identify 66 distinct communities, based on the frequency of the message exchange between individuals. Correlation between actual teams and organizational units and the clusters derived by e-mail analysis was surprisingly high. The HP Labs team also identified leaders from the e-mail traffic, which mostly corresponded to widely recognized leaders within the organization.

7. Conclusions and Further Work

We have identified four areas where our work can be applied

1. By locating COINs, organizations can learn about innovations which are underway. This enables them to spot hidden business opportunities and also cut the time to market for new inventions.

2. By supporting hidden COINs and making them transparent, organizations can become more efficient in working together. They can better identify their knowledge sources and streamline communication processes.

3. Because key contributors can be identified through transparent COINs, organizations have a better chance to identify and reward leaders and important collaborators.

4. By making the communication flow transparent, a more open working environment can be created, generating additional trust among its members.

The next step in our work will be to improve the temporal visualization and analytic capabilities of our tools. We will then undertake further analysis of these W3C and gather date from other kinds of networks as well. We hope to gain new insights into network evolution and examine member roles in more detail.

Acknowledgements

References

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

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

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

Freeman, L.C. Centrality in social networks: I. Conceptual clarification. Social Networks, 1979, 1, 215-239.

Freeman, L.C. Freeman, S.C. 1980. A semi-visible college: Structural effects of seven months of EIES participation by a social networks community. In Henderson, M.M. MacNaughton, M.J. (eds) *Electronic Communication: Technology and Impacts*, 77-85. AAAS Symposium 52. Washington DC: American Association for the Advancement of Science.

Gloor, P. Laubacher, R. Dynes, S. Zhao, Y. Visualizing Interaction Patterns in Collaborative Knowledge Networks in Medical Applications. Proc HCII 2003, Crete, June 25-27, 2003.

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

Gladwell, Malcolm. 2002. The Tipping Point: How Little Things Can Make a Big Difference. Back Bay Books

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 2002.

Wasserman , S., Faust, K. 1994. *Social Network Analysis : Methods and Applications*. Cambridge University Press.

Thomas M. J. Fruchterman and Edward M. Reingold. *Graph drawing by force directed placement*. Software: Practice and Experience, 21(11), 1991.

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

Wasserman , S., Faust, K. 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press.