

A Dual-View Approach to Interactive Network Visualization

Galileo Mark Namata
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 with each other. In this paper, we propose an approach which uses 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 types.

Categories and Subject Descriptors

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

Keywords

Social Networks, Coordinated Views, Interactive Graph Visualization, Exploratory Data Analysis, Information Visualization

1. INTRODUCTION

There is an increasing amount of network data made available each day, especially on the Internet. The data range from explicit social networks defined by social networking sites such as MySpace¹ and Friendster² to 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. The approaches include filtering, clustering techniques and different ways of zooming and laying out networks in a display [2, 7, 8, 18, 22, 27]. 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 between 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.

¹<http://myspace.com>

²<http://www.friendster.com>

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.

The rest of the paper is organized as follows. In section 2, we motivate this problem by describing it in the context of exploring an email communication network of a large corporation. We review relevant work in network visualization and coordinated views in section 3. In sections 4 and section 5, we describe our approach and illustrate our approach in a tool called DualNet. We then present the results from a case study of our approach involving an email collection in section 6 and section 7. Finally, we discuss future work in section 8 and provide our conclusions in section 9.

2. MOTIVATING PROBLEM

Recent studies estimate that the number of email messages sent in a day exceeds 2.25 billion [29]. Because it is such a major form of communication, there is significant interest both processing it more efficiently and archiving and analyzing large email collections for both historical and legal reasons. Toward this latter goal, there have been a number of tools created to work with email archives [14, 17]. In the field of information visualization, the standard approach of representing the communications as a node-link graph, with nodes representing individuals and edges representing email exchanges, has been successful at various tasks including highlighting anomalies in traffic counts and illustrating word usage in communications [14, 12, 20]. Although these tools have provided insight into the email collections, they view the communication network only as a single, large network. These tools ignore the fact that communication networks have a great deal of structure, consisting of many subnetworks resulting from differences in node types (i.e., email address, person) and attributes, as well as differences in relationship types (i.e., friendship, subordinate-manager) expressed by the edges. Moreover, they do not allow users to navigate and compare these subnetworks to provide context to the analysis of the larger graph.

An illustration of the utility of multiple network views can be seen in the following example from the widely studied Enron email collection [21]. In this collection, documentation about the collection lists Mark Taylor, a senior level executive, as one of the most active participants in the communication network. Looking solely at the communications graph, this identification is correct. As any executive would, Mark Taylor sends a large quantity of emails and a quick overview of the email content shows content consistent with this hypothesis. Simultaneously looking at a different view of the network, however, with the network arranged to emphasize the line of authority, we see that Mark Taylor seems to communicate frequently with both high-level executives and low-level employees. Moreover, by looking at the net-

work using another visualization displaying the different divisions within a company, we find that Mark Taylor also emails two very different Enron divisions frequently. Looking at these views for other individuals in the organization, specifically with other senior level executives, we can establish this is uncommon.

In this case, identification of this anomaly using multiple views of the same network revealed that there were *two* Mark Taylors, a Mark E. Taylor and a Mark A. Taylor, in the company. Mark E. Taylor was a Vice President in the ECT division while Mark A. Taylor was an employee from the legal department of the EES division. The high similarity of their names resulted in the two individuals, and their email accounts, being mistakenly merged in the distributed Enron collection.

3. RELATED WORK

In this section, we discuss related work in network visualization and discuss the strengths and limitations of current approaches. We then discuss related work on the benefits of applying multiple coordinated views in data exploration.

3.1 Network Visualization

Network visualization is a challenging task that becomes increasingly difficult as network’s size and density increases [2]. The most commonly used approach to address these problems is by applying techniques to layout the nodes and edges [10, 31, 6, 2] based on network properties and tasks. For example, force directed layouts [10] are useful for spreading out the network, as well as identifying clusters and outliers in the data. Similarly, semantic substrates are useful for networks which can be separated into meaningful regions [2]. Layouts still degrade with very large and dense networks so there has also been work in reducing the number of visible nodes and links that need to be displayed.

