Two-mode Social Network Analysis as Exploratory Tool for CSCW: Technology Adoption and Use

Steve Abrams

sabrams@ics.uci.edu

Social network analysis is typically used to understand relationships between people and people, what Wasserman and Faust term "one-mode" networks. This is not surprising, given its historical roots in sociology and anthropology. CSCW, however, is not only concerned with people, but also with the things people do. As such, two-mode (or greater) network analytic methods may be more useful in developing understanding of issues of interest to our community. In this paper, I apply such methods to explore the adoption and use of a technology system in its early deployment.

Since January 2002, I have been studying collaborative, conceptual space mission design work at NASA (Mark 2002) as a research assistant for Dr. Gloria Mark. In June 2002, we were fortunate to be able to observe the inaugural design session of a large-scale, conceptual mission design team, distributed among three NASA field centers and one non-NASA site (Mark, Abrams et al. 2003). A design "session" here refers to a set of three, 3-hour periods of real-time interaction, conducted over a 3- or 4- day period. In 2004, we observed this large-scale team in 4 similar design sessions. In all sessions, a videoteleconferencing service provided the main audio and video channels between sites. In addition, facilities were provided to enable individuals to conduct conversations (called "sidebars" by participants) with others at remote sites, as well as within a site. These sidebar discussions are a central feature of the design process as this is where much coordination occurs and many aspects of the design are scoped, problems identified, and solutions negotiated before being communicated to the larger team.

The Site 1 team was the largest team, comprised of approximately 25 engineers, scientists, and managers and 2-3 infrastructure support personnel. The team at Site 2 had 9 engineers and 3 infrastructure people in 2002 and 3 engineers/2 infrastructure people in 2004. The Site 3 team had 9 engineers/2 infrastructure people in 2002 and 6 engineers/2 infrastructure in 2004.

For the 2002 design session, we were allowed to record distributed sidebar discussions as well as 3.8 hours of individual conversations (captured via wireless microphones) for 20 team members (total of 76 hours) at one site. Colocated sidebars were numerous, spontaneously formed, and typically brief (average: 75 sec). Distributed sidebars were assigned by the collaboration facilitator, required the assistance of at least one infrastructure support person to initiate, averaged almost two minutes of idle time before work could commence, and was much longer (average: 642 sec).

As a partial result of our analysis of the 2002 session (Mark, Abrams et al. 2003) the NASA site leading the 2004 sessions developed a new technology system to enable distributed sidebar conversations without the problems we identified. Called the "Pvoice system" (PV), it provides a Web-based interface to a telephone exchange and enables individuals to easily know who is accessible at any particular moment, to quickly establish a sidebar conversation, and to monitor multiple sidebar conversations simultaneously (see Fig. 1). The fundamental goals, as articulated by the system's designer, were to reduce the amount of time required to establish a distributed sidebar and to make distributed interaction more similar to colocated interaction.

We weren't allowed to record the actual conversations in 2004, though the Pvoice system enabled us to listen to them and make notes. The setup time – the delay between clicking on the interface to hearing the remote end – was reduced to less than one second. More articulation work, compared to the distributed sidebars in 2002, was observed to occur via this system. Though technical problems created widespread frustration during the first 2004 session in which it was deployed, subjects reported considerable satisfaction with the system in the three subsequent sessions.

Qualitatively, the goals appeared to have been met but the first research question, in applying a social network analytic methodology, was whether or not the team interaction using the voice system in 2004 was more similar to the distributed interaction of 2002, the collocated interaction of 2002, or something intermediate. A second research question was to determine how the actual use evolved over the three subsequent sessions.

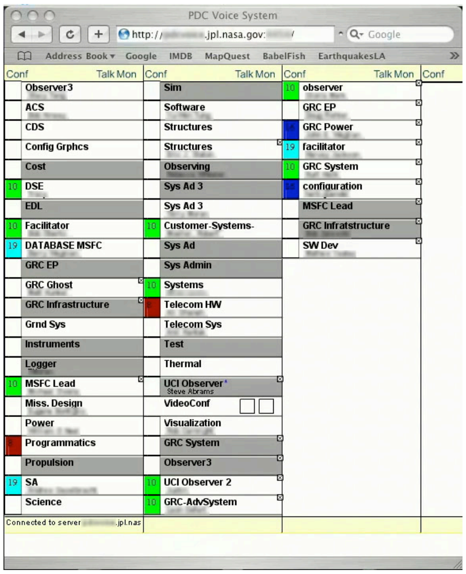


Figure 1 Screenshot of Web GUI to Pvoice system showing four sidebars (colors).

A third research question emerged in the course of observations of sessions in 2004. At the site with the most team members (approx. 30), there appeared to be a zone extending approximately 20 feet from the front of the room where the study facilitator led the session and the video view of the remote sites was displayed. Team members within this zone were observed to be much more attentive of the facilitator than those outside the zone. This led to the question of whether or not team members within the zone were also more attentive to the remote sites via the voice system. It should be noted here that team members at this site sat at fixed, specific workstations (according to their area of expertise) within the room. Although it might be expected that team members relevant to the current mission under design might be located "close to the action" at the front, hence in the zone, that was not the case. At least three seats (Power, Structures, and Configuration) outside the zone were relevant to the mission and had counterparts at other sites with whom they could reasonably be expected to communicate using the voice system. But did they?

