

Visualizations of Collaborative Information for End-Users

Rebecca Xiong*

Microsoft Research
One Microsoft Way
Redmond, WA 98052

Marc A. Smith

University of California, Los Angeles
Department of Sociology
405 Hilgard Ave
Los Angeles, CA 90024 USA
+1 213 342 9834
smithm@ucla.edu

Steven M. Drucker

Microsoft Research
One Microsoft Way
Redmond, WA 98052
+1-425-703-5850
sdrucker@microsoft.com

ABSTRACT

There is a growing need for methods and tools to illuminate the social contexts of interaction environments created by the World Wide Web, Usenet newsgroups, email lists, and other network interaction media. We present here a framework for creating visualizations of the social connections created in and through network interaction media. Using graph-drawing methods, visualizations can be created for a range of systems that link people to people and people to objects through networks. As an example, we present an application of our methods to the Usenet to illustrate how visualization can improve existing systems. We propose that users of network interaction media can benefit from visualizations that illuminate the interaction context generated by the rich interconnections between groups, conversations, and people in these media.

KEYWORDS

Information visualization, interaction context, collaborative filtering, network diagrams, graph drawing, supporting on-line interaction, social network analysis

INTRODUCTION

Network interaction media like email and email lists, conferencing systems and bulletin boards, chats, multi-user dungeons (MUDs), multi-user games and graphical worlds are used singly and in concert to create an exploding number of social cyberspaces that allow groups of people to gather on-line and interact. Within these gathering points millions of social groups have emerged and formed around nearly every conceivable topic, interest, and issue.

While network interaction media are increasingly popular, there are a number of problems facing their users. The interaction context, or information about the kind of space, group and activity taking place, is often missing or ambiguous in the spaces created by these systems. This makes finding groups to participate in and people to interact with difficult. The history and structure of groups

often become visible only after extended participation. As a result, many of these systems remain complex and confusingly tangled spaces in which it is difficult to navigate and participate.

We are interested in ways of untangling these spaces by clarifying the interconnections that structure them. In the following we will explore ways of illuminating collaborative information, the interconnection between people and objects. We describe four different views of this information and the advantage of each using a simple example of a book purchase database. We then discuss related work in the area of social and collaborative filtering and information visualization. We apply these principles to the Usenet to reveal patterns and connections that are difficult or impossible to see in existing interfaces. These visualizations can help to answer central user questions about the relationships among and within newsgroups, identify context and patterns of activity, and enable them to make more effective use of these systems.

Four views of collaborative information

Collaborative information is the product of relationships between people and objects, which could be topics of discussion, database records, books of interest, Web pages, and other forms of data. To use books as an example, connections between people and books can be based simply on who buys what book. This information is commonly used through collaborative filtering [3, 12] techniques to recommend books to other users, for example, at the Amazon.com Web Site [1].

These kinds of relationships can be represented in four different views:

- **Object-to-Object** – displays relationships between objects created by user activity
- **Objects-to-People** – starting from objects, see who are the people relating them
- **People-to-Objects** – starting from people, see which objects connect them

* Current address: MIT Lab for Computer Science, becca@graphics.lcs.mit.edu.

- **People-to-People** – how people relate to each other based on interest in similar books

While simple sets of relationships can be presented in tabular form the relationships created in most collaborative information systems rapidly becomes complex and impossible to comprehend unless visualized. We present a tabular representation (table 1) as contrast.

PERSON	BOOKS PURCHASED
PersonA	Book1, Book2, Book3
PersonB	Book1, Book2, Book3
PersonC	Book1, Book4
PersonD	Book4, Book5
PersonE	Book4, Book5

Table 1. Tabular presentation of book purchases

These relationships are best illustrated using network diagrams, or graphs, as shown in Figures 1-4. In these figures, rectangular nodes represent objects (in this case books) and circular nodes represent people. The lines connecting them represent various relationships. Examining each in turn:



Figure 1. The Object-to-Object Diagram

The **Object-to-Object Diagram** shows the relationship among objects based on user preferences. In Figure 1 for example, the nodes represent books, and the links represent how often the same user purchases both books (the more often, the stronger the connection, and the shorter the link). By looking at the diagram, one can infer that book1 and book2 are more closely related to each other than to book5. In fact, book1, book2, and book3 forms a cluster because they are all connected to each other. They may represent a set of books on similar topic.

Using this information, users can see how a set of books may be related to each other and find other books on similar topics based on books they already know.

