

Loom2 – Intuitively Visualizing Usenet

Sociable Media, MIT Media Lab
<http://smg.media.mit.edu/projects/loom2/>

Judith Donath
Professor
judith@media.mit.edu

Hyun-Yeul Lee
Researcher
hyun@media.mit.edu

danah boyd
Researcher
danah@media.mit.edu

Jonathan Goler
Researcher
jagoler@media.mit.edu

Abstract

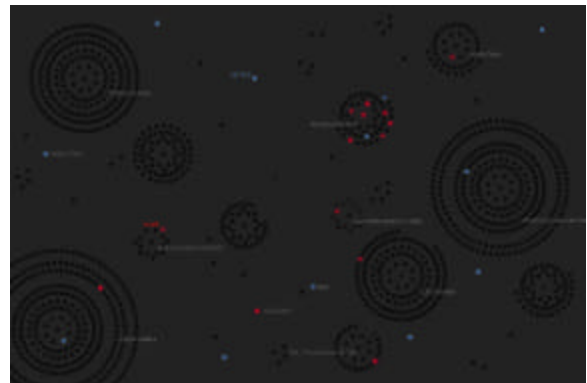
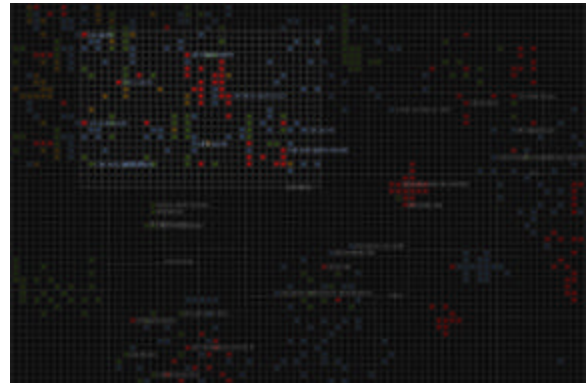
We are currently developing a system for visualizing Usenet newsgroups at a variety of scales. A macro/landscape view depicts many newsgroups and the relationships among them; a medium view depicts the interactions within a single group; a close-up view depicts the individual in the context of the conversational situation. Our research goals include designing intuitive visual representations of social information and furthering our understanding of what are the most socio-culturally significant patterns in the domain of online conversations.

Introduction

Our goal in visualizing social patterns is to help users to understand the community in which they are a part and to find appropriate forums for interaction.

Most data visualization systems are designed to represent numerical data accurately and compactly; ideally, one could go from numerical to graphical representations and back again with no loss of accuracy. Our approach, which we call social visualization, is somewhat different: our goal is to convey the meaning of the data as intuitively as possible. These visualizations are thus inherently designed to be interfaces for the participants in the community rather than for observers, marketers, etc.; their function is to provide a vivid sense of this abstract space, rather than to accurately delineate its statistical features. Ideally, reading the visualization should be as understandable as surveying a natural social crowd, making it simple to determine the level of activity and types of interactions.

The process we are using to create this system integrates computation, sociology, and interaction design. Innumerable patterns can be detected in the many statistical measures derivable from an analysis of Usenet postings, yet only a few are sociologically significant. Similarly, there are an infinite number of ways one could visually represent the statistical data. We believe that by constructing a system that allows one to easily explore different mappings of the Usenet universe we will achieve a much greater understanding of which measures are indeed socially meaningful. A deeper look at the social meaning of the data guides the visual design.



Figures 1-3: An example macro, conversational and individual layout

The Usenet community

We have chosen to use a selection of Usenet newsgroups as our data for several reasons. Usenet is publicly accessible and quite familiar to many researchers in this field; as we progress in our work, we will be able to easily find test subjects and design critics who can judge whether and how these visualizations improve the experience of exploring this social space. Also, Usenet consists of a large number of structurally similar (they share the same interface) groups that are formed around very different topics and have spawned very diverse communities and cultures.

Millions of people from around the world [Whittaker] have participated in the thousands of Usenet newsgroups. They use various browsers, most of which allow the user to navigate by choosing a group from a list of available groups and then browsing thru messages which may be arranged by conversational thread, date or author. The sheer scale of the Usenet universe, combined with the utilitarian interface afforded by most browsers, makes it difficult to sense the important social patterns that exist in this environment.

At the macro level, different groups have very different cultures: some are supportive and friendly, others are argumentative and hostile, some are excellent sources of information, others have devolved to a state where they contain only ads for porn and get rich quick schemes. At the micro level, there are complex and subtle social interactions as participants take on different roles and as they work, consciously or not, to shape the culture of the group in which they are participating.

These patterns and affinities make up the social structure of Usenet, yet they are very difficult to perceive given the current browsing tools.

Usenet is increasingly notorious for its low “signal to noise” ratio. One of our larger goals in visualizing the community is to create an interface that encourages greater cooperation among the participants. For instance, we may surmise that reputation building is a key reason why people spent the time and energy to answer the questions of virtual strangers. A visualization that highlights a group’s core members – say, those who participate frequently and both respond to and are responded to by others (especially in a positive manner, if we do some basic textual analysis) – arguably would further encourage cooperation.

We are interested in portraying both the social structure of various newsgroups as well as individual identities. At a macro view, we want the user to be able to understand what social elements make the various newsgroups unique. Once inside a newsgroup, the user should be able to get a sense of the individuals who are involved in conversation. While this information is not actually absent in Usenet browsers, it is difficult to perceive given the current layout in a short period of time. We intend to highlight this information to put a social focus on Usenet.

