PeopleGarden: Creating Data Portraits for Users

Rebecca Xiong, Judith Donath

MIT Media Laboratory

20 Ames Street

Cambridge, MA 02139 USA

{becca, judith}@media.mit.edu

PeopleGarden URL: http://graphics.lcs.mit.edu/~becca/papers/pgarden

ABSTRACT

Many on-line interaction environments have a large number of users. It is difficult for the participants, especially new ones, to form a clear mental image about those with whom they are interacting. How can we compactly convey information about these participants to each other? We propose the *data portrait*, a novel graphical representation of users based on their past interactions. Data portraits can inform users about each other and the overall social environment. We use a flower metaphor for creating individual data portraits, and a garden metaphor for combining these portraits to represent an on-line environment. We will review previous work in visualizing both individuals and groups. We will then describe our visualizations, explain how to create them, and show how they can be used to address user questions.

KEYWORDS: Information visualization, data portraits, user-centered visualization, interaction context.

1. INTRODUCTION

On-line interaction environments such as Web-based message boards, chat rooms, and Usenet newsgroups have become widely popular. As the number of participants rises, it is increasingly difficult to distinguish individual users and to comprehend the overall interaction context. Often, users must piece together a holistic view of other users through extended participation. This makes it difficult for new users to find appropriate groups to participate in and people to interact with. Even existing users may find it difficult to keep track of an ever-changing population.

What is missing is effective communication of information about participants to each other. Figure 1 lists four questions users might ask about the underlying social environment (the quantitative measures in parentheses provide insights to these questions):

1. Do participants here really get involved? (Post frequently or only once in a while?)

2. How much interaction is there? (Do people respond to each other?)

3. Do participants here welcome newcomers? (Do newcomers get many replies?)

4. Who are the experts? (Who has been here for some time or posted many messages?)

Figure 1. Users' questions about a discussion group.

A simple solution to seeing information about users is to sort all the postings by author. Deja's Web interface for the Usenet news shows author's previous postings [Deja99]. While such views offer some information about author's level of involvement or interests, users still need to search for responses from other authors or for comparison with other authors of the group.

What is needed is the ability to look at different patterns in the archives of postings, such as users' contribution, the rate of responses, the interaction among users, etc.. Although such patterns are hard to see in a mass of archival data, by visualizing the salient data, the key roles and relationships can be readily perceived.

Our solution is the *data portraits* (see Figure 2). Unlike photo-realistic portraits, which show physical features such as gender, age, or race; data portraits are abstract representations of users' interaction history.
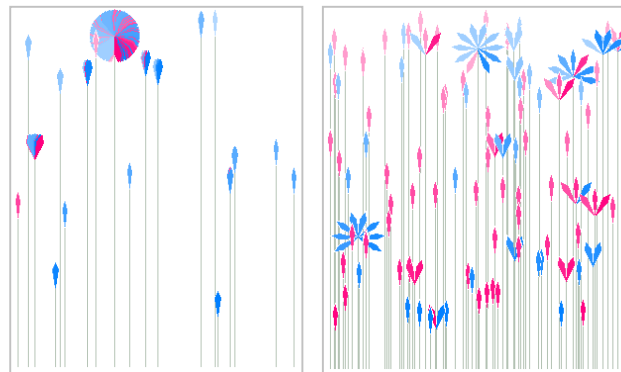

Figure 2. What can data portraits like the ones above tell us about discussion groups? A group with a dominating voice vs. a more democratic group.

In following sections, first we discuss previous people-centered visualizations in more detail. In Section 3, we will define data portraits and discuss how they can be created. Section 4 describes in detail PeopleFlower, our method for creating data portraits. Section 5 describes how PeopleGarden uses a series of PeopleFlowers to represent the different users of an on-line environment. Sections 6 and 7 shed light on the details of implementation and design rationale. Finally, Section 8 describes the future work and Section 9 concludes.