My ongoing analysis is based on screen video captures of the changing state of the Web interface (see Fig. 1), recorded at one frame per second, of the four distributed design sessions in which this system was deployed. These videos were manually coded into a database system by viewing the video in fast forward until a change in the view was detected. Each change recorded was coded for time of occurrence, the subject making the change, the action performed (e.g. "Creates Sidebar," "Joins Sidebar," "Leaves Sidebar," "Switches to Another Sidebar), and a unique sidebar ID relevant to the action. Activity timelines were generated for each individual actor and sidebar and these were reviewed to locate actions that were missed in the initial coding. Technical difficulties prevented the coding of Days 1 and 3 of the second distributed session.

Pajek (Batagelj and Mrvar 1998) network files were generated from this coding database, for both the 2002 and 2004 sessions, and analyzed. A two-mode representation – actors and sidebars – was chosen to enable analysis of both modes and edges were weighted by the duration (in seconds) of an actor's participation in a sidebar. Since sidebar content couldn't be recorded, a "tie" (edge) in this network represents an "opportunity to overhear or provide information to other team members," instead of a characterization of the conversation. Centrality, with this definition of a tie, is more reflective of awareness of, and engagement with, the state of the evolving design interaction rather than any specific communicative activity. Centrality measures for 1-mode networks offer insights at the individual actor and overall network levels. Two-mode networks, because they associate subsets of actors with subsets of events, afford an additional level of inquiry intermediate between actor and network (Faust 1997).

For the first research question, the network–level measure of network betweenness centrality (Wasserman and Faust 1994) was chosen. Network betweenness centrality (NBC) is widely understood and used, although it is arguably not the optimal metric. As Wasserman and Faust note, all betweenness centrality (BC) measures assume that all geodesics in a network are equally likely to occur. In organized human activity, interaction is more structured than this. Information centrality, which allows for weighting edges according to the probability of interaction, would be the optimal metric here, but the method of determining edge weight probabilities isn't clear. Another alternative metric, planned for future analysis, is eigenvector centrality in which important sidebars are those in which important actors participate and actor importance is dependent on the importance of the sidebars in which they participate.

NBC here is calculated for both the 2-mode networks, as well as for the 1-mode networks generated from them. Such generated networks are "co-occurrence," or "co-participation" in this case, networks wherein the ties between nodes of one mode (e.g. actors) are linked through co-occurrence via a node in the other mode (e.g. sidebars). As shown in Table 1, the NBC of the Pvoice system is not significantly different from the NBC of the collocated interaction (p=0.8625) and is significantly different from the NBC of the distributed interaction (p=0.0450).

For the second research question, visual inspection is helpful. As above, the collocated and distributed sidebars from the 2002 session serve as a baseline for comparison. In 2002, the collocated interaction was extensive (498 sidebars, totaling 9.2 hours of conversation), but disjointed (see Appendix, 2002.A.c.2). Most sidebars were dyadic (degree centrality 2.20 but, visually, obvious in the black and grey sidebar nodes connecting to only two colored nodes), occurring between individuals doing similar kinds of work (same color vertices). The distributed interaction in 2002 was even more disjointed, with a subset of 5 actors having very high betweenness values and the remainder having very low betweenness values. In these sidebars, the actors with high betweenness have influence over how information is shared within the network; inspection of the timestamps in the sidebar labels (omitted for clarity in the Appendix graphs) give a sense of the direction of potential information flow (bearing in mind the earlier definition of a tie in this study). The difference between these two modes of interaction, in 2002, is reflected in their very different network betweenness measures: averaging 0.4080 for collocated sidebars and 0.1584 for distributed ones. In comparison, use of the Pvoice system results in an average network betweenness of 0.3904.

	Betweenness Network	**Betweenness Actors**	**Betweenness Sidebars**
2002.A2.c.2	0.4343	0.5876	0.1103
2002.A2.d.2	0.1756	0.1769	0.1858
2002.A3.c.2	0.3817	0.4782	0.0724
2002.A3.d.2	0.1412	0.1515	0.1632
Avg. Coloc.	***0.4080***	***0.533***	***0.091***
Avg. Distrib.	***0.1584***	***0.164***	***0.175***
2004.A3.PV	0.2296	0.1460	0.1438
2004.B1.PV	0.2586	0.2564	0.1420
2004.B2.PV	0.3976	0.1230	0.4561
2004.B3.PV	0.2632	0.0843	0.1783
2004.C2.PV	0.5631	0.0857	0.1997
2004.D1.PV	0.4490	0.1672	0.1395
2004.D2.PV	0.4019	0.0555	0.2272
2004.D3.PV	0.5606	0.0992	0.4109
Avg. PV	***0.3904***	***0.1272***	***0.2372***
p-Values			
PV × Coloc.	**0.8625**	0.0000	**0.1562**
PV × Distrib.	0.0450	**0.4550**	**0.5193**

Table 1. Network-level measures of betweenness (both 2-mode and generated 1-mode) and density (generated 1-mode). P-values are determined by a two-tailed t-test (alpha=0.05) comparing the mean betweenness measures of the Pvoice system with the collocated and distributed interactions. Bold-faced p-values indicate where mean NBCs are equivalent.