Zooming [4] and fisheye [23] techniques have been used to look at specific parts of a network with greater detail than others. Multiscale [13] algorithms have also been used to create abstractions of the network to improve runtime and display. These techniques are usually only useful when looking at items closely positioned in a layout and do not allow you to select specific subsets. Moreover, these techniques can obscure the global structure of the network [26].

Filtering has also been used with great success. Vizster [15] does filtering on a social network by only showing a few nodes initially and allowing the user to add and remove nodes and edges from there. Although this allows a user to focus in on specific nodes, it does not allow the user to get an overview of the whole collection, to see outliers and clusters, as advocated by the widely used Visual Information Seeking Mantra [28], “Overview first, zoom and filter, then details on demand.” In contrast, NicheWorks [30] begins with an overview first and uses dynamic query filters to reduce the size of the network. NicheWorks is successful at exploring networks with hundreds of thousands of nodes. Although filtering is unavoidable for most large networks, we believe we can reduce the amount of filtering needed by using coordinated views to show more information in parallel.

There has also been work on alternate ways to show networks besides the commonly used node-link representation. Ghoniem et al. [11] shows that matrix representations of networks are more useful for certain network exploration tasks like finding specific nodes and links. Treemaps [18] and SpaceTree [27] have been used successfully by applications showing tree data, while Netlens [19] has been used to iteratively navigate actor-content networks.

The previous approaches listed in this section individually have specific strengths and weaknesses on specific types of networks. We believe that the combination of these approaches in multiple coordinated views will allow us to apply the most effective techniques to different views and subsets of the network to further enhance exploration.

3.2 Multiple Coordinated Views

Multiple coordinated views are a powerful approach to exploring information. Coordinated views have been shown to improve user performance and allow discovery of unforeseen relationships [3, 25]. Consequently, a number of tools have been created using multiple coordinated views. One of the most successful is the commercial product Spotfire [1] whose use of coordination between common visualizations like scatterplots and charts has been used successfully in the biological domain. Spotfire, however, does not support visualization of network data. SocialAction [26] is an application that supports multiple coordinated views on networks. The application is specific to social networks, though, and only uses coordinated views to highlight an item selected in a ranked list, scatterplots and overview matrix.

The most similar work to our approach that we are aware of is PairTrees [22] which uses coordinated views between treemaps, SpaceTree [27] and geographical maps to display a federal statistics dataset. PairTrees is specific to visualizing hierarchical network data while we apply multiple coordinated views on general networks.

4. 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 con-

trols. For example, a tree subnetwork should be displayed using a visualization for that specific type such as treemap [18] or SpaceTree [27]. Larger, more connected subnetworks might be better displayed using a node-link diagram using a force-directed layout. We note that the visualizations do not have to be of distinct, non-overlapping subnetworks. For example, multiple node-link diagrams of the same subnetwork, can re-represent the same subnetwork with the same layout but using different node colors to represent different node attributes. Such a display would allow users to see more attributes in parallel than a single display could support and may lead to identifying interesting correlations among the different attributes.

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.

5. 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 [16]. 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 four tabs (Network, Filters, Properties, Search) provide users with information and controls for manipulating the network. The Network tab provides users with options for displaying the network. The Graph Type option allows users to choose how to visually represent the network. For this version of DualNet, we limit ourselves to node-link representations of the data and thus Graph Type lists only layout options appropriate for node-link diagrams. We support many of the popular network visualizations including force directed [10], radial tree [31] and node-link tree layouts [6]. The option, Node Types, allow for representing the nodes as clusters based on a single string attribute. For example, multiple nodes representing email addresses can be merged to one name node by using the name of the individual who owns those email addresses. The Node Color option allows users to specify varying colors for different string or numeric attribute values. In the case of numeric values, the color represents the magnitude of the values with blue representing low values and red representing high values. The Node Size option supports changing the size of each node based on the values of numeric attributes. The same set of options are also available for edges in this tab.