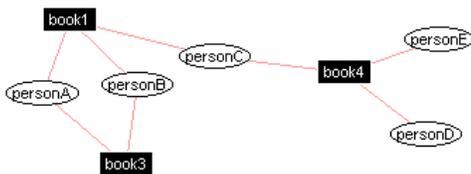


Figure 2. The Objects-to-People Diagram

The **Objects-to-People Diagram** in Figure 2 makes explicit the intermediary links that connected objects of interest from the previous figure 1. The rectangular nodes represent selected books and the circular nodes represent the people who have purchased these books, which creates

the link between book and person and other books. By looking at the diagram, one can infer that personA and personB are both interested in book2, and may share interests in a similar set of books, whereas personC is interested in somewhat different books.

If a user is also interested in book1, then the user might share interests similar to personA, personB, and person C. But if the user is more interested in book4, than personC, personD, and personE might be people of similar interest. This can be a way for users to seek out people with similar interests in books with whom to interact.

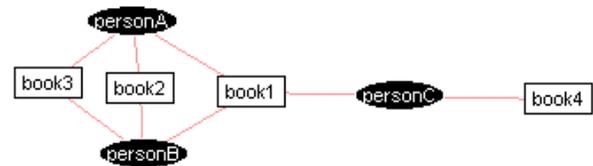


Figure 4. The People-to-Objects Diagram

We can also go the opposite way. The **People-to-Objects Diagram** shows the relationships from people to objects. For any set of people a user expresses interest in we can display how different objects relate them to one another, as shown in Figure 4. The black nodes represent selected people and the links connect them to the books they have purchased. By looking at the diagram, one can infer that book1, book2, or book3 is closer to personA's interest than book4. So if a user's interest is more aligned with personA's interest, he might be interested in looking at the first 3 books.

If a user knows a group of people who share her interest, and another group which does not share her interest, this view can help separate out the books based on the preference of people whom she knows.

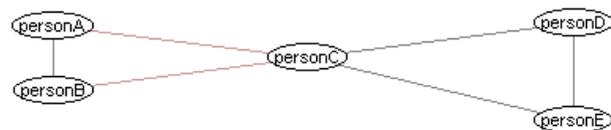


Figure 3. The People-to-People Diagram

Finally, we are interested in showing the relationship between people themselves based on their preference for objects, in the **People-to-People Diagram**. In Figure 3 for example, the nodes represent people, and the links represent how often each person purchases the same books (the more often, the stronger the connection, and the shorter the link). So by looking at the diagram, one can infer that personA, personB are more closely related to each other than to others. In fact, personA and personB or personD and personE may be a pair of people who share similar book purchases, and may share a lot of interests in common. PersonC, on the other hand, seems to share diverse interests with a lot of people.

This information shows the relationships among people, giving a nice peek into the social interaction context. So instead of seeing a group of people and not knowing their connections, and beginning any interaction blindly, users can now have an explicit view of people's connections, and use it to locate people to contact.

Notice the parallel and dual nature between the diagrams. The Object-to-Object Diagram uses the purchases of users to separate out the relationships between books, whereas the People-to-Person Diagram uses shared book purchases to separate out the relationships between people. Also, the Objects-to-People Diagram uses known relationships between selected objects to separate out the people, whereas the People-to-Objects Diagram uses known relationship between selected people to separate out the books. Object-to-Object and People-to-Object Diagrams both help users to locate objects. While the People-to-People and Objects-to-People views both help to locate other people

So far, we have presented the four views in the most basic format. Each of these can be enhanced by overlaying additional data using colors, sizes of nodes etc. These views can be integrated for more effective navigation of the underlying information.

Related Work

Our work is greatly influenced by recent developments in collaborative filtering, as well as work in more established fields such as social network analysis and information visualization.

Social network analysis is the study of the structure and dynamics of relationships between people. Originating with Jacob Moreno's sociometric research [9] in the 1920's social network analysis has demonstrated the dense interconnections between people that allow any two people on Earth to connect through six or fewer intermediaries [8]. Network analysis has gone on to illuminate the empirical structure of local communities [16] and inter-office patterns of communication [7].

Collaborative filtering originated from the Information Tapestry Project at Xerox PARC [3] for e-mail filtering, and has been applied to other domains such as Usenet postings [12]. By using recommendations from other users for objects, a program can make recommendation for new users.

Information visualization techniques are typically used for the visualization of large amounts of commercial, financial, scientific and technical data for data analysis and mining. Its main audience has been data analysts who have the time and effort to sieve through data from different angles. We are more interested in using visualization as a simple interface to help the average user use collaborative information.