In comparing the visualizations of PV use (see Appendix) over its deployment,an interesting feature was identified. In early deployment, several sidebars are observed with similar node betweenness centralities. This reflects individual team members using the voice system to initiate, conduct, and close sidebar discussions in similar ways. Over time, however, two sidebars grow in betweenness to dominate the network. These sidebars are established early in each day's sub-session and, in almost all cases, they persist throughout the sub-session.

These "standing" sidebars were initiated and perpetuated, in all cases, by two different types of team members. One obvious type is a sidebar dedicated to the infrastructure support personnel, providing a common communication channel to quickly resolve problems, just as had been formally organized in 2002. In contrast to that inaugural session, however, this informal use was emergent over time and not through any observed or reported inter-site agreement. The second type was initiated daily, and primarily attended by, team members whose assigned role required a systems-level focus (in contrast to the subsystem orientation of the rest).

While the systems sidebar made sense, given the continuously iterating design state, the infrastructure sidebar did not. From our ethnographic observations, we knew the infrastructure people generally sat in front of a computer during these sessions and they were always accessible via the voice system (through wearing wireless headsets). From our observation notes, I found that these sidebars, in contrast to the rest, were silent much of the time. Why maintain a connection to a silent, standing sidebar when one is a mouse-click away from participating in a sidebar?

A possible explanation is that team members began using this communication system as a social space, as in the Thunderwire system (Ackerman, Starr et al. 1997) for the infrastructure and systems people and an awareness system by other team members. Although a temporal analysis is beyond the scope of this paper, the coding of PV data made it apparent that some of these other team members repeatedly check the status of these standing sidebars to learn whether anything is happening behind the scenes. The pattern was to connect to a sidebar and quickly (within ~5 seconds) disconnect and move on to check another sidebar.

One negative impact of these standing sidebars emerged in which non-systems/ infrastructure team members would sometimes see someone they wanted to talk to and, not realizing they were joining an existing sidebar, would begin to talk about things irrelevant (and distracting) to the others listening to that sidebar. Another negative impact emerged when team members often removed their telephone headsets without disconnecting from a standing – but silent – sidebar. According to the web interface, they appeared to be accessible when, in fact, they weren't. This was also observed with Thunderwire and described as "the most troublesome situations" (Ackerman, Starr et al. 1997) in the use of the system. We observed several instances of team members trying to initiate discussion with another who was apparently accessible and failing. In at least five instances during the last 2004 session, remote team members contacted another person at a site, via the Pvoice system, and asked them to track the other person down and get them to put their headsets back on.

For the third research question, visualization also proved useful. A separate Pajek partition file (*.clu) was generated for each network to identify those within the zone at the site in question and those outside the zone. Each network is drawn, in Pajek, with its associated partition file. In the figures shown (see Appendix, 2004.* graphs), the actors of interest are the yellow ones (in the zone) and the green ones (outside the zone). Sidebars are colored according to whether they involved only team members at one site (grey) or team members from multiple sites (black). The first step, in analyzing each network, is to confirm that yellow and green nodes are both present (in the first column, drawn with equal node sizes). The second step is to count the number of greys and blacks to which each yellow and green is tied. The final step is to re-draw the network with a betweenness centrality measure (second column of graphs in Appendix table) and see whether yellows dominate the greens in aggregate size or not. In each network, the yellows do dominate the greens.

In Appendix graph 2004.A3, there are approximately equal numbers and sizes of the yellow and green nodes, but most of their betweenness values are due to their connections to the grey nodes at the periphery. In other words, participation in sidebars is approximately equal, but those within the zone are primarily engaged in inter-site sidebars while those outside the zone are primarily engaged in intra-site sidebars. A similar situation is seen in graph 2004.B1. Exceptions are seen in graphs 2004.B2, 2004.D2 and 2004.D3, where the same team member, “1struct,” is seen interacting with only black nodes. In 2004.D3, there is an additional exception. One explanation for the increase in later exceptions could be that team members outside the zone are learning the benefits of monitoring the inter-site, standing sidebars form team members in the zone.

I've demonstrated some benefits of a basic application of two-mode network analytic methods to a CSCW-oriented situation, but more is possible. I've found it useful to explore a Galois lattice representation

(Wasserman and Faust 1994) of a binary two-mode network (specific observed actors/events v. categories generated via Grounded Theory) with an Open Source tool called "Concept Explorer" <http://sourceforge.net/projects/conexp/ >. In the context of this paper, a Galois lattice representation can identify superset-subset relationships between subgroups of actors and sidebars and can generate association rules characterizing these relations ships, e.g. “in 8 cases, 88% of the time, actor1 and actor3 were found in association with actor7, actor9, and actor10. Such information could be used to weight edge probabilities in the application of information centrality, for example, to an anlysis.