The Filters tab gives the user the ability to filter the network

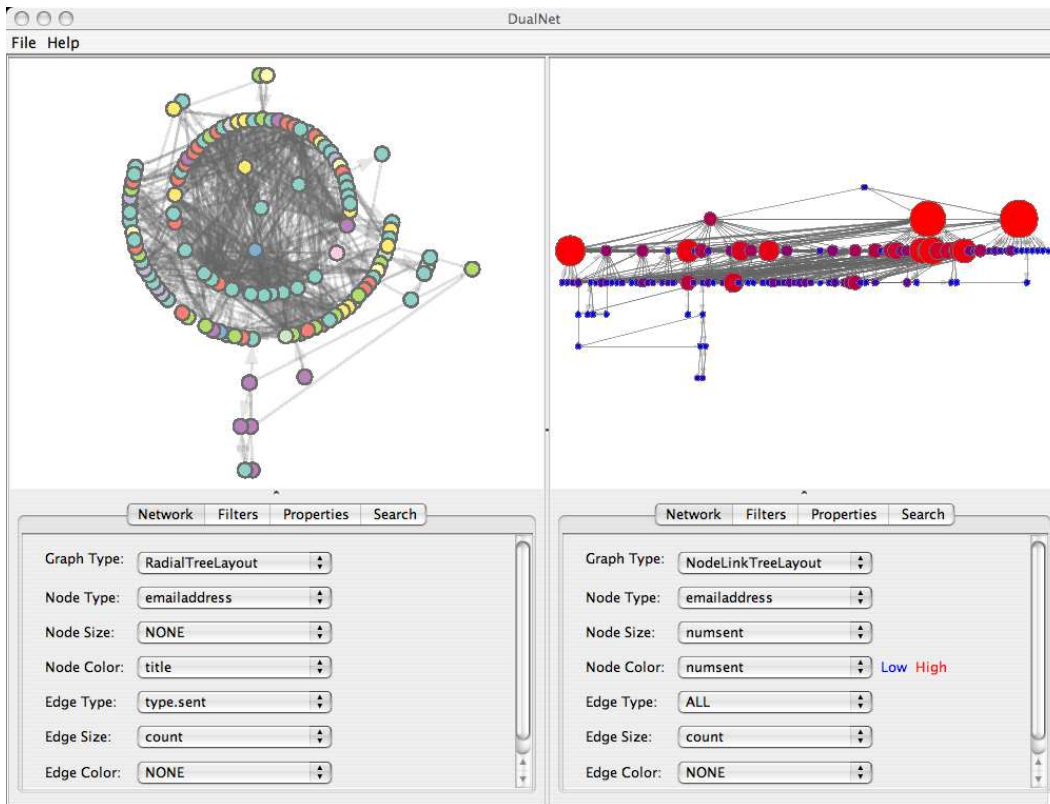


Figure 1: An email communication network represented in two linked views using DualNet. The left panel is a Radial tree view with the email nodes colored by title and size varied by the amount of messages sent. The right panel is a Node-Link tree view with color and size varied by the amount of messages sent.

based on attributes of the data. Options include filtering by node degree and by value range for any numeric attributes of the nodes and edges. DualNet also provides a mechanism for re-centering the display as appropriate in some layouts. For example, when using the radial tree layout, re-centering will place the currently selected node as the root in the center of the display. There is also a “Show Only Highlighted” option in the filters tab. When selected, the current display will only show nodes and edges corresponding to the set of nodes and edges selected in the other view.

The Properties tab is used to display additional information about the currently selected node or edge. For nodes, the Properties tab shows all the attributes of the node. For edges, it shows the edge properties and also the properties of the nodes corresponding to that edge.

The Search tab allows users to search for a specific node, by attribute value, in the data. The nodes matching the queries are shown in the results list and are highlighted in pink in the display when selected. DualNet also supports searching over supplementary data relevant the network at hand. For example, the evaluation described in the next section uses a network from an email archive. DualNet allows users to search for all emails sent by a single individual or as a result of an exchange between two individuals. In this case, a list of email messages is returned by the search and the selection

of one email in the list displays that email’s contents in a separate window.