Our work makes extensive use of networks, or directed and undirected graphs. Networks are used extensively for a

narrow range of applications, most notably in telecommunications management and social network analysis. Unlike trees, networks may contain cycles, which are important for showing complex, non-hierarchical information. To limit the amount of information shown at once, and give user control over what they see, we have also used traditional information visualization methods like filtering. [10].

Our work can be seen as a new way to visualize collaborative filtering information through the Object-to-Object Diagram, which shows how close objects are to each other based on user preferences. Alternatively, our work can be used as way to perform visual collaborative filtering, by directly looking at other people's preferences through the People-to-Object Diagram.

There has been some recent work in simple visualizations of complex relationships that is close to our work. Judith Donath's "Visual Who" displays the interconnections between people based on mailing list membership [2]. A set of mailing lists are selected as anchors, the names of people are then placed around the anchors based on how similar they are to the members of these lists. Dynamic animation of changes in information and user interaction made the display easy to use. The visualization is similar to our Object-to-People Diagram, which lays out people based on closeness to objects.

Loren Terveen and Will Hill's Auditorium Visualization for Inter-site Clan Graphs displays the relative importance of different Web sites based on a user's query [14]. Web sites are laid out in a series of concentric semicircles based on importance. They used dynamic ordering of Web sites to allow user customization based on interest. This is similar in spirit to our Object-to-Object view with a weighted display, as described later.

RazorFish's [11] design for an interface to the Smithsonian Institution catalog and ThinkMap's [15] Visual Thesaurus both use networks to display the interconnections between objects, though this information is not generated through people. Henry Kautz design for ReferralWeb, on the other hand, illustrates the interconnections between people, without showing the objects.

USENET and NETSCAN

As an illustration of our approach, we visualize relationships between groups, posters, and threads in the Usenet. Usenet is often seen as an example of the tragedy of the commons [4, 6], a public resource that has been ruined by overuse and pollution. Spam, off-topic and poorly targeted messages, huge message volume and limited tools for generating an overview of activity have made Usenet an increasingly frustrating system for many users. Many people abandon the system as a result. However, recent studies have shown that Usenet is still a robust and growing environment [13, 17]). While Usenet has not been consumed by its tragic qualities, better tools

for navigating and visualizing this space may allow people to return and restore this public commons. Collaborative filtering has been applied to news articles, such as in [12]. Our work presents a different approach.

The Netscan Usenet analysis system is located at <http://netscan.sscnet.ucla.edu>. The Netscan data collector constantly examines an ongoing stream of Usenet messages and extracts information from each message's header. Netscan collects and archives the contents of the From, Newsgroups, Subject, Date, Organization, Lines, Message-ID, and References lines. Netscan builds a number of databases that organize this data, which can be queried to generate a range of measures of activity and relationships in selected newsgroups. These measures are assembled into reports and graphical illustrations that are accessible via a standard web browser.

Visualization

We are interested in visualizing connections between newsgroups and among topics and participants within newsgroups. Our goal is to support users in the process of gaining general awareness of the structure of a newsgroup by determining the groups it neighbors and the range of opinions expressed in the group. This information should aid the process of asking appropriate questions that are addressed to the proper group and to people within the group as well as evaluating the contributions of other participants. As a result, we believe that these visualizations will encourage participants to contribute more effectively as well as more efficiently draw resources from these groups.

Successfully completing these tasks depends on a fairly fine-grained understanding of the newsgroup(s) of interest, their topics of discussion, and the relationships between participants; exactly the kind of collaborative information that connects people and objects. To support these tasks we have created four different views:

- Inter-Group View
- Inter-Thread View
- Thread-to-Poster view
- Inter-Poster View

Examples of these are shown in Figures 5-8. Each view is a network diagram with nodes representing either objects or people, and links representing different connections.

The Inter-Group View (as shown in Figure 5) displays all the newsgroups linked to a selected newsgroup (or set of newsgroups) by cross-posts, or messages sent to more than one group. This is an example of an Object-to-Object Diagram. Two groups are closely linked if there are many shared cross-postings from people. Strong ties may indicate similarities or shared interests between two groups.

The Inter-Thread View (as shown in Figure 6) displays all the threads in a particular newsgroup linked by shared

posters. Chains of posts that answer previous posts define threads. By looking at the topics of the threads, one can get a sense of the topics being discussed within the newsgroup.