2. PREVIOUS WORK IN VISUALIZING PEOPLE

Information visualization has traditionally been used for the analysis of large amounts of commercial, financial, and scientific data. With the decreasing cost of graphics hardware, it has increasingly been used to help people to interactively filter through information for everyday tasks such as selecting movies [Ahlberg94, Card99]. We examine here people-centered visualizations for showing data about a person, illuminating relationships among people, and for representing individuals for on-line interaction.

2.1 Visualizing individual's history

There has been relatively little work focusing on visualizing information about individuals. One notable exception is Plaisant et al's LifeLines, which shows detailed life history of a person using time-lines [Plaisant96]. It is effective for detailed examination of an individual's medical or criminal records, but is not meant for comparing a group of people.

2.2 Visualizing relationships among people

Comparisons of groups of people have been the focus of social network analysis, which studies the dynamics of relationships among people using a network diagram: each person is represented as a node, links between nodes represent various connections between people.

Originating with Jacob Moreno's sociometric research [Moreno36] 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 [Milgram67]. Network diagrams have also been used to illuminate the empirical structure of local communities, inter-office patterns of communication, and collaboration patterns for papers [Wellman79, Krebs96, Kautz98]. When there are a large number of links, the network diagrams can get fairly complex. These visualizations are useful for sociologists to analyze group dynamics, but are less suitable for average users to gain a sense of the connections among people.

A related piece of work is Judith Donath's VisualWho[Donath95], which uses the positions and shades of people's names to convey their affiliation and presence. The emphasis of the visualization is more on showing relationships among people, and less on information about individuals.

2.3 Representing people

A third area of related work is the design of avatars in virtual environments. Avatars include such varied representations as smiley faces, cartoon characters, 3D models, or photographic representation of users [Damer97, Kurlander96].

Avatars have been used in a variety of 2D and 3D chat rooms and virtual environments. Steve Benford's avatars show where in the virtual environment a user is looking at and acting upon [Benford94]. Paul Rankin's Star Cursor uses the idea of inner vs. outer space to convey interests of users [Rankin98]. Fernanda Viegas' Chat Circle represents users as circles, and changes the size of a circle if the user sends out a message [Viegas99].

These avatars show the current state of the individuals, but do not change over time based on users' past interactions. They do not convey how long the person represented has been in an environment, or whether he/she is a prolific contributor to the conversation.

3. THE DATA PORTRAIT

While there are previous visualizations that convey a user's history, show how users are related, or are compact enough to represent individuals in an on-line environment, we have not found a single visualization that combines all of the above. Our solution is the *data portrait*, or user visualization based on interaction data, both to represent individual users and to compare multiple users.

In real life, successful portraits show the physical likeliness of people; they also use objects in people's lives to illustrate their interests and status [Art97, Brilliant93]. This second aspect is an inspiration for our work. The on-line equivalent of one's objects is data about one's past interactions. For a Web-based message board, where users can post messages or respond to others' messages, a user's data objects are the set of all the messages posted by him/her. In a chat room, the data objects are the messages sent by each user.

A data portrait is a visualization of the user based on his/her data objects. Different portraits can be created for different users depending on the number and attributes of the data objects. Possible attributes for data objects such as messages include time of creation, size, addressee, whether an initial message or a response, or number of responses from others. Various visual parameters such as the size, hue, saturation, or shape of the data portrait can then be used to represent these attributes.

The challenge in creating successful data portraits is thus two-fold: 1) selecting those attributes about data objects that best convey information about a user, and 2) deciding how to visualize these attributes legibly and intuitively.

Over time, a user's portrait changes based on his changing interaction; the portrait will thus always show the most up-to-date data about its user. Unlike avatar representation, these portraits are based on real interaction history.

A collection of data portraits can be used to represent the participants of an on-line environment. Such an overview allows users to see overall trends in the data, such as the level of user participation or the amount of interaction. Because data portraits are generated by simple, uniform rules, they form small multiples that can be compared and contrasted [Tufte91].

As with any visualizations of user information, user privacy is a concern. Our focus is on using information visualization within the context of a community. The visualizations are of publicly available data about the people in the community and their relationships; the target audience is also the members of that community.