DualNet also supports a number of standard controls in the network display itself. Hovering over nodes and edges highlights those items in yellow, while tool tips on nodes will display the node value specified in Node Type. DualNet also supports showing nodes as labels, instead of circles. This is done automatically when there are only a few nodes (less than 25) in the display. In addition, users can navigate each view with the mouse by panning and zooming in addition to dragging individual nodes. Finally, we support brushing between the two panels. Selection of a node or edge in one panel will highlight the selected item in green and corresponding items in the other panel in pink.

6. EVALUATION

DualNet was evaluated on a case study using the well studied Enron email collection [21]. For our evaluation, we focused on a subset of the collection containing email addresses from 2000-2001, selected specifically because we have documented information about the titles and positions for the individuals who used these email addresses [9, 24]. 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

Table 1: Enron Email Address Attributes

<i>Attribute</i>	<i>Description</i>
emailaddress	Internal Enron employee email address Removed @enron.com suffix for readability
name	Name of the person using the email address
title	Title of that individual using the email address from 2000-2001
numsent	Number of emails sent by the email address in 2000-2001
numreceived	Number of emails received by the email address in 2000-2001
numtotal	Total number of emails sent and received by this email address in 2000-2001
mgremail	Direct manager of the person using the email address

Table 2: Enron Email Exchange Attributes

<i>Attribute</i>	<i>Description</i>
type	Edge type (to, from, total)
count	Number of emails of this type exchanged in 2000-2001.
issubmgr	Set to yes if this edge represents communication sent by a subordinate to its manager

and direction of email communications between nodes. The node and edge properties we used are given in Table 1 and 2, respectively.

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 identifying and using the underlying social network of the communications [9, 24]. The user from University of Maryland also has experience in network visualization on the specific task of entity resolution [5].

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.

7. 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 inter-

face 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. We present those results in this section, as well as additional critiques and requested features in section 7.5.

7.1 Network Comparison

One feature the users liked was the ability to focus on two different parts of the same network and compare them side by side, whether by zooming in to one area of the network or filtering the two networks differently. In Figure 2, 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, illustrated by the lack of green nodes on the left. This provides insight to how interactions between different titles differ in the network.

7.2 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 3 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 different titles, given one group of unknown 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

