

A Dual-View Approach to Interactive Network Visualization*

Galileo Mark Namata Jr.
Dept. of Computer Science
Univ. of Maryland
College Park, MD 20742
namatag@cs.umd.edu

Lise Getoor
Dept. of Computer Science
Univ. of Maryland
College Park, MD 20742
getoor@cs.umd.edu

Brian Staats
Dept. of Computer Science
Univ. of Maryland
College Park, MD 20742
staatsb@mail.umd.edu

Ben Shneiderman
Dept. of Computer
Science/HCIL
Univ. of Maryland
College Park, MD 20742
ben@cs.umd.edu

ABSTRACT

Visualizing network data, from tree structures to arbitrarily connected graphs, is a difficult problem in information visualization. A large part of the problem is that in network data, users not only have to visualize the attributes specific to each data item, but also the links specifying how those items are connected to each other. Past approaches to resolving these difficulties focus on zooming, clustering, filtering and applying various methods of laying out nodes and edges. Such approaches, however, focus only on optimizing a network visualization in a single view, limiting the amount of information that can be shown and explored in parallel. Moreover, past approaches do not allow users to cross reference different subsets or aspects of large, complex networks. In this paper, we propose an approach to these limitations using multiple coordinated views of a given network. To illustrate our approach, we implement a tool called DualNet and evaluate the tool with a case study using an email communication network. We show how using multiple coordinated views improves navigation and provides insight into large networks with multiple node and link properties and types.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Visualization]: User Interfaces

*A full version of this paper and additional information on DualNet is available from <http://www.cs.umd.edu/linqs/projects/dualnet/>.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIKM'07, November 6–8, 2007, Lisboa, Portugal.

Copyright 2007 ACM 978-1-59593-803-9/07/0011 ...\$5.00.

General Terms

Design, Human Factors

Keywords

Social Networks, Coordinated Views, Interactive Graph Visualization, Information Visualization

1. INTRODUCTION

There is an increasing amount of network data made available each day ranging from explicit social networks defined by social networking sites such as MySpace and Friendster to implicit network data generated from communication and collaboration graphs such as email communications and instant messages. Due to the increasing abundance of these networks, there is growing interest in visualizing, exploring, and analyzing these networks for historical (i.e., exploring government email communications), financial (i.e., viral marketing) and legal (i.e., looking at the emails of investigated companies) reasons.

Information visualization is useful for leveraging the perceptual abilities of humans to quickly explore and understand large amounts of data in parallel. However, the task of visualizing network data is difficult by its nature. Unlike other forms of data, visualizing network data not only involves visualizing various attributes for each data point, but it also involves visualizing the links between the data points and the attributes of the links. As a result of trying to show all these pieces of information, visualizations of network data are often cluttered, suffer from occlusion and illegible labels, and are difficult to explore.

Various approaches have been proposed to address these problems including filtering, clustering techniques and different ways of zooming and laying out networks in a display [1, 5, 6, 8]. These approaches, however, focus on a single view of the network, limiting the items, links and attributes that can be shown. Moreover, past approaches only allow the user to do comparisons of different views and aspects of the data by running multiple copies of the same, or different, tools side by side. Comparisons in this manner result in significant overhead from repeated context switching be-

tween dissimilar visualizations and controls. Also, separate tools, with no direct way to follow a data item in one tool to corresponding data items in the next, make it difficult to perceive correlations from the visualizations.

To address these problems, we propose an interactive, coordinated view approach to visualizing networks. Instead of a single monolithic network view, we propose to have multiple, customizable network views. Each view is designed to show different subsets and representations of a network most appropriate to the task and data (i.e., use tree representations for acyclic subsets of the network). We also propose to create a link between these independently controlled network views to identify corresponding representations of a data item across all views. We demonstrate the utility of our coordinated view approach for analyzing a large corporate organization hierarchy and email collection.

2. OUR APPROACH

Given the limitations of using a single representation in network visualization, we propose an approach to network visualization using multiple, interactive, coordinated views of different overlapping subnetworks of a given network. Specifically, we propose that network visualization tools follow the following guidelines.

First, rather than treating the network data as a single indivisible collection, the tool should allow the user to reduce the size of the network by selecting meaningful subnetworks of the overall collection. The subnetworks can be a natural subset of the data (i.e., subnetworks created by selecting nodes belonging to a given group defined in the network), a manually generated subset of interest for a task (i.e., the set of interesting nodes and edges users have selected explicitly), or the subnetworks resulting from merging or splitting data items (i.e., multiple email addresses merged to represent the person using them).