For valued, two-mode network data, correspondence analysis is suggested (Wasserman and Faust 1994) for exploring event profiles of actors and vice-versa. For example, to refine the analysis above, I'll use the existing timestamped network files to generate second-by-second Pajek time networks which I'll analyze using correspondence analysis. This will change the unit of analysis from "actors participating in a sidebar" to "actors participating in a sidebar *at the same second*." Similarly, a correspondence analysis of betweenness measures might identify groups of users who use the Pvoice system in similar ways over time. Although UCINET (Borgatti, Everett et al. 2004) performs this analysis, I've found a Microsoft Excel add-on named "XLSTAT-Pro" <http://www.xlstat.com/products-xlstat-pro.htm > much easier to use and, when used with a related add-in – XLSTAT-3Dplot <http://www.xlstat.com/products-xlstat-3DPlot.htm> – visualization and exploration becomes enjoyable.

Ackerman, M. S., B. Starr, et al. (1997). "Hanging on the ‘wire: a field study of an audio-only media space." ACM Transactions on Computer-Human Interaction **4**(1): 39-66

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

Borgatti, S., M. Everett, et al. (2004). UCINET 6, Analytic Technologies.

Faust, K. (1997). "Centrality in Affiliation Networks." Social Networks **19**: 157-191.

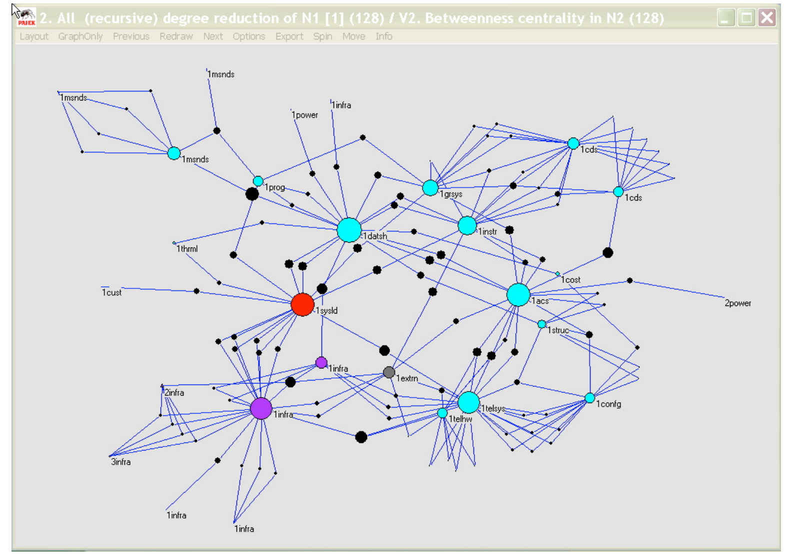
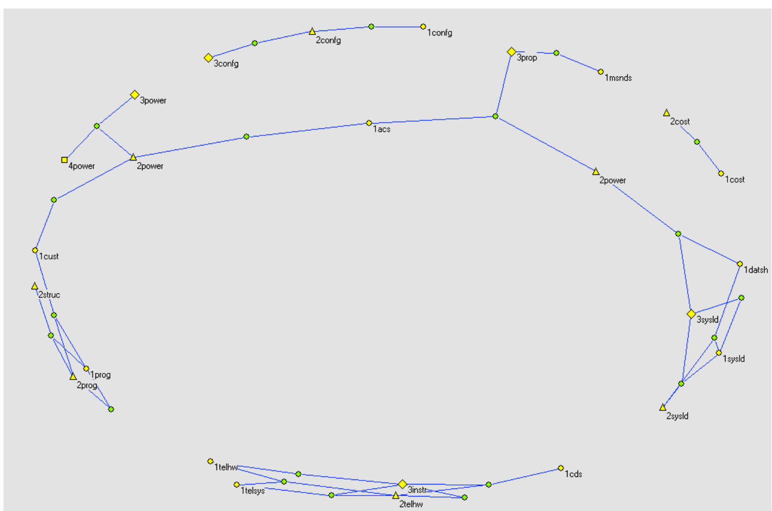
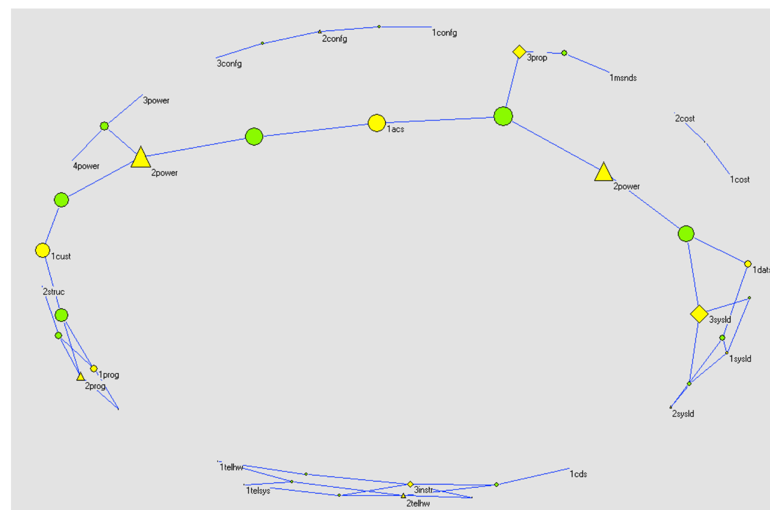
Mark, G. (2002). "Extreme collaboration." Communications of the ACM **45**(6): 89 - 93.