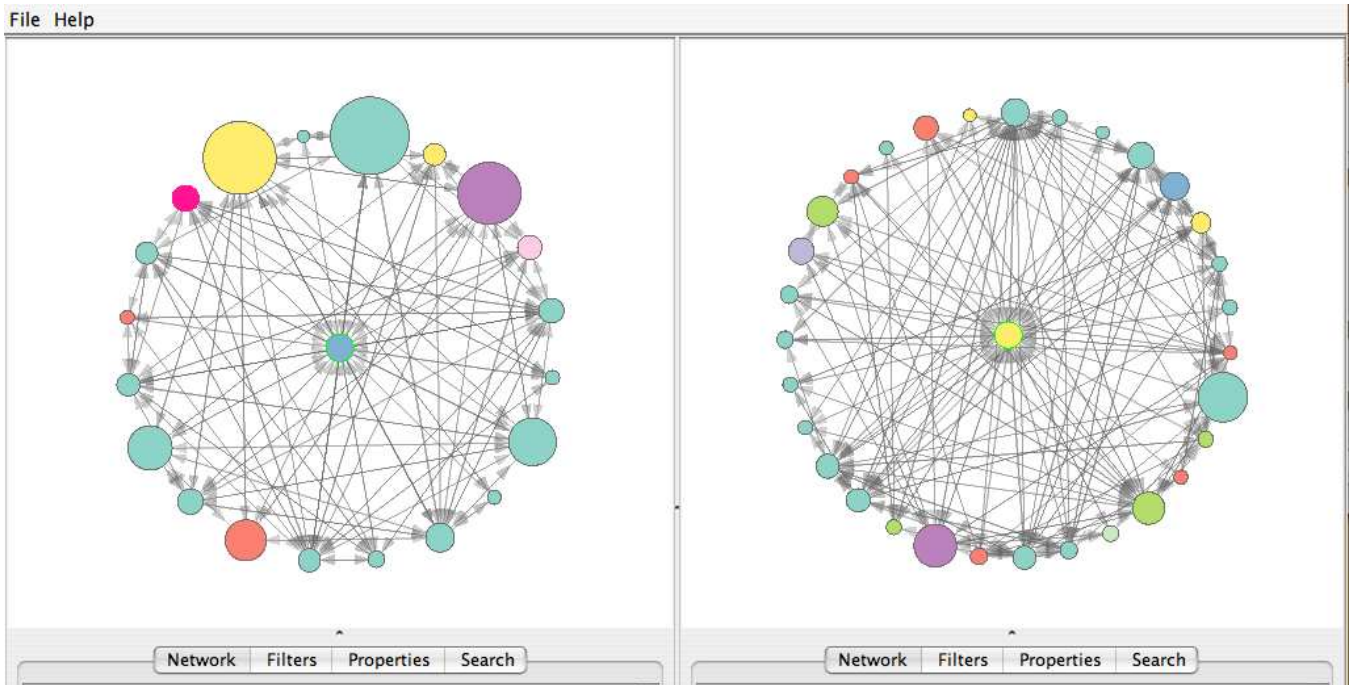


Figure 2: Comparison of the executive committee (left), and vice president ego networks (right) show differences in the size and the type of individuals each title communicates with