Social questions surrounding Usenet

What would a landscape of newsgroups look like? What are the significant interaction patterns within an online conversation? What defines a person’s identity in the context of their Usenet participation? These questions address the core issues that a participant would be interested in knowing in order to have an intuitive sense of the space. Thus, they are the primary questions underlying this project. To address them, we have been formulating a

number of more specific – and at least in approximation, quantifiable – questions. These include:

People and Participation:

How many members post in a given group? In order to answer this question, we are using the email address from which the user posts their message. We distinguish between “general membership” and “active membership.” General membership consists of all unique users who post to a given newsgroup. Active membership considers the frequency with which an individual posts original questions, responses to other questions, follow-ups to other’s responses to her/his original post, and frequency over time compared to the current time. This creates a quantitative method of saying how active an individual is within the context of one group.

Do a few individuals dominate the group? We are interested in knowing if a group is comprised of a few loud individuals or a multitude of quieter individuals. By considering how active individuals are within a group and how many posts a group has, we determine how frequently a user posts in relationship to each other.

How are newcomers treated? When someone who has not been active in previous conversations posts an initial post, does this evoke responses or is the individual ignored?

How active is a group? In order to give a numerical definition to activity within a group, we are interested in knowing both the level of activity of the people as well as the frequency of posts and the lengths of those posts.

How many groups are people involved in? When we are looking at individuals, we are interested in knowing how many groups they are involved in and what the overlap of people is like between groups. For each group that a user participates in, we compare how frequently a user posts to the number of messages in a group.

Conversational style and tone:

What is the style of a given group? How long is an average post – i.e., do people usually rant or give quick responses? How long is a typical thread? How many people get involved in a typical conversation? These questions give meaning to the style of the group.

Depth vs. breadth of conversations? Do the posts in the group consist of question/answer or is there more dialogue? We determine this based on number of threads and number of responses to a given thread. In addition to considering the direct number of postings, we are also concerned with how many people are involved. For example, a long dialogue between two people is weighted differently than a conversation with responses from a wide variety of people.

What is the tone of a given newsgroup? Using simple text analysis, we are trying to determine if there is a tone to a group. Is there quite a bit of screaming? What about flaming? How frequently are advertisements posted? Is there name calling? If there are quick responses, are they answers to questions or harsh responses? What about the meanings of rants – do they represent a long discussion or a soliloquy? Most of these problems we have not begun to determine how to answer, but we will be using some form of natural language processing.

Is there a great deal of cross-posting? When a message is posted to multiple groups, or cross-posted, it means that the message in that post is generic enough to be posted to multiple places. This affects the tone of a group.

For each message, we are able to see which groups it was posted to.

Time and Aging:

How old is the group? Given enough data, we want to be able to determine age through initial posting. Current browsers delete old messages. Options for getting this information include *deja* and the group's FAQ.

How have the posting and people patterns changed over time? If possible, we would like to analyze an individual's participation over time. When does a user post? Does a user only post on weekends? Are they posting less frequently than they used to?

Data Collection and Analysis

In order to visualize Usenet data, we gather a subset of the newsgroups from a local newsgroup server and store parsed messages into an Oracle database. We are only using a fraction of the newsgroups that are available, limiting ourselves to newsgroups posted in English and with a minimum number of posts in the last month. We then parse the messages for people, time, group, parent message and lexical information. Using this information, we analyze how the data to answer some of the questions mentioned above. After we give quantitative components to the information, we construct a visualization to give intuitive meaning to this statistical data. Our visualization is constructed in a modular fashion so as to allow for different types of visual/data pairings.

Visualization

While the data gives us a starting point, our goal is still to give an intuitive representation to the space and feel of the Usenet community. Our visualization is structured into three components. At the macro level, we are viewing all of Usenet, where each newsgroup is given a representation. This landscape view is intended to give an understanding of how newsgroups are related and give some basic information about what differentiates newsgroups so that the user can explore the next level. After choosing a given newsgroup, the focus of the visualization switches to the types of conversations that are present. The conversation view gives some basic information about the individuals and their participation. In addition, this view amplifies what makes a given newsgroup unique, by allowing the individual to visually explore the dynamics between the people and the messages posted. The final type of visualization we address is person specific. At this level, the user can analyze what an individual's dynamic is, by seeing their patterns, types of comments and ways of responding. This three-tiered approach gives a social overview to the community that has evolved in Usenet.

We are also looking at what types of social spaces exist in real life and how they feel visually. By looking at newsgroups as a series of different communities within a mega-city and developing ways to visually differentiate them, we are hoping to give a recognizable visual language to our visualizations.

Related issues of interest to us

We are interested in participating in the workshop in order to learn what other researchers are doing with

community data, with a particular focus on how they are analyzing and distributing the information.

Are other groups working on visual approaches to community data? How have users responded to feedback through their own data? What social and ethical concerns should we consider as we approach this topic?

What approaches are people using to collecting data that is not produced on their own systems? In older communities such as Usenet, it is difficult to get older data since each newsgroup can be stored on any server and the administrator is in control of that particular group.

We are quite interested in discussing the social aspects of online interaction with other researchers and hope to be able to participate in CSCW's Community Data Workshop.

References

Whittaker, S. Terveen, L., Hill, W., and Cherny, L. (1998). The dynamics of mass interaction, In *Proceedings of Conference on Computer Supported Cooperative Work*, 257-264.



Figure 4. Early Loom2 sketch exploring the macro landscape view.



Figure 5. Other early macro-level sketches