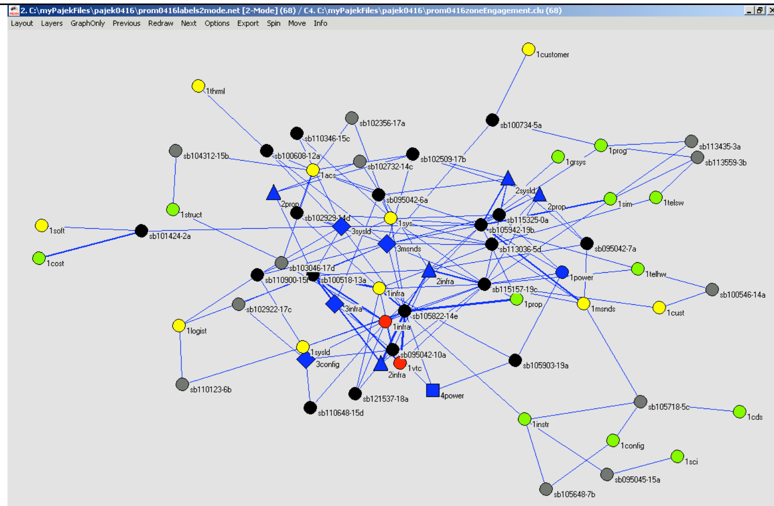
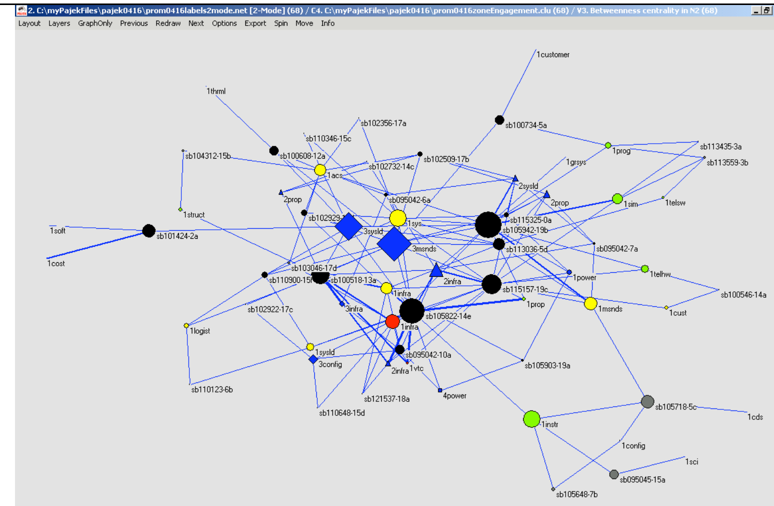
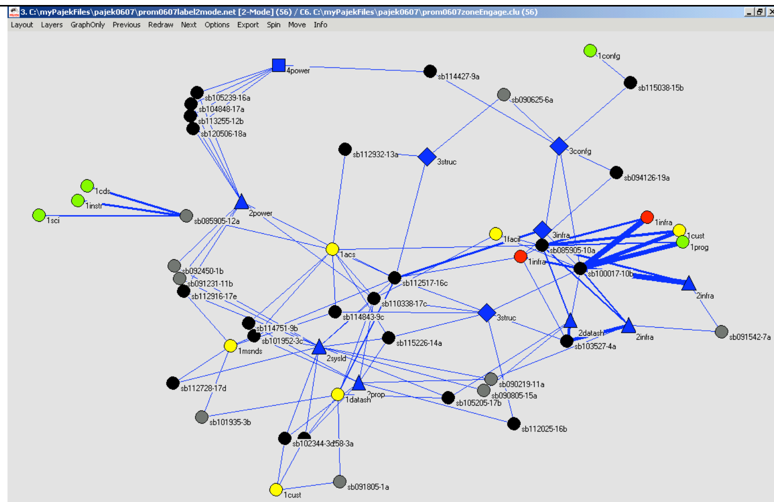
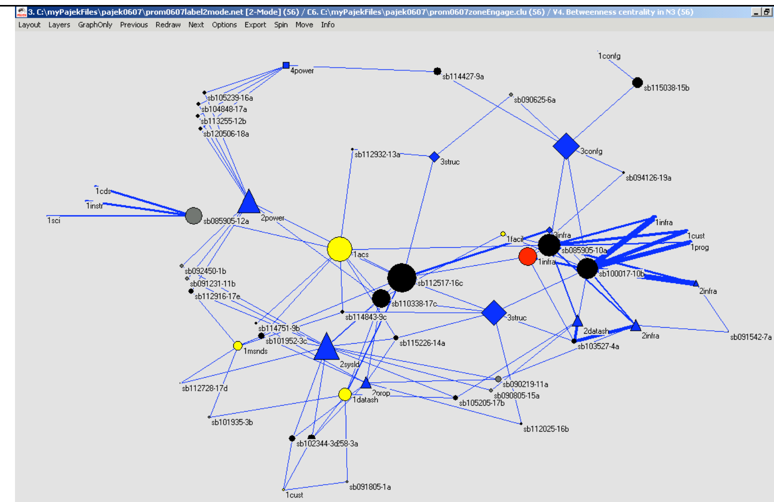
Mark, G., S. Abrams, et al. (2003). "Group-to-Group Distance Collaboration: Examining the “Space Between”." Proceedings of the 8th European Conference of Computer-supported Cooperative Work (ECSCW’03), Helsinki, Finland: 99 - 118.