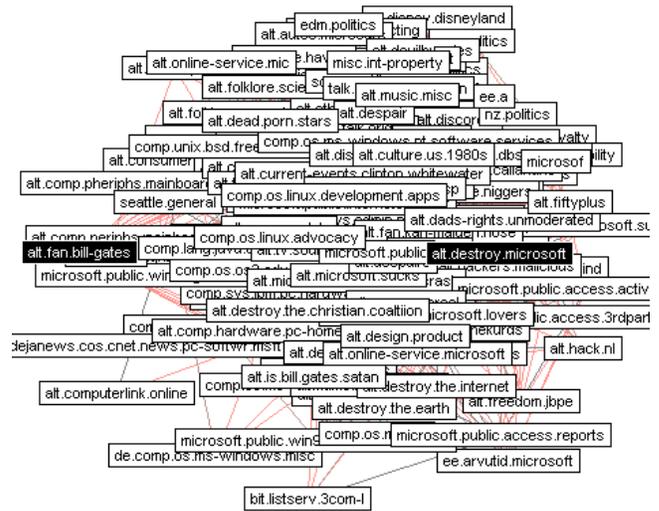


Figure 5. Inter-Group View

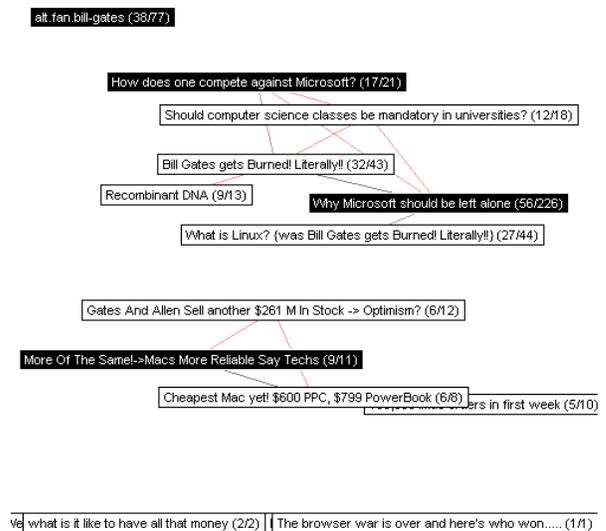


Figure 6. Inter-Thread View

The Inter-Thread View is another example of an Object-to-Object Diagram. Two threads are closely linked if they share many of the same posters. Threads that do not share posters with any others are isolated and displayed at the bottom. These may represent topics that do not generate large amount of interest from the newsgroup.

The Thread-to-Poster View (as shown in Figure 7) displays selected threads in a particular newsgroup and clusters the respective posters around each. This is an Object-to-People Diagram. By following links from topics of interest, one can find people who are interested in these topics. This is can be a first step to finding people of shared interests for interaction.