Next, network visualization tools should allow any combination of subnetworks to be represented in separate, interactive visualizations where each subnetwork can be displayed and manipulated with the appropriate representation and controls. For example, a tree subnetwork should be displayed using a visualization for that specific type such as treemap [5] or SpaceTree [8]. Larger, more connected subnetworks might be better displayed using a node-link diagram using a force-directed layout.

Finally, the tool should support linking between the different network views. Selection of nodes and edges in one view should highlight the corresponding set of nodes and edges in the other views. This allows users to cross-reference visualizations to see how different subnetworks and attributes correlate with each other. For example, if one visualization groups people with the same manager as a single node and the other shows each person as a node, selection of a manager node in the first view will highlight all the people nodes, with that same manager, in the second view. The same is true in the reverse, where selecting a person in the second visualization will highlight the node representing the manager of that person in the first view.

3. IMPLEMENTATION

To test the utility of our approach, we implemented a tool called DualNet. DualNet is a Java application built using the open source Prefuse information visualization toolkit [4].

As seen in Figure 1, DualNet consists of two panels, showing dual views of the network. Each panel is identical, consisting of a display panel on the top and four control and information tabs on the bottom. The controls and filters for each panel are independent; options and filters applied on one panel will not be applied to the second panel. The current implementation of DualNet supports interactive dragging and zooming using mouse controls, various node-link layouts, node and link size, shape and color controls, common filter controls, as well as the ability to search, the contents of the emails used in our evaluation.

4. EVALUATION

DualNet was evaluated on a case study using the well studied Enron email collection [3, 7]. For our evaluation, we focused on a subset of the collection from 2000-2001, selected specifically because we have documented information about the titles and positions for the individuals in these email addresses [3, 7]. We also used counts of the email communications between all those individuals in 2000-2001. The network graph constructed from this data set resulted on a network consisting of 119 email addresses and 1140 directed edges representing the amount and direction of email communications between nodes.

We performed a case study of the tool with an Associate Professor at University of Maryland, College Park and a researcher at Johns Hopkins University Applied Physics Laboratory. Both users have done extensive work on the Enron dataset, particularly with regards to exploiting, identifying and using the underlying social network of the communications [3, 7, 2]. The users were given a 30 minute preview of the tool and then allowed to explore the tool for as long as they wanted, but no less than 30 minutes. During their exploration, we were available to answer any questions about the tool. The users were instructed to provide detailed feedback of bugs, feature requests, comments and criticisms of the tool. They were also requested to comment on various patterns and anomalies highlighted by the tool, as well as point out any interesting results from their exploration.

5. RESULTS

The overall feedback for the tool was positive. Both users felt that a multiple view approach was definitely appropriate for network data, specifically in the case of communication networks where there are often many different edge types. They felt that a dual interface was a cleaner and more natural approach to showing a large number of node and edge attributes. Moreover, they felt that the multiple interfaces were useful in iteratively navigating the graph to nodes and edges of interest. A reference was made to how the interface is similar to a “bird’s eye” view of the data, common in many image processing applications. They felt our interface was more powerful, however, since our view was customizable, not restricted to be of the same type, and can be at different levels of abstraction. We also received feedback about interesting aspects of the tool and data from our demonstration and from their experience with the tool which we present in this section.

5.1 Network Comparison

One feature the users liked was the ability to focus on two different parts of the same network and compare them side

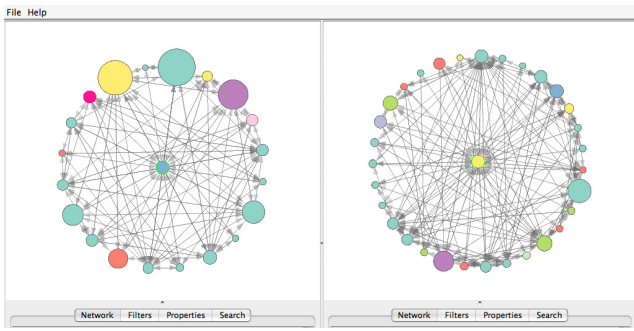


Figure 1: Comparison of the executive committee (left), and vice president ego networks (right) show differences in the size and the type of individuals each title communicates with