4. PEOPLE FLOWER: VISUALIZING A USER

We will now discuss in more detail PeopleFlower, our visualization of the individual. Each user is represented as a flower, as shown in Figure 3. His/her data objects are represented as petals of the flower, arranged by time in a clockwise fashion.

Figure 3. Basic PeopleFlower.

Visual artists have long emphasized using representations that have the same level of visual and conceptual complexity as the objects they represent [Arnheim89]. We therefore wanted a simple object that can easily deal with a changing number of components. We also like the organic nature of a flower, and the suggestion that it changes over time, as users do. The flower metaphor was chosen for its simplicity and intuitiveness. Other visualizations that use radial layout to display data include radar sweep visualizations and Kiviat Diagrams [Jefferies97, Harris97]. For more technical discussions on creating organic or non-photo-realistic rendering, see [Kowalski99, Raskar99]

For simplicity, we will focus our discussion on a Web-based message board. Our findings can be generalized to

visualizations of Usenet newsgroups, Web page traversals, or Chat room interactions with few modifications. We will discuss more in Section 8.

Each PeopleFlower represent a user of the message board, and its petals his/her postings. The different visual parameters of a petal can be used to represent different attributes about each posting. We found the time of posting, the amount of response, and whether a post starts a new conversation to be the three most valuable attributes in conveying social information about the user, as demonstrated below.

4.1 Time since posting

Time of posting is an important attribute for showing posting history. PeopleFlower indicates time in both the ordering and the saturation of petals. A message posted far in the past would be less saturated and less noticeable. As a user posts more messages to the board, his or her PeopleFlower will change, as shown in Figure 4.

Figure 4. A user's PeopleFlower changes over time.

A few things to note here:

- The numbers of petals increases as more messages are posted, just as a flower opens up.

- Older petals move to the left as newer petals are added to the right, to maintain symmetry. The overall shape of a flower is determined by how many petals it has.

- Each petal fades over time showing time since posting.

- A marked difference in saturation of adjacent petals, seen in the rightmost flower, denotes a gap in posting.

4.2 Response to posting

PeopleFlower can display not only a user's own posting pattern, but also the amount of feedback from other users of the board. This is important for illustrating the amount of interaction in an environment. We have used pistil-like circles on top of the petals to denote responses. Each circle indicates one response to a post, as shown in Figure 5.

Figure 5. Three users with same amount of postings over different durations, and with different amount of responses.

While the three users in Figure 5 posted the same amount of messages, they posted at different time interval, as indicated by the saturation of the petals. The amount of responses also varies greatly.

- User A posted a series of messages in a short interval, and has not posted for a while.

- Initial response to A was high, over time it decreased.

- Users B and C both posted their messages over longer intervals, but C has gotten much more response.

4.3 Initial post vs. reply

Why do the amounts of responses differ for these three users? Figure 6 shows the same PeopleFlowers with petals colored by whether the post starts a new conversation (i.e. initial posts versus replies). This scheme has been used in the Loom visualization for newsgroups [Donath99].

Figure 6. The same three users from Figure 5, this time with magenta representing initial posts, blue replies. (See http://graphics.lcs.mit.edu/~becca/papers/pgarden .)

We gain more insight into the situation:

- User A started with an initial post (shown in magenta), perhaps a question or a new topic of discussion. There were many responses. Then User A replied to others' postings, and the level of response gradually dropped.

- User B only responded to others' posts, as indicated by all blue posts.

- Users C made both initial posts and replies. This user likes to both start new conversation and interact with others. User C received the most replies.

This example seems to suggest that one way to get more response from others would be to start a new conversation. See [Whittaker98, Smith99] for more studies about the interaction dynamics of Usenet newsgroups.

5. PEOPLE GARDEN: VISUALIZING GROUP OF USERS

Having described our representation of individual users, we will now turn to placing these individual representations in the larger context of the on-going discussions, i.e. the message board. Figure 7 shows a section of a sample board with about 1200 messages posted by 150 authors over the period of two months.