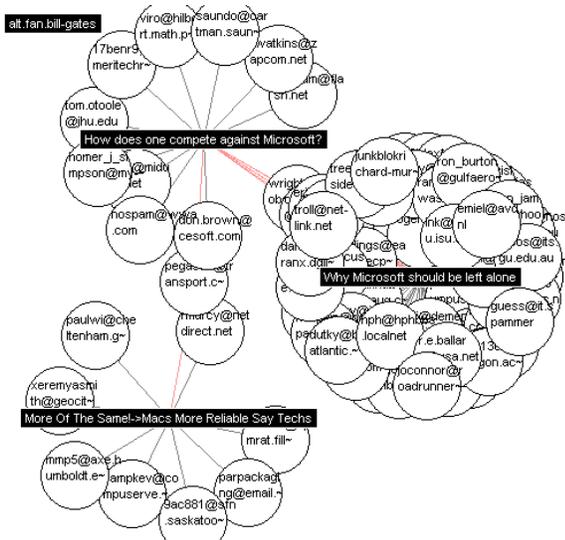


Figure 7. Thread-to-Poster View

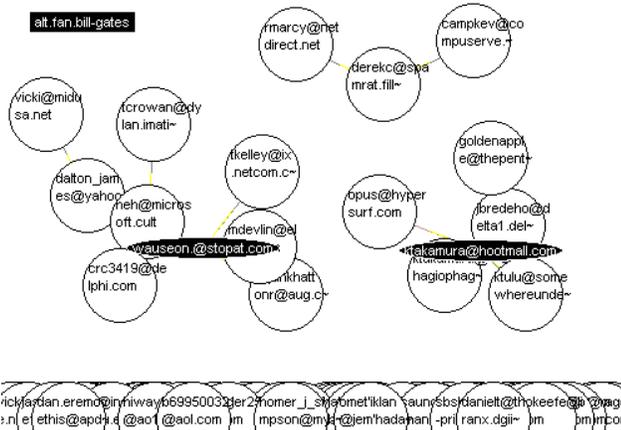


Figure 8. Inter-Poster View

The Inter-Poster View (as shown in Figure 8) displays selected posters in a particular newsgroup linked by shared posts. This is a People-to-People Diagram. This view shows essentially who responds to whom. The people who are highly connected have interaction with many people, and might thus be among the core group. Some posters post infrequently but are responded to by many others. Other people may post frequently but receive few or no responses. By looking at the diagram, one can gain valuable insight about the amount and type of the interaction going on in a group, and also about the status and behavior of individual participants.

Interface

Our goal is to provide a visual interface to Usenet that can be useful to the average user of newsgroups. We have applied the following guidelines in our design: simplicity and clarity, dynamic retrieval and display, interactive user-control, and overlay of information on demand.

The Netscan network visualization we have created uses a simple spring-based model for layout. Each node exerts a repelling force on other nodes, and each edge exerts forces

on the nodes it connects when its length differs from its preferred rest length, determined by the connection strength. User can directly select a node, move it, and see the connected nodes and nearby nodes move accordingly to minimize the forces on them. This is a simple, intuitive way to sort through the data and interact with the display.

Because a large amount of nodes may need to be displayed we provide other ways to filter the information visualized, differentiate better among the different nodes, and select views of interest to the individual user. The Netscan visual interface uses corresponding controls as shown in Figure 9. In the following, we examine the important ones in order.

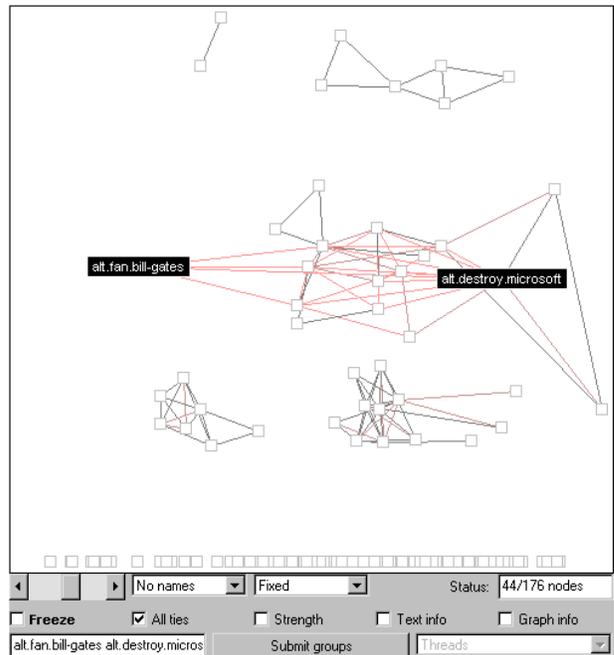


Figure 9. Netscan Control Panel and Network Display

The slider at the far left is used to filter out weaker tie strengths. When the slider is at the far left, the display contains all the ties between all of the groups related to the core newsgroup. At the opposite end only the strongest ties (in the inter-group view, pairs of groups sharing the most number of crossposted messages) will be displayed. The nodes that are no longer connected fall to the bottom of the display. The number of nodes currently connected to other nodes is shown in the status field to the right.

The NAME pull down menu contains three options: ALL NAMES, which displays all names cut to 20 characters and thus fit in boxes of uniform size; LONG NAMES, which displays full names; and NO NAMES, which displays a simple square or circle for each node. The last selection can reduce the clutter on the screen and make the network structure easier to see, as shown in Figure 9.

The FIXED pull down menu contains three options: FIXED, FREE, and WEIGHTED. These options provide three varying approaches to the placement of the nodes. The first two allow nodes to arrange themselves freely with

the initially selected groups either fixed or not fixed. The last selection places the more well-connected nodes, or those with higher degrees, higher. It is useful for identifying the more important or central nodes.

The nodes in the display may continuously jiggle around to find the optimum layout. This can be distracting. Checking the FREEZE box stops this process and allows the user to manually rearrange the nodes to better display the relationships of interest.