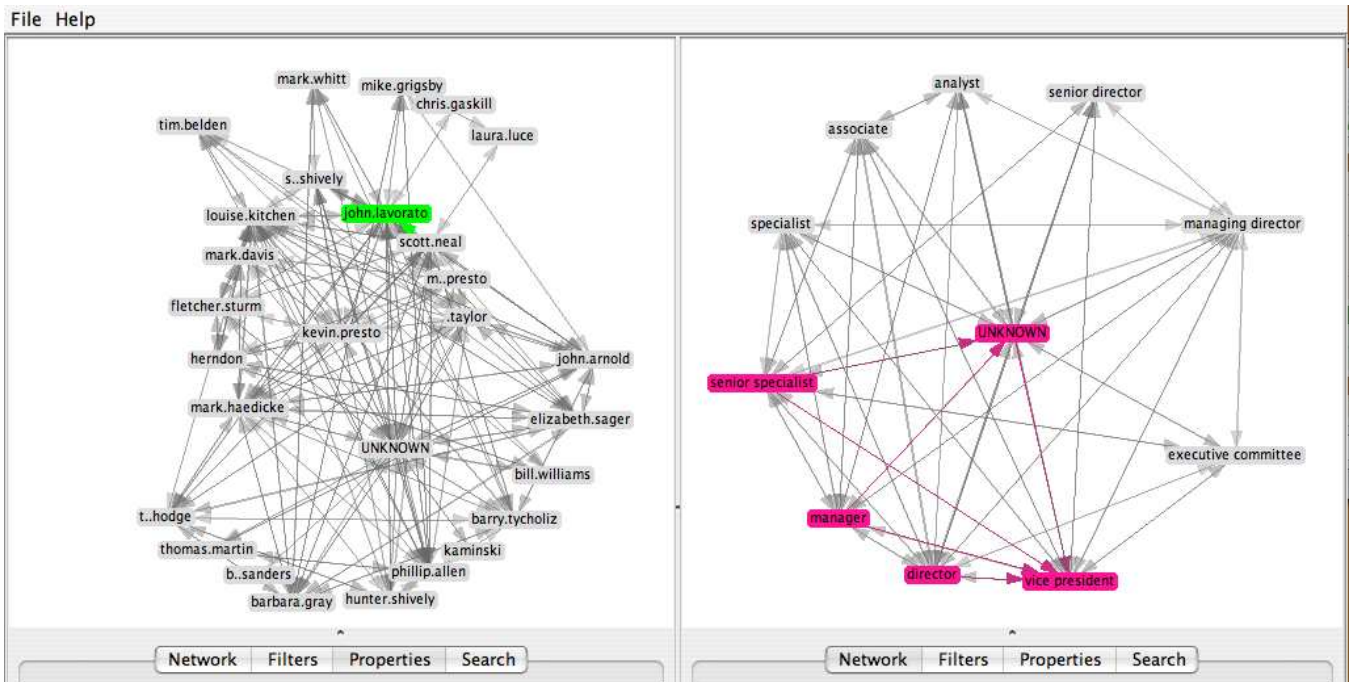


Figure 3: Selection of "john.lavorato," green (left), from a graph of managers highlights the titles of individuals who report to John Lavorato (right)

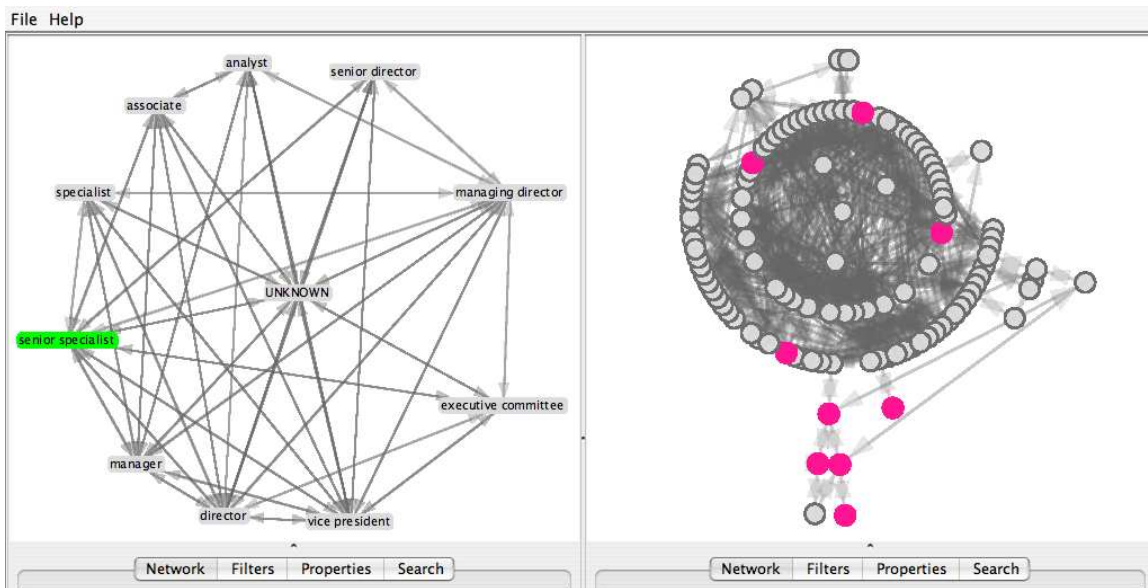


Figure 4: Applying different graph layouts and node types allows the user to see that the senior specialists (left) are outliers in the communication graph (right)