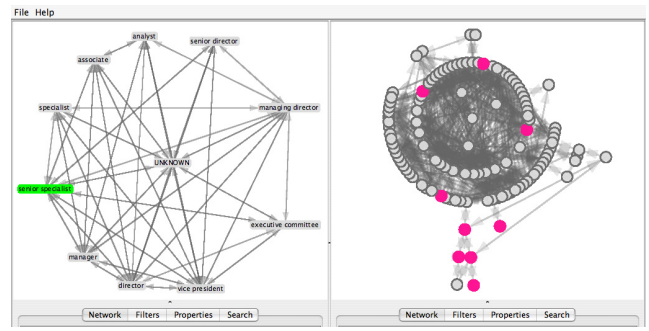


Figure 3: Applying different graph layouts and node types allows users to see that senior specialists (left) are outliers in the communication graph (right)

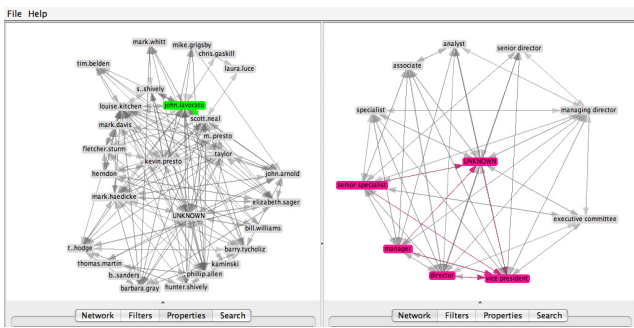


Figure 2: Selection of “john.lavorato,” green (left), from a graph of managers highlights the titles of individuals who report to John Lavorato (right)

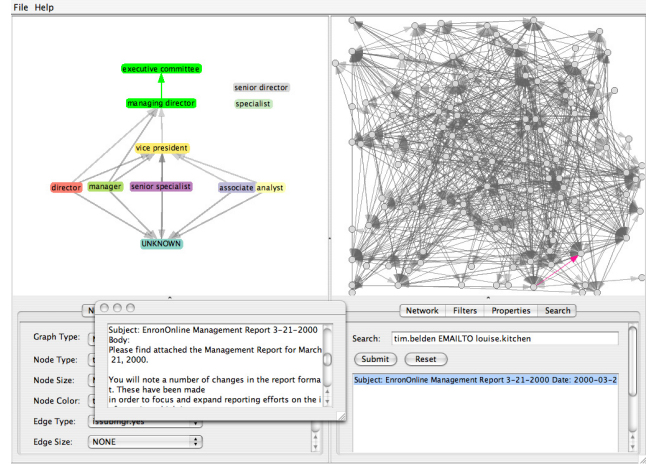


Figure 4: Title nodes shown in a tree to lines of authority (left) and email communication graph (right), separated using a force directed layout, highlight upper management communications

by side, whether by zooming in to one area of the network or filtering the two networks differently. In Figure 1, we give an example showing the ego networks of two individuals. An ego network for a node is a commonly used subnetwork in social network analysis consisting of a node and its local neighborhood, all edges and the nodes to which the node is directly connected. In this example, ego networks for two individuals were generated by representing nodes as email addresses and edges as all communications between those email addresses. All nodes and edges are removed using the node filters except those within a distance of one from the root. The node whose ego network we are interested in is set as the root in a radial tree layout. The node color corresponds to the 11 titles held by the owners of the email addresses.

The ego network on the left corresponds to an Executive Committee member, and, on the right, the ego network corresponds to a Vice President. Initially, we expected the two networks to be very similar since they are both upper management positions. We find however that they vary greatly. The Executive Committee member has a smaller ego network compared to the Vice President. Also, the Executive Committee member and Vice President communicate with different types of people. For example, the Executive Committee member rarely has contact with Specialists, illus-

trated by the lack of green nodes on the left. This provides insight to how interactions between different titles differ in the network.

5.2 Multiple Representation

A second feature our users liked was the ability to represent the same graph multiple ways and see the correspondence in the views. One representative example is shown in Figure 2 where the left view clusters nodes based on their direct manager while the right view clusters them in terms of the title. In this visualization, users are able to see which titles report to which managers. For example, selecting *john.lavorato* on the left, highlighted as green, we see the titles *Senior Specialist*, *Unknown*, *Manager*, *Director* and *Vice President* highlighted in pink on the right. This shows that in our collection, individuals of at least four known titles report directly to John Lavorato. Moreover, we note that this gives users additional information about who John Lavorato is, pointing out that this person must be in upper management. Using the search capability of the tool

to find and view all the emails sent by John Lavorato, we were able to establish that he is indeed upper management, specifically Executive Committee member.