The ALL TIES check-box toggles the display from a hub and spokes to a network. The hubs are the nodes users have selected based on their interest. When it is checked, ties between the hub and spoke groups are displayed along with all the connections between spoke groups.

The STRENGTH check-box toggles the display of the number of messages that tie any two groups together. The TEXT INFO and GRAPH INFO show additional information about the nodes, such as group size, using text and graphics (size and color), respectively.

USER SCENARIOS

The different scenarios our visualization addresses can be divided into three kinds -- questions about groups, threads, and people:

1. What groups are related to those that I am interested in? Which group to post or follow? Which group has the most activity? How about the kind of interaction in the group? How are groups similar or different?
2. What are people talking about? How are different subjects related to each other? Who are the people interested in these topics?
3. Whose message to follow? What percentage of the group interact with each other? Who are the core people in each group? How much do they interact with each other?

Groups

For user scenarios such as general awareness about a subject, it is important to find out the relationship among groups, and to find out how groups are different.

Say we are interested in what people think of Microsoft, and have picked two core groups of interest: alt.fan.bill-gates and alt.destroy.microsoft. The resulting diagram in Figure 5 contains many edges/nodes. These represent the different groups that have cross-posts with the core group. In other words, people who are interested in the core groups are also interested in these other groups. Many of these edges are weak, and consist of only 1 or 2 postings. We can filter out the weaker links using the slider. As shown in Figure 10A, clusters of newsgroups that tightly bind to each other start to appear. Some of these make clear sense, for example, the cluster in the upper right is about politics, and the cluster in the lower right is related to jokes, but other cluster are less obvious, suggesting interesting connections.

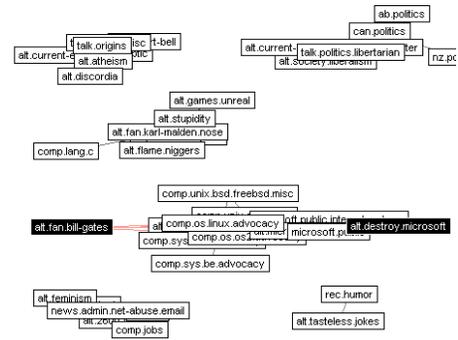


Figure 10A. Inter-Group View, filtered



Figure 10B, Inter-Group View, further filtered

Figure 10B shows the strongest connections to the group alt.destroy.microsoft. It gives the user a fairly good indication of the interests of the people involved in the selected target group.

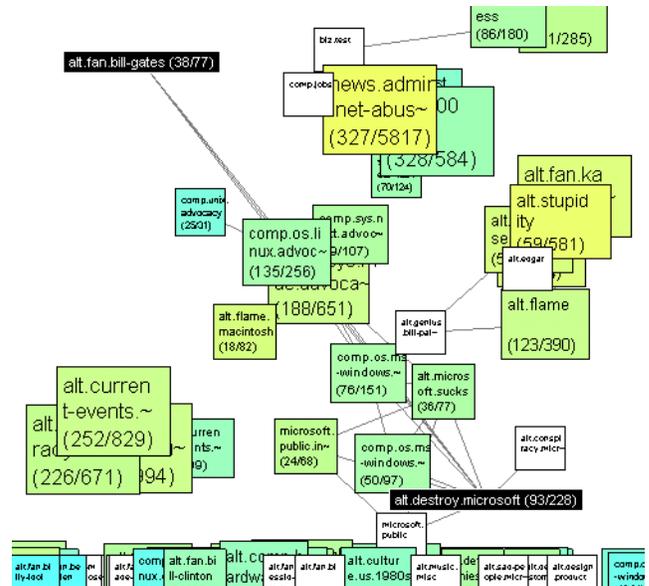


Figure 11. Inter-Group View with poster/post information.

To determine which newsgroup to post or read, users need to know the amount of interaction and overall activity within each group. For that, users can overlay information on the Inter-Group View, as shown in Figure 11.

The numbers in the parentheses represent number of poster and posts, respectively. The size of the box shows the amount of postings, and the color represents the ratio of the number of posters to posts. The size of the groups shows the amount of postings. The poster-post ratio is an important hint to the nature of intra-group interaction. A group with ratio of one (in blue) could be an announcement group, because each person posts only one message: not

much interaction between people. A group with ratio close to zero (in yellow) could be a moderated group with one person or a few people posting all the messages. A group with ratio somewhere in the middle (in green) may indicate a group with healthy interaction.

Threads

Once a user has pinpointed a newsgroup of interest, it is important to know what goes on within the selected group. What are the topics being discussed? Which topics are most popular? Which topics are of general interest?