Currently, users need to read through a large amount of text to determine who's who in the message board. PeopleGarden is designed to facilitate this task. While PeopleFlower specifies how to represent individual users,

Colorado and stuff - Jen 21:50:54 4/23/99 (1)
    Re: Colorado and stuff - Dan 08:57:48 4/24/99 (0)
Apology for "provoking" - Cathy 22:00:48 4/22/99 (2)
    Re: Apology for "provoking" - James 12:47:31 4/23/99 (1)
    Re: Apology for "provoking" - Mia 06:05:28 4/24/99 (0)
L/R Brain info. continued... - Kam 18:44:28 4/22/99 (1)
    Re: L/R Brain info. cont... and.... - Cathy 21:02:30 4/22/99 (0)

Figure 7. Traditional view of a message board.

PeopleGarden specifies how to layout the flowers for overview and comparison of users. We have used the garden metaphor because a healthy garden has certain properties that we can use to represent a healthy discussion group. For example, a garden with more bright flowers indicates a discussion group with more new posts. It looks healthier than one with faded flowers representing a group with mostly old posts.

Figure 8 shows a PeopleGarden created for the same message board. The height of a flower represents how long a user has been in the board, as indicated by time of first posting. This uses the metaphor that a flower planted earlier is taller than one planted later. A few patterns here:

- Many people have participated at the board since the beginning, and have been prolific contributors since.

- A large number of casual participants dropped by to make a few posts.

- There are more users joining the board in the first half than the second.

We can also notice outliers such as the big red flower two-thirds from the top and two-thirds from the left of Figure 8. While most people have more to say the longer they have been at the board, this relative late-comer has really bloomed and contributed much to the discussion.

Now let us return to the four user questions we have raised at the beginning of the paper about on-line environments:

Q1. Do participants here really get involved? (Post frequently or only once in a while?)

Q2. How much interaction is there? (Do people reply?)

Q3. Do participants here welcome newcomers? (Do newcomers get many replies?)

Q4. Who are the experts? (Who has been here for some time or posted many messages?)

The PeopleGarden in Figure 8 can help to answer a couple of these questions.

A1. There are many users here with varying level of involvement. Many of them have also posted continuously, denoted by smooth change of saturation across a wide range.

Figure 8. A PeopleGarden showing messages from a message board with 1200 postings over a 2-month period. Height of flower denotes amount of time a user has been at the board.

A4. The most prolific contributors also tend to be those who have been here the longest, and can be found at the top of the figure.

Figure 8 is useful for answering questions about general posting trends, since all the flowers are displayed at once. For detailed questions about the level of responses to individuals, another view is required.

Figure 9 shows a different PeopleGarden that lays out flowers sorted by the amount of petals. Posts are colored differently based on whether it is an initial post. The responses for each post are also shown. A smaller message board is used here to show the details more closely. Figure 9 can help to answer the remaining questions:

A2. People in this board respond to each other quite a bit. In fact, most posts are responses shown in blue, and more than half the posts are replied to, as indicated by the gold pistils.

A3. Newcomers shown at the bottom have a 50-50 chance of getting a response, only slightly worse than the old-timers. Also, their initial posts are much more likely to get a response.
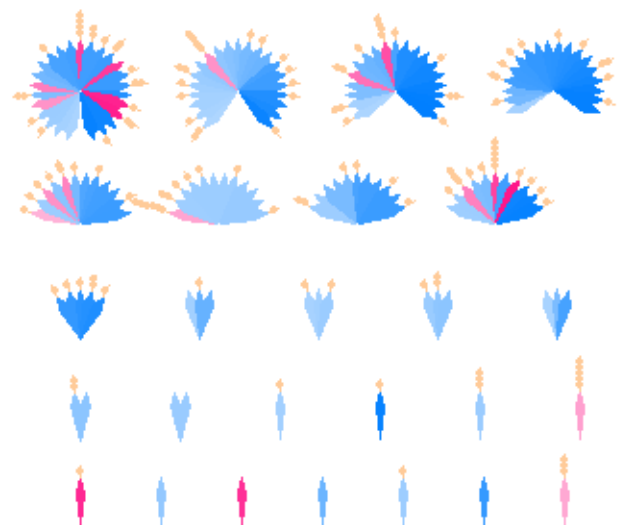

Figure 9. A PeopleGarden sorted by amount of postings. Magenta denotes initial postings, and blue replies. (See http://graphics.lcs.mit.edu/~becca/papers/pgarden for colored picture.)

