Social Network Visualization: Can We Go Beyond the Graph?

Fernanda B. Viégas
fviegas@media.mit.edu

Judith Donath
judith@media.mit.edu

MIT Media Laboratory
20 Ames Street
Cambridge, MA
02142 USA

Abstract

Visualizations have played an important role in generating new insights in social network analysis. We suggest that such visualizations can be of interest not only to analysts and researchers but also to the people whose data is being analyzed. In this paper we briefly talk about two visualizations of email that we developed to give people a better sense of their email archives and social networks. One visualization shows a traditional network graph with email contacts as nodes. The second visualization depicts the temporal rhythms of interactions in dyadic relationships between ego and individual contacts. While observing and interviewing users of these systems, it became clear that, when used in tandem, these visualizations complemented and clarified each other's depiction of a person's social network. Based on our experience with these two systems, we propose that visualizations of social networks that are aimed at end users ought to go beyond the graph paradigm. We posit that basic cartographic principles – such as adaptive zooming and multiple viewing modes – provide system designers with useful visual solutions to the depiction of social networks.

1. Introduction

Visualizing social networks is more than simply creating intriguing pictures, it is about generating *learning* situations: "images of social networks have provided investigators with new insights about network structure and have helped them communicate those insights to others" [1]. Such network images are created mainly in two ways: the first one is by drawing graphs made up of nodes and connecting lines. The second way is to devise a matrix where rows and columns stand for people and the numbers in each cell stand for the social connections between the people. In practice, however, most social network applications have focused on the graph representation [1].

In this paper we present two visualizations of email that show different aspects of a person's social network. The first application, *Social Network Fragments* (*SNF*), is a traditional graph visualization that highlights clusters of contacts derived from the TO and CC lists in email archives. The second visualization, *PostHistory*, depicts the temporal rhythms of interactions between ego and individual contacts over time. We discuss some of the lessons learned from watching users interact with these systems. Even though only one of these systems illustrates the conventional social network visualization – a graph of connected nodes – it is the joint use of these two applications that allows users to more fully understand the makeup of their networks.

Based on this experience and additional informal observations, we posit that there is a need to rethink the ways we create computer-generated representations of social networks. By critically pondering the shortcomings of current depictions of social networks we hope to create more legible ways in which to represent these complex systems.

2. *SNF* and *PostHistory*

SNF and *PostHistory* focus on two major dimensions of email archives: people and time. Even though both visualizations reveal aspects of the email social landscape of the user, they do so in very different ways. *PostHistory* focuses on the social world of dyadic email relationships whereas *SNF* explores the groupings of people that emerge within a person's social network as seen through email exchanges.

In *SNF*, a person's email archive is mined for header data such as CC and TO lists. The system uses this information to derive a matrix of connections between all the recipients of email messages. In the tradition of social network visualizations, the system uses this matrix as input for a spring system algorithm, which attempts to maximize the ideal position of all people on

Fig 1. PostHistory interface with calendar panel on the left and contacts panel on the right. Names on the right panel move higher to reflect more intense email exchanges with ego. As time progresses and the intensity of exchange changes, names either slide back down or stay stationary.

a 2D plane. Those with tight bonds are pulled towards one another; those who do not know each other – i.e. people who have never appeared in the same CC or TO lists – are repelled. The graphical layout determined by the spring system algorithm is the basis for the user interface (Fig. 2).

PostHistory is a visualization that focuses on time and rhythm, where the variations in long-term email exchange are revealed to the user. The interface is based on the metaphor of a calendar (Fig. 1). The system visualizes the amount of email exchanged over time between ego and each different contact, revealing large concentrations of interaction during certain periods in contrast to times when almost no email was exchanged. For a more in-depth explanation of how both visualizations work, please refer to [3].

PostHistory focuses on the users' direct interactions with each of the contacts in their email world. In contrast to this dyadic focus, *SNF* reveals a world of social collectivities where friends are linked to other friends, work colleagues might be linked to some of ego's family members, etc. Because of their distinct approaches to email, these visualizations generate different insights, which frequently complement one another as will be discussed in the next session.