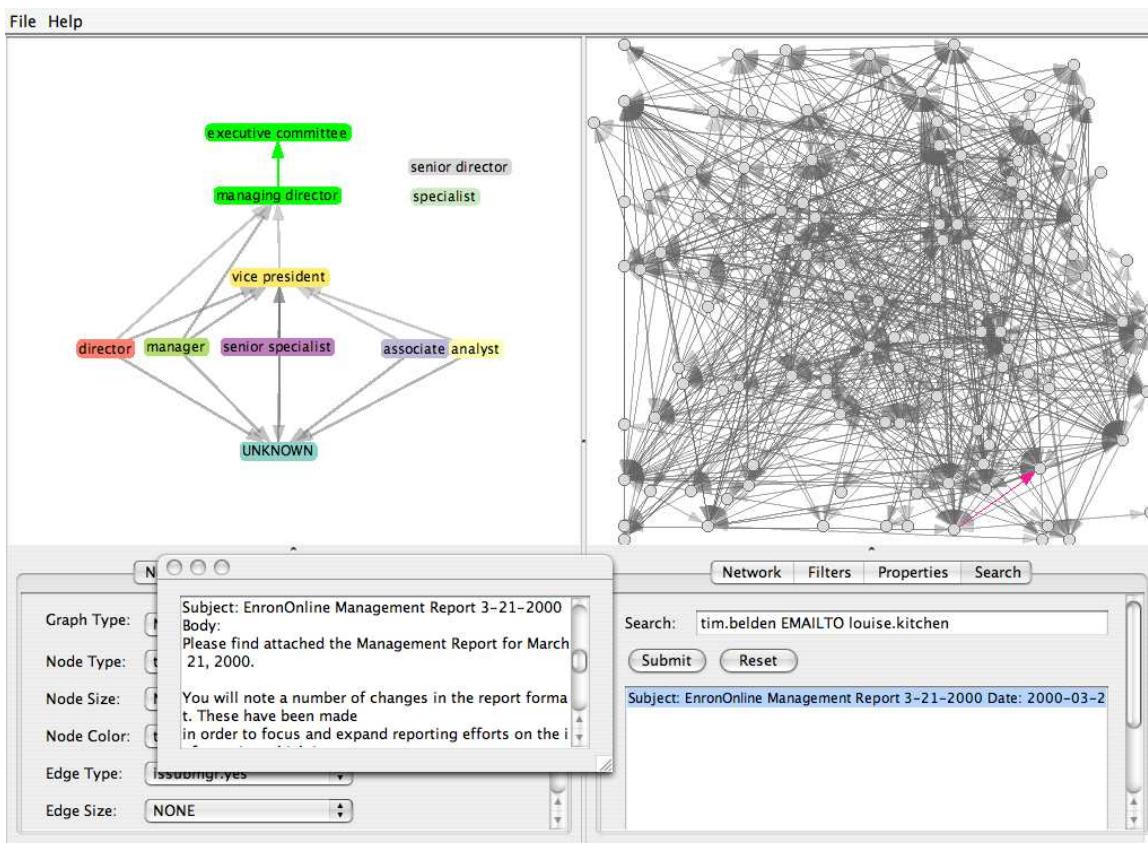


Figure 5: Title nodes shown in a tree to lines of authority (left) and email communication graph (right), separated using a force directed layout, quickly highlight upper management communications

upper management, specifically Executive Committee member.

7.3 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 4, 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.

7.4 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. This is the first time title and direct report information were put together for the analysis of this collection and this was something which both users were eager to see. The combination 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 [24], 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.

7.5 Criticisms

We received some criticism about the tool which we discuss here. First, although they felt that the tool allowed for general exploration, they felt that more specific data representations would make the tool more effective. For example, in case of this data set, developing our own layout algorithm, rather than using general network layout algorithms, to automatically arranging the nodes in terms of power, with higher ranked individuals shown above their subordinates, would be useful. Another criticism was that the tool currently only allow one node or edge to be selected at a time. Our users felt that the capability to select whole groups of nodes would allow for better analysis. On the same note, they felt that the color highlighting may not be enough, especially for larger networks where nodes may be very small. Providing a link or line between the two views or using some sort of animation, linking corresponding selections, might be needed. We plan on addressing these issues in future versions of our tool.

8. FUTURE WORK

In addition to the suggestions described above, we also plan on making additional changes to make the tool more general. The current version of DualNet only allows for node-link diagram visualizations. We would also like to explore other representations such as matrix-based [11] and treemaps [7]. Similarly, we want to allow for more than just two views of the network to give greater flexibility to users. The current implementation of DualNet provides users with many parameters for each view. A useful addition would be to provide a save feature to allow users to save the state of each view given the current setting of parameters.