5.1 Comparing different groups

PeopleGarden applied to multiple groups shows how the interaction in these groups differs. For example, Figure 10 shows the PeopleGardens for two groups with around 200 posts each, but very different interaction. Height of a flower again denotes how long a user has been at the board. The URLs for the two groups are respectively

- http://graphics.lcs.mit.edu/~becca/enneagram/movieboard

- http://www.usscreen.com/message/general/

Figure 10a) shows a group with one dominating voice with many initial posts (in magenta) and more responses (in blue). There is a trickle of new participants throughout the time period, most of whom tend to post replies. This is more a discussion group with predominantly replies in blue. Figure 10b) shows a more democratic group with more equal participation. There are a larger number of participants, each of who posted between 1 and 10 messages. People tend to post more initial posts than replies. This suggests that this group is more of an announcement group.

6. IMPLEMENTATION

PeopleGarden has been implemented as a Java applet. Each petal of a PeopleFlower is a 4-sided polygon. To create a flower, draw each petal in a clockwise fashion. The spacing between petals is the same for all flowers in a garden. The flower with the largest number of petals almost completes a circle.

PeopleGarden lays out the individual flowers by setting their center point. For example, the sorted view reduces row by row the spacing between the center points of flowers. Thus, the later flowers with less information will not take up disproportional amount of screen real estate.

7. DISCUSSION

Different encodings of the display and layout of PeopleFlowers can be used for expressing different data. How does one decide which visual encoding to use? While we have not yet conducted formal user testing, we have collected informal feedback from people. We found that people prefer encodings with some basis in reality. For example, most users found it easy to tell how long a user has participated in the message board based on the height of the flower.

For encodings not based on physical reality, simpler is usually better. For example, the two-tone coloring of petals is fairly easy to comprehend. But when more than two colors are used, the display becomes too cluttered. Also, we have experimented with showing time of posting based on angular position of petals, but this produced irregular flowers that were too complex to compare at a glance.
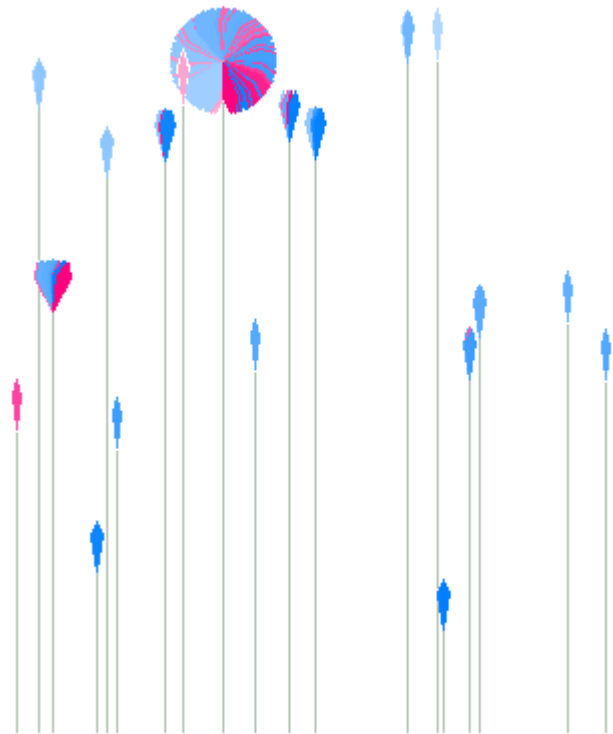