3. Small Evaluation: Case Studies

While developing *PostHistory* and *SNF*, we ran a small ethnographic evaluation with ten users. All ten people had their own email data visualized. Out of the test users, two became our case studies because they

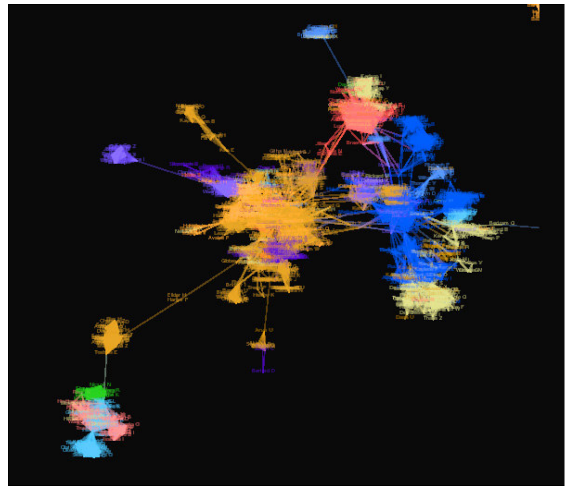

Fig 2. A complex cluster of contacts in SNF. The colors indicate that the cluster includes people from different contexts of ego's social life: family, school friends and work colleagues.

provided us with extensive personal email archives that spanned five years (from 1997 to 2002). Both of these people had their archives visualized by the two applications and both took advantage of this simultaneous access when interacting with the systems.

When looking at *SNF*, most users first passively observed the visualization animate over time, adjusting to the movement of the vast constellation of names displayed on the screen. Users would then focus on graphically interesting clusters and start exploring, usually by zooming in to see the various names in a tight group.

In *PostHistory*, users saw the "rise and fall" of many relationships: "I loved to see the pattern of my relationships with various lovers: intense conversation, then stability, then slowed down conversation and then !bam! no conversation (a.k.a. breakup)." *PostHistory* also highlighted the core group of relations and how this core evolved over time. "Seeing my [contacts] in *PostHistory* makes me aware of how many people overall I know, and how few of them really count. It's fascinating to see how some of the stronger names (higher up on the screen) stay around for a long time, bobbing up and down occasionally; how some of them faded away slowly while others crashed instantly."

Users readily utilized the visualizations to revisit past experiences and to reflect on their relationships with others. We were surprised to find that users felt comfortable sharing the visualizations with friends. Not only did users share the specific portions that concerned their friends, but also entire visualization overviews.

In both of our case studies, users made extensive use of both systems simultaneously, going back and forth between them (we had the applications running side by side on different computers). For example, when users spotted an interesting cluster of people in *SNF,* they would turn to *PostHistory* to locate the patterns of intensive email exchange that made those people's names coalesce into a single cluster. One of our users was able to trace how she got involved in a legal action concerning a group of people she met over one summer. She first saw the tight cluster of names in *SNF* and turned to *PostHistory* to confirm when her exchanges with that group of people had taken place. In such cases, users repeatedly used one system to confirm and contextualize the other.

4. Thinking critically about graphs

Overall, *SNF* qualifies as a successful visualization on the grounds that participants generally understood what the system showed them and they were able to learn new things about their email-based social network. Nevertheless, users expressed concerns about some aspects of the visualization. First and foremost, all users were overwhelmed by the high number of names displayed on the screen and the difficulty of reading them before zooming into specific clusters. Users also complained that they did not understand the static nature of people's location in *SNF*'s space; they wanted people's positions to change over time as they participated in different groups. While the coloring of names was effective for relational purposes, users were confused about its meaning and accuracy.

Perhaps more importantly, however, was the fact that users who had the advantage of having their email simultaneously visualized both by *PostHistory* and *SNF* made heavy use of the systems in tandem. These users were able to make better sense of the structure of their social network on *SNF* than users who did not have access to *PostHistory*. As mentioned earlier, users would go back and forth between *PostHistory* and *SNF* in order to clarify and contextualize what the visualizations were showing them. One obvious explanation of this outcome is the fact that we usually understand things better when we have more than one way of looking at them. In other words, if a user is given any two different visualizations to look at a dataset, chances are she will have more insights about the data than if she had had access to only one visualization to start with.

Even though this might be the case in general, we believe it is of particular significance in getting people to understand graphs of their own social network. We believe that complementary ways of looking at social network data – ways that go beyond the regular "social network graph" metaphor – are crucial for revealing how different clusters came to be and what the network structure implies.

Even though end users have been exposed to images of social network graphs coming from the most varied venues – visualizations of social networks on Friendster and instant messaging systems are examples of this phenomenon – there have not been any studies that look at how much meaning people derive from looking at those images. Informal observations have shown us that, even though people find these graphs interesting and enjoy looking at the names they display, the images are usually hard to interpret. One of the main problems is the occlusion of names/nodes due to the staggering number of nodes and the tight clusters – figure 2 in this paper is a good example of this problem. The other problem with regular social network graphs is that a lot of times it becomes difficult to understand who is connected to whom inside clusters. Because regular spring algorithms work based on the distance between nodes and not on how the graph is ultimately drawn on the screen, no attention is paid to fundamental things such as how many lines end up crossing each other or how many names get written on top of one another. In addition, most graphs make use of straight lines to connect nodes when we know, from studies of visual perception, that it is much easier to perceive connecting lines that are smooth and curvilinear [4].