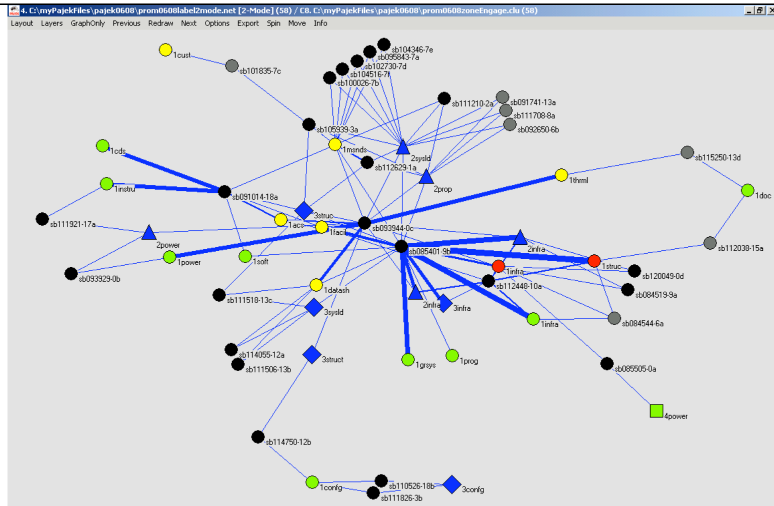
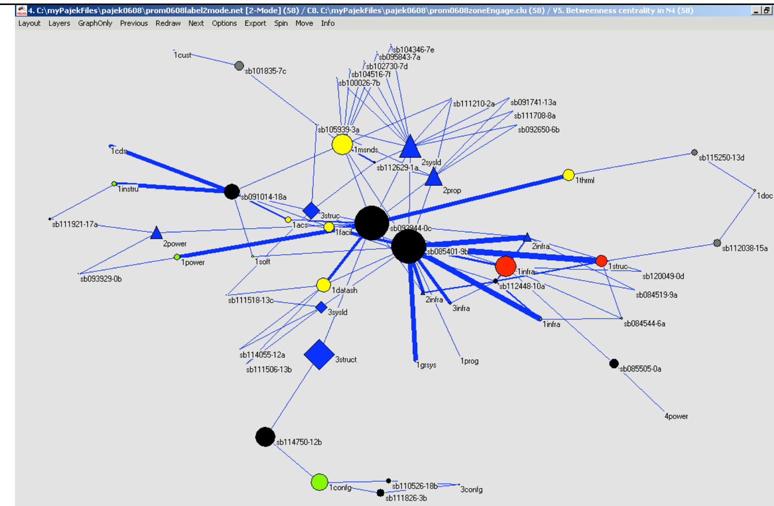
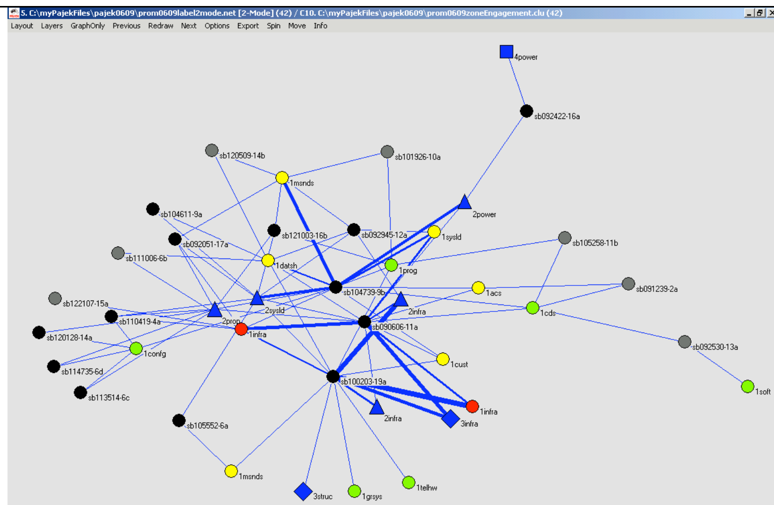
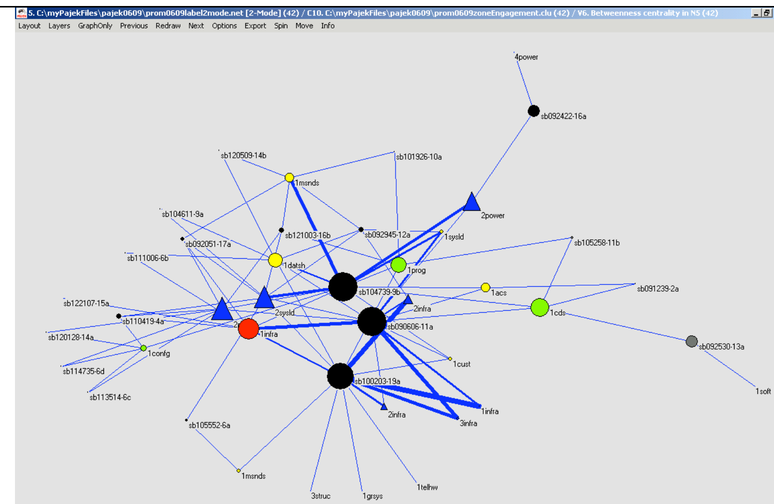
Wasserman, S. and K. Faust (1994). Social Network Analysis : Methods and Applications, Cambridge University Press.