Figure 10 a) A group with one dominating voice (http://graphics.lcs.mit.edu/~becca/enneagram/movieboard, July '99) with a large number of replies in blue. Height of a flower denotes how long a user has been at the board.
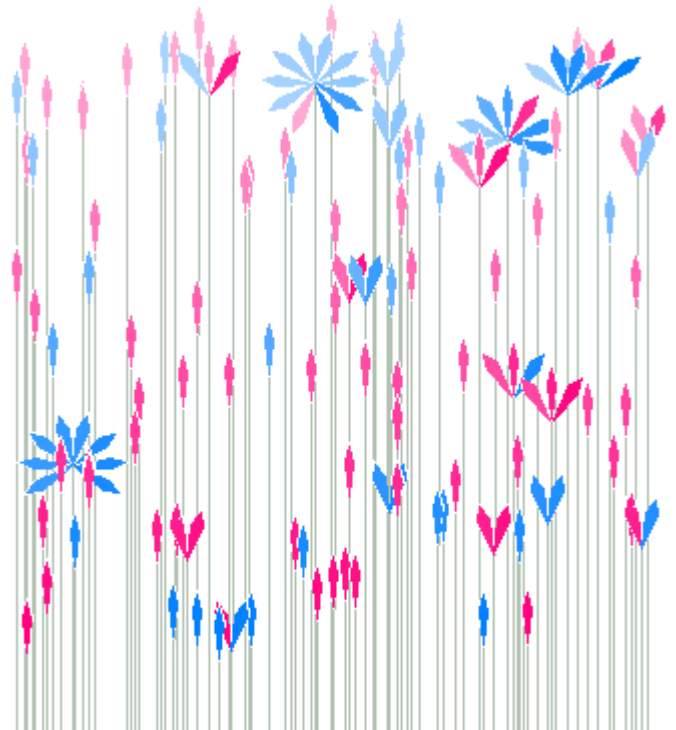

Figure 10 b) A more democratic group (http://www.usscreen.com/message/general/, July '99) with a large number of initial posts in magenta. Height denotes how long a user has been at the board.

How does PeopleFlower scale as the number of posts for a given user increases? We found that the petals can still be distinguished clearly when the ends of adjacent petals are more than one pixel apart. However, pistils representing responses cannot overlap, or the result would be illegible. One solution would be to use each petal to represent the average or median value of multiple data objects.

Since our visualization is based on users' actual history, we need to be concerned with their privacy. The visualizations presented here use only publicly available posting histories. For more sensitive information, a possible solution is to not specify user identity, or use only pseudonyms. Our visualization can still show patterns about an interaction environment without identifying individuals.

8. APPLICATIONS AND FUTURE WORK

PeopleGarden can be readily applied to other threaded discussion space such as Usenet newsgroups. It can also be applied to other interaction spaces such as chat rooms with some modifications. One small difference is that there are no clear distinction between initial post and response. We can use instead different categorizations of chat messages such as question versus a statement. A related application is representing users' traversal history through a Web site.

There are several directions of future work for PeopleGarden. We would like to conduct formal user testing for more detailed answers about which encoding works best for people and further evolve our designs.

We would like to dynamically update PeopleFlowers of users in Web boards, and chat rooms. One can watch the growth of a PeopleGarden over time as users interact more with each other.

9. CONCLUSION

To interact effectively on-line, one needs to know about other participants. Unfortunately, existing interaction environments require extensive participation for users to form a holistic view of the interaction context. We believe that these environments can be enhanced by re-introducing the social context present in physical interactions.

We have presented the data portrait as a novel method for conveying information about participants of an on-line environment. PeopleFlower visualizes a user's posting history; PeopleGarden uses a collection of these flowers to represent users of a message board. The various views of PeopleGarden can be used to successfully answer user questions about the board. PeopleGarden can be extended to other on-line environments to illustrate their social context. We believe that these data portraits can help users to understand an interaction space, see how it changes over time, and provide a mirror for the level and nature of the interaction in that space.