In short, current social network graphs are drawn from a mathematical instead of a visual-perception standpoint. For this reason, even though these graphs embody lots of interesting data, it can be hard to perceive much of the information they contain.

It is no wonder that the social network pieces done by Mark Lombardi [2] have gained such an enthusiastic following. What Lombardi did that our current computer-generated social network graphs fail to do is address matters of layout. On Lombardi's drawings lines hardly ever cross, nodes do not overlap one another, and connections are smooth and curvy. Another important aspect of these drawings is the fact that they do not try to display millions of nodes at the same time. Some of Lombardi's more complex drawings show hundreds of nodes at once but even then, the elements are laid out in such a way that the patterns of the network are kept legible.

5. Graphs as Infrastructure & Cartographic Principles

Do the problems mentioned above mean that graphs should be disregarded altogether as representations of social networks? Not necessarily. As long as we approach graph drawing critically, taking into consideration the perceptual issues that are involved in trying to draw thousands, sometimes millions of connected nodes on low resolution, small computer screens, graphs remain a possibility.

A promising venue for graphs is to serve as infrastructure for more sophisticated social network visualizations. One of the most useful features of social network graphs is that they create an overall "map" of the network; in other words, they establish a sort of geography where each node is given a position in a 2D plane. This initial "terrain" can then be used as the skeleton upon which the actual visualization is overlaid.

But what kind of visualization would constitute a meaningful addition to this base?

One possibility is to create a cartographic visualization. Cartography is a field that deals with remarkably voluminous amounts of data. Yet cartographers have managed to organize geographic data in a manner that communicates effectively. Maps are the result of a delicate balance between data analysis, synthesis and representation. Two of the basic principles employed in cartographic maps can be easily adapted for social network maps: (1) adaptive zooming, and (2) multiple viewing modes.

Adaptive zooming describes the adjustment of a map, its contents and the symbolization to target scale in consequence of a zooming operation. In a social network map this might refer to the changes that take place when a user zooms, from the overall map, into a specific cluster of contacts. A lot of times clusters are so dense and so tight that nodes end up almost completely on top of each other and names of people (supposing each node represents a person) become illegible. In such cases it is useless to make the visualization program draw every single node and name seeing how they will not be readable to the user. In *SNF*, for instance, users had the ability to zoom into clusters. This simply meant that the computer would render the same lines and nodes it had rendered before, only now in a bigger size. Adaptive zooming is different; it actually renders different *representations* of the same thing at different levels. Therefore, we can create a visualization that has different levels of details at each level of zoom. At the overview level, clusters might not be drawn as detailed compilations of multiple nodes and connecting lines but rather as "regions" on the map – just like a city might show up as a simple dot on a map of an entire country. As users zoom into a cluster region, individual nodes and lines become legible. Moreover, structural elements such as cluster density and recency might be conveyed by the use of color saturation or brightness of each cluster region.

Multiple viewing modes refer to the notion that the same "map base" can show different kinds of information at different times. For instance, a world map may convey geographic information such as the location of mountains and rivers, or it may depict human-related information such as political boundaries, national gross product indices, etc. These are very different kinds of maps but the information they display is shown using the same basic "structure" consisting of the location of continents and oceans around the globe. By the same token, social network visualizations should be able to represent different kinds of information overlaid on the same graph structure: from the location of specific people to the multiplexity of connections between nodes. This ability to utilize the same basic structure – the computer-generated graph – to convey different kinds of information will certainly aid users in making sense of the social networks being visualized.

6. Conclusion

In many ways, graphs are the obvious solution for trying to visually represent a network structure. No other graphical system functions as much as a full-time symbol of connections between nodes. Nonetheless, given the limitations of resolution and screen size of computers, social network graphs usually end up cluttered and rather illegible. This is not so much due to technical limitations in the algorithms that generate these graphs as much as it is due to a general neglect of principles of good layout and visual perception on the part of the designers of these systems. We suggest that graphs should be used as the underlying infrastructure for more sophisticated visualizations of social networks. By applying some of the basic principles of cartography to the design of social network visualization, we are bound to create more legible representations that capture this social phenomenon.

References

1. Freeman, L. C. (2000) Visualizing social networks. Journal of Social Structure, 1(1).

2. Lombardi, M. http://www.albany.edu/museum/www museum/work/lombardi/

3. Viégas, F., boyd, d., Nguyen, D., Potter, J., & Donath, J. Digital Artifacts for Remembering and Storytelling: PostHistory and Social Network Fragments. In Proceedings of HICSS-37, January 2004.

4. Ware, C. Information Visualization: Perception for Design. Morgan Kaufmann Publishers, 2000.