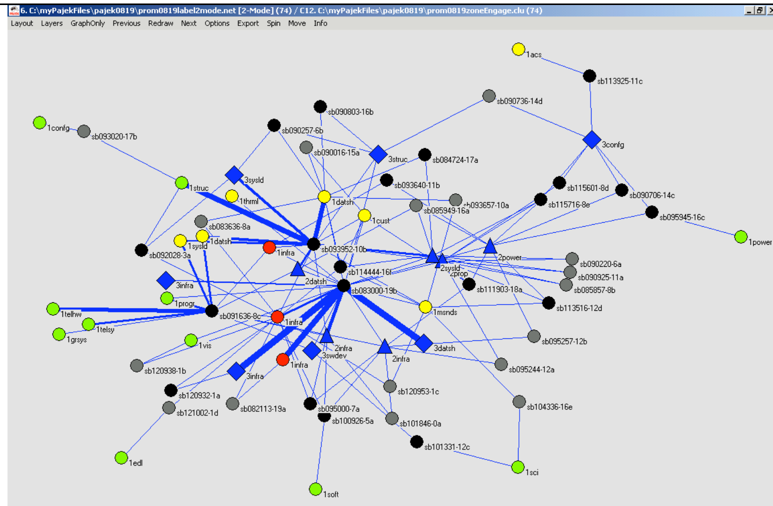
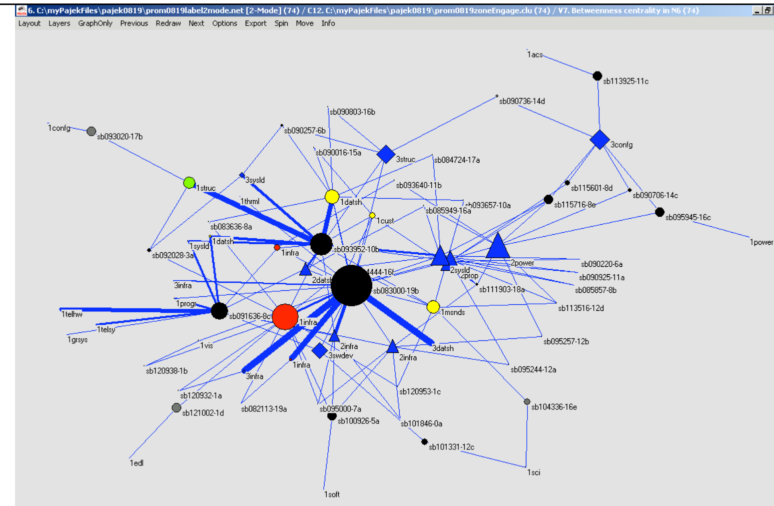
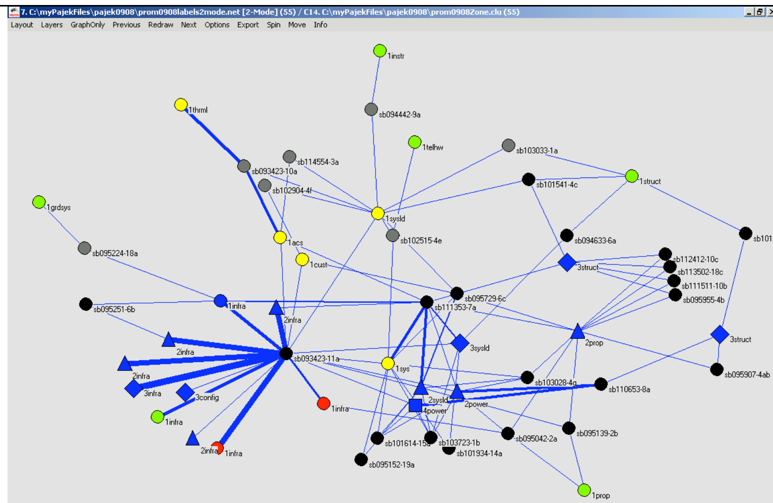
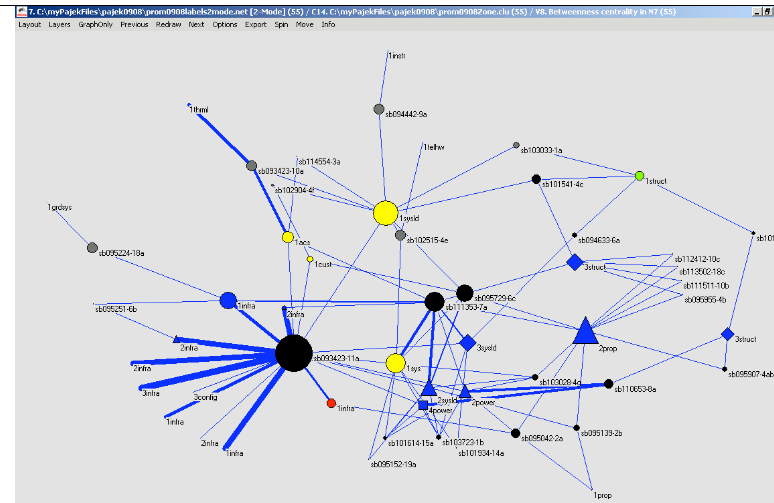
APPENDIX