The Inter-thread View, shown in Figure 6, gives users a fairly good sense of the topics under discussion for the group alt.fan.billgates. Notice for example, one can see two clusters formed: one about Microsoft in general, and one about Mac. By overlaying information about the number of posters and posts of the threads, we can identify topics of hot discussion by multiple people, just as in the Inter-Group View. We can select particular threads of interest, and look at who are the people involved in the discussion by turning to the Thread-To-Poster View in Figure 7. This view also shows who are the most active posters within a discussion thread. It is a good way to identify people of shared interest to interact with.

People

We can examine at the people aspect in more depth. What percentage of the group interacts with other members? Who are the core members? How often do they interact with each other? The Inter-Poster View is designed to answer these questions, as shown in Figure 12.

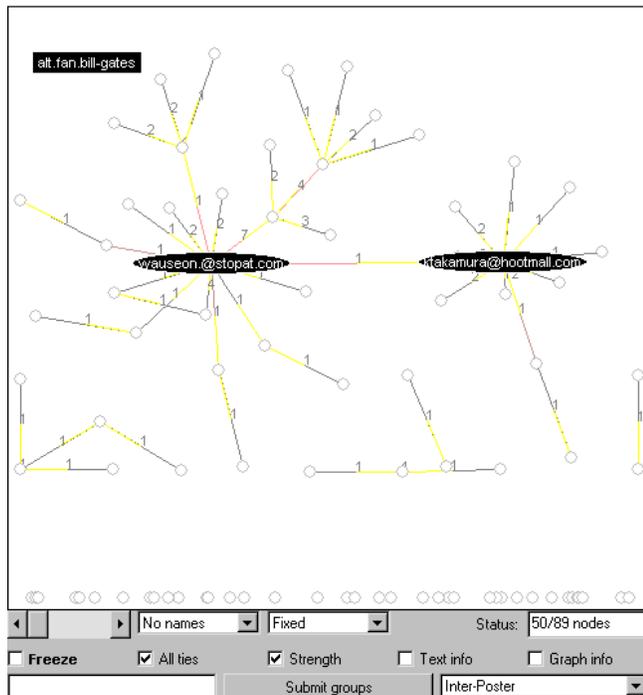


Figure 12. Inter-Poster View with NO NAMES

Immediately, one can see that about half the people, or 50/89, actually talk to each other, as indicated by the status

field. The rest are not connected and shown in the bottom of the display. These messages have never been replied to. In contrast to the views of interconnections between newsgroups, which are bi-directional, edges in the Inter-Poster view have a direction, if one person responds to another the reciprocal is not necessarily the case. In the display, the yellow end of the lines connecting posters indicates the poster being responded to. The display makes it clear that there are two central hubs near the center of the display. The one on the left both responds to and is responded by others – perhaps an expert in the group; the one on the right is responded to by many others – perhaps someone who started a very interesting discussion, or someone being particular annoying.

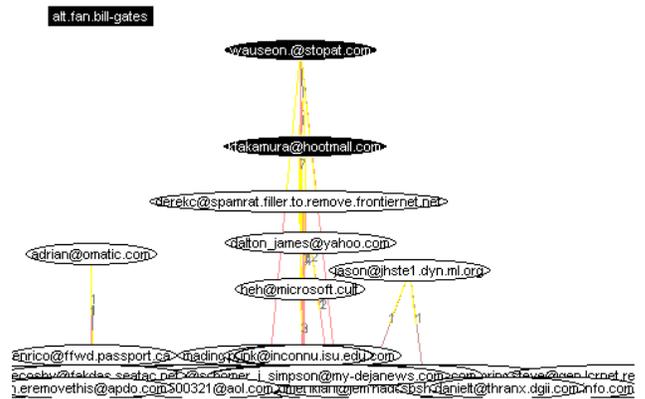


Figure 13. Inter-Poster View with LONG NAMES and WEIGHTED display.

To clearly identify the active participants of the group, we can use the WEIGHTED display, as shown in Figure 13. The nodes with the highest degree, or largest number of connections, float to the top. We can immediately see who forms the core of the discussion group. Based on this information, the user can have a much better sense of the people involved in the discussion, and become a more informed participant.

DISCUSSION

Currently, many network interaction systems are predominantly textual and require users to engage in a great deal of clicking and branching in order to assemble piece-by-piece a holistic view of the interaction space. We believe that these spaces need significant enhancement to reintroduce the social context present in physical interactions that is missing in these media.