Color pictures and an actual demo can be found at

http://graphics.lcs.mit.edu/~becca/papers/pgarden

REFERENCES

[Ahlberg94] Ahlberg, C. and B. Shneiderman, Visual information seeking: Tight coupling of dynamic query filters with starfield displays. Proceedings of CHI '94.

[Arnheim89] Arnheim, R., *Visual Thinking*, University of California Press, 1989.

[Art97] *The Art Book*, London: Phaidon Press, 1997.

[Benford95] Benford, S., J. Bowers, L.E. Fahlén, C. Greenhalgh, and D., Snowdon, User embodiment in collaborative virtual environments, Proceedings of CHI '95.

[Brilliant93] Brilliant, R., *Portraiture,* Harvard University Press, 1993.

[Card99] Card, S.K., J.D. MacKinlay (ed.), B., Shnieiderman (ed.), *Readings in Information Visualization: Using Vision to Think*, Morgan Kaufman, 1999.

[Deja99] Deja News Web site, http://www.deja.com/, 1999.

[Damer97] Damer, B., Avatars! : Exploring and Building Virtual Worlds on the Internet , Peachpit Press, 1997.

[Donath95] Donath, J., Visual Who: Animating the affinities and activities of an electronic community, at ACM Multimedia 95, Nov 5-9, San Francisco, CA.

[Donath99] Donath, J., K. Karahalios, and F. Viegas, "Visualizing conversations". In Proceedings of HICSS 32, January 5-8, 1999.

[Harris97] Harris, R. L., *Information Graphics: A Comprehensive Illustrated Reference: Visual Tools for Analyzing, Managing, and Communicating*, Management Graphics, 1997.

[Jefferies97] Jeffery, C., "A menagerie of program visualization techniques", in Software Visualization: Programming as a Multimedia Experience, MIT Press, 1997.

[Kautz98] Kautz, H and B. Selman, Creating Models of Real-World Communities with ReferralWeb, Working notes of the Workshop on Recommender Systems, held in conjunction with AAAI-98, Madison, WI, 1998.

[Kowalski99] Kowalski, M., Markosian, L., et al, Art-Based Rendering of Fur, Grass, and Trees, in SIGGRAPH 99.

[Krebs96] Krebs, V., Visualizing Human Networks, Release 1.0, 12 February 1996.

[Kurlander96] Kurlander, D., T. Skelly, and D. Salesin, "Comic chat", Proceedings of SIGGRAPH '96.

[Milgram67] Milgram, S. The Small World Problem, Psychology Today, 1 1967.

[Moreno36] Moreno, J. *Who Shall Survive: Foundations of Sociometry, Group Psychotherapy, and Sociodrama*, New York: Beacon Press.

[Plaisant96] Plaisant, C., et al., LifeLines: Visualizing personal histories, In Proceedings of CHI '96.

[Rankin98] Rankin, P.J. et al, Star Cursors in Content Space: Abstractions of People and Places, In Technical Sketches of SIGGRAPH '98.

[Raskar99] Raskar, R. and M. Cohen, "Image Precision Silhouette Edges", Symposium on Interactive 3D Graphics, Atlanta GA, April 1999.

[Smith99] Smith, M., "Invisible Crowds in Cyberspace: Measuring and Mapping the USENET", *Communities in Cyberspace*, Routledge, 1999.

[Terveen98] Terveen, L., and W. Hill, Finding and Visualizing Inter-site Clan Graphs, Proceedings of CHI '98.

[Tufte90] Tufte, E.R., Envisioning Information, Graphics Press, 1990.

[Viegas99] Viegas, F., and J. Donath, "Chat Circles", in Proceedings of CHI '99.

[Wellman79] Wellman, B., The Community Question: The Intimate Networks of East Yorkers, American Journal of Sociology, 84, 1979.

[Whittaker98] Whittaker, S., L. Terveen, W. Hill, L. Cherny, The Dynamics of Mass Interaction, CSCW '98.