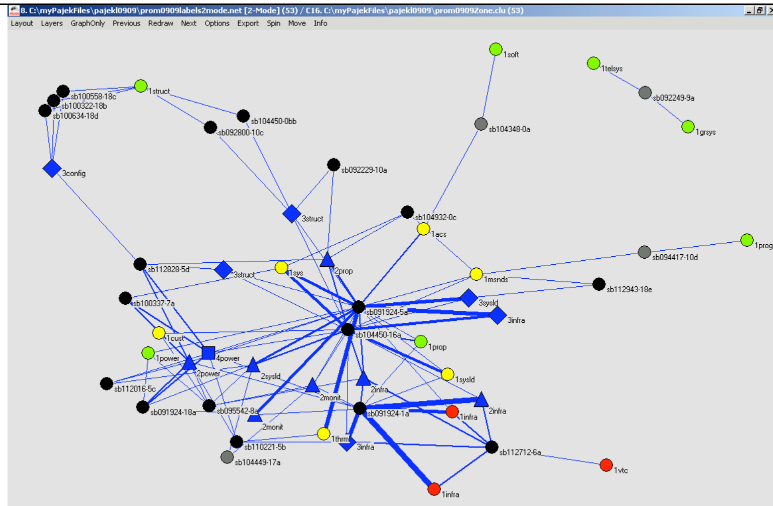
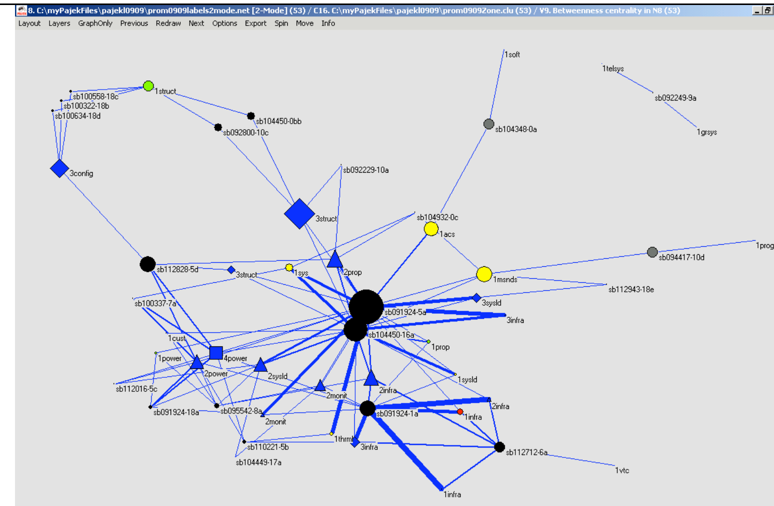
All figures below were produced using Pajek's Fruchterman-Reingold 2D layout algorithm.

	Network with equal node sizing	Network with nodes sized by their betweenness centrality
2002 Coloc	TBD – 2002.A.c.1	2002.A.c.2 
2002 Distrib	2002.A.d.1 	2002.A.d.2 

2004.A3 **Day 4 (of Pvoice system deployment)**		
2004.B1 **Day 56**		

2004.B2 **Day 56**		
2004.B3 **Day 57**		

2004.C2 **Day 127**		
2004.D1 **Day 146**		

2004.D2 **Day 147**		
2004.D3 **Day 148**