Our work presents a first step in this direction. Our current results suggest a number of further enhancements. For example, the layout algorithm is a simple scheme that works well for reasonable number of nodes. When there are much more nodes, a better layout scheme is needed. Graph layout is a NP-complete problem with its own research area. We hope to continue to adapt development in this area to enhance user-oriented visualization.

Currently, users must know the name of newsgroups they are interested in visualizing. Integration with full text indexes, like Deja News and Alta Vista's Usenet index could significantly improve the system's capacity to select groups and individuals.

In addition, we wish to improve the transitions between views and the integration of additional information about the objects displayed. In particular we are exploring ways of integrating representations of activity over time, as animation or as a histogram.

We would also like to allow users to navigate or go on automatically piloted "tours" of the interaction space based on a range of parameters like following the strongest, weakest or most recent tie.

We believe that other forms of collaborative filtering systems, such as chat spaces, web based discussions, and, as demonstrated, complex databases, like the Amazon.com book and music holdings can be usefully visualized.

Another direction is to integrate multiple media, for example pairing a synchronous chat with each Usenet newsgroup so concurrent users can interact directly while using the system. We see a way to support a chain of steps starting with searches, leading through navigation of graphical displays and culminating in direct interaction between people.

In all this, we need to be concerned with the privacy of our users. It is not as much an issue with our Newsgroup visualization, since posters' email addresses are already public. In general, many views of our system can still provide a lot of information without specifying the user identity. Knowing the other interest of others who share our interest gives enough hints regardless of whether we know their names.

CONCLUSION

We have presented a framework for generating visualizations for collaborative information. The four different diagrams (Object-to-Object, Object-to-People, People-to-People, and People-to-Object) show different interconnections between people and objects. We have applied it to the Usenet groups and shown how it can be used to find newsgroups and people of interest. This technique can be applied to a range of different collaborative information to help users understand the context of online social interaction environments.

ACKNOWLEDGMENTS

We thank Josh Berman, Susan Dumais, Eric Horvitz, Yezdi Lashkari, and Tamara Munzner for their helpful discussions. Netscan was generously supported by the Microsoft Virtual Worlds Group.

REFERENCES

1. Amazon.com Inc. Web Site: <http://www.amazon.com>.
2. Donath, J., Visual Who: Animating the affinities and activities of an electronic community, at ACM Multimedia 95, Nov 5-9, San Francisco, CA.
3. Goldberg, D., Oki, B., Nichols, D., Terry, D.B. Using Collaborative Filtering to Weave an Information Tapestry. Communications of the ACM, December 1992, Vol 35, No12, pp. 61-70.
4. Hardin, G. The Tragedy of the Commons, Science 162: 1243-48, 1968.
5. Kautz, H and Selman, B., Creating Models of Real-World Communities with ReferralWeb", Working notes of the Workshop on Recommender Systems</i>, held in conjunction with AAAI-98, Madison, WI, 1998.
6. Kollock, P and Smith, M. Managing the Virtual Commons: Cooperation and Conflict in Computer Communities. In Computer Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives, Susan Herring, ed. John Benjamins, 1996.
7. Krebs, V. Visualizing Human Networks, Release 1.0, 12 February 1996
8. Milgram, S. The Small World Problem, Psychology Today, 1 1967.
9. Moreno, J. Who Shall Survive, New York: Beacon Press.
10. Mukherjea, S., and Foley, J. D., Visualizing the World Wide Web with the Navigational View Builder, WWW '95.
11. RazorFish Web Site: <http://www.razorfish.com>.
12. Resnick, P., Neophytos, I., Mitesh, S. Bergstrom, P. and Riedl, J. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In Proc. of CSCW '94: Conference on Computer Supported Cooperative Work (Chapel Hill, 1994), Addison-Wesley.
13. Smith, M., Netscan: Measuring and Mapping the Social Structure of Usenet, 17th Annual International Sunbelt Social Network Conference, San Diego, California, February 13-16, 1997.
14. Terveen, L., and Hill, W., Finding and Visualizing Inter-site Clan Graphs, CHI '98.
15. ThinkMap Web Site: <http://www.thinkmap.com>.
16. Wellman, B., The Community Question: The Intimate Networks of East Yorkers, American Journal of Sociology, 84, 1979.
17. Whittaker, S., Terveen, L., Hill, W., Cherny, L., The Dynamics of Mass Interaction, CSCW '98.