5.3 Outlier Detection

Through the use of the coordinated views, the users felt that certain outliers in the data became more noticeable. In initial explorations of the data, we set the node colors to represent different titles and asked our users to count how many titles there were and how many of each title were available in the graph. With eleven titles spread all over a network graph of 119 nodes, they couldn't get a sense of how many of each title there was by just looking at a single view. However, by using coordinated views, setting the left panel to represent nodes as titles and keeping the right panel as email addresses, they were able to count exactly how many titles there were. By clicking on a specific title, the users could see the corresponding email addresses with that specific title stand out from all the other nodes. In the case of Senior Specialist, highlighted in green on the right of Figure 3, not only could they easily see that there are nine individuals, highlighted in pink on the right, with the title of Senior Specialist, they were also able to see that this title is an outlier in the communication graph. In the right panel, we see that Senior Specialists make up five of the six nodes, isolated in the bottom of the display, who only email one other.

5.4 Hierarchy Generation

The most convincing and compelling results during the evaluation of DualNet are with respect to what the tool shows users about the hierarchy of these communications. The combination of title and direct report information shows users a glimpse of how the management structure in Enron worked. We set the node type on the left panel to cluster nodes by title and to only show edges representing emails to direct reports (i.e., directed edges starting from node A to node B means node B is the manager of node A). We observed the subnetwork had no loops and following the convention of displaying lines of authority with organizational charts, we layout the network using a top-down node-link tree. On the right panel, we keep the nodes as email addresses and the edges as all communications between those edges. In order to minimize node overlaps and make better use of the available display space, we use a force directed layout in the right panel.

Given these settings, the tool shows on the left that the highest title in our collection is Executive Committee. This also shows that unlike initial assumption that only Directors report to Vice Presidents cited in one of the user's works [7], users found that the role of Vice President, for Enron, was more dynamic. Four different title types report to Vice Presidents. Looking further, users found that, for some reason, Associates, the lowest title defined in Enron documentation, also report to Vice Presidents. By clicking on that link, shown in green, and looking at the corresponding highlights on the right display, shown in pink on the lower right, users can see that there is only one pair of individuals, Tim Belden and Louise Kitchen, with this type of relationship. The users were then able to search for these email communications for further analysis.

6. CONCLUSIONS

Visualizing network data is a challenging task with many important applications. Static (printed) drawings can sometimes be helpful, but support for discovery requires an interactive solution to reduce complexity and enable users to selectively display components of interest. Previous approaches to an interactive solution involve zooming, filtering, clustering and layout techniques to reduce the number of overlaps and minimize the amount of data to fit in the space available. These approaches, though good, are still limited in the number of attributes they can display at one time and do not allow comparisons of different subsets and aspects of the data. To address these limitations, we proposed an approach using multiple coordinated views of the same network. Illustrating our approach using a tool called DualNet, we have shown that multiple coordinated views improve navigation and exploration of network data. We have also shown how using multiple coordinated views provide insight into the network data that would have been difficult to discover with a single view.

7. ACKNOWLEDGMENTS

We thank Chris Diehl from the Johns Hopkins Applied Physics Laboratory for the Enron organization hierarchy information, for participating in our evaluation, and for his input in the design of our system.

8. REFERENCES

- [1] A. Aris and B. Shneiderman. Network visualization by semantic substrates. *IEEE Transactions on Visualization and Computer Graphics*, 2006.
- [2] M. Bilgic, L. Licamele, L. Getoor, and B. Shneiderman. D-dupe: An interactive tool for entity resolution in social networks. In *Proceeding of IEEE Symposium on Visual Analytics Science and Technology*, 2006.
- [3] C. Diehl, L. Getoor, and G. M. Namata. Name reference resolution in organizational email archives. In *Proc. of the SIAM Conf. on Data Mining*, 2006.
- [4] J. Heer, S. K. Card, and J. A. Landay. Prefuse: a toolkit for interactive information visualization. In *Proc. of the SIGCHI Conf. on Human factors in computing systems*, 2005.
- [5] B. Johnson and B. Shneiderman. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. In *Proc. of the 2nd Conf. on Visualization*, 1991.
- [6] B. Kules, B. Shneiderman, and C. Plaisant. Data exploration with paired hierarchical visualizations: initial designs of pairtrees. In *Proc. of the 2003 Annual National Conf. on Digital Government Research*, 2003.
- [7] G. M. Namata, L. Getoor, and C. Diehl. Inferring formal titles in organizational email archives. In *Proc. of the ICML Workshop on Statistical Network Analysis*, 2006.
- [8] C. Plaisant, J. Grosjean, and B. B. Bederson. Spacetree: Supporting exploration in large node link tree, design evolution and empirical evaluation. In *Proc. of the IEEE Symposium on Information Visualization*, 2002.