We would also like to do a more controlled study comparing our approach to other network visualization tools to discover the strengths and weaknesses of each. Finally, we want to explore other richer and larger network data to evaluate our approach on those networks.

9. 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.

10. 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. We also thank Mustafa Bilgic, Robin Berthier, Tugrul Ince, David Rouff, and Elena Zheleva for their helpful comments on earlier drafts.

11. REFERENCES

- [1] C. Ahlberg. Spotfire: an information exploration environment. *SIGMOD Record*, 1996.
- [2] A. Aris and B. Shneiderman. Network visualization by semantic substrates. *IEEE Transactions on Visualization and Computer Graphics*, 2006.
- [3] M. Q. W. Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the working conference on Advanced visual interfaces*. ACM Press, 2000.
- [4] B. B. Bederson, J. Grosjean, and J. Meyer. Toolkit design for interactive structured graphics. *IEEE Transactions on Software Engineering*, 2004.
- [5] 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.
- [6] C. Buchheim, M. Junger, and S. Leipert. Improving walker's algorithm to run in linear time. In *Proceedings of the International Symposium on Graph Drawing*, 2002.
- [7] M. Burch and S. Diehl. Trees in a treemap: visualizing multiple hierarchies. In *Proceedings of the Conference on Visualization and Data Analysis*, 2006.
- [8] S. K. Card, B. Suh, B. A. Pendleton, J. Heer, and J. W. Bodnar. Time tree: Exploring time changing hierarchies. In *Proceedings of the IEEE Symposium On Visual Analytics Science And Technology*, 2006.
- [9] C. Diehl, L. Getoor, and G. M. Namata. Name reference resolution in organizational email archives. In *Proceedings of the SIAM Conference on Data Mining*, 2006.
- [10] P. A. Eades. A heuristic for graph drawing. In *Congressus Numerantium*, 1984.
- [11] M. Ghoniem, J.-D. Fekete, and P. Castagliola. A comparison of the readability of graphs using node-link and matrix-based representations. In *Proceedings of the IEEE Symposium on Information Visualization*, 2004.
- [12] F. Grippa, A. Zilli, R. Laubacher, and P. A. Gloor. Email may not reflect the social network email may not reflect the social network. In *Proceedings of the International Conference NAACOS*, 2006.
- [13] D. Harel and Y. Koren. A fast multi-scale method for drawing large graphs. In *Proceedings of the 8th International Symposium on Graph Drawing*, 2000.
- [14] J. Heer. Exploring enron. <http://jheer.org/enron>, May 2007. [Online; accessed 20-May-2007].
- [15] J. Heer and D. Boyd. Vizster: Visualizing online social networks. In *Proceedings of the 2005 IEEE Symposium on Information Visualization*, 2005.
- [16] J. Heer, S. K. Card, and J. A. Landay. Prefuse: a toolkit for interactive information visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2005.
- [17] IBM. Commonstore. [Online; accessed 20-May-2007].
- [18] B. Johnson and B. Shneiderman. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. In *Proceedings of the 2nd Conference on Visualization*, 1991.
- [19] H. Kang, C. Plaisant, B. Lee, and B. B. Bederson. Exploring content-actor paired network data using iterative query refinement with netlens. In *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*, 2006.
- [20] P. S. Keila and D. B. Skillicorn. Structure in the enron email dataset. In *Proceedings of the SDM Workshop on Link Analysis, Counterterrorism and Security*, 2005.
- [21] B. Klimt and Y. Yang. Introducing the enron corpus. In *Proceedings of the Conference on Email and Anti-Spam*, 2004.
- [22] B. Kules, B. Shneiderman, and C. Plaisant. Data exploration with paired hierarchical visualizations: initial designs of pairtrees. In *Proceedings of the 2003 Annual National Conference on Digital Government Research*, 2003.
- [23] J. Lamping and R. Rao. The hyperbolic browser: a focus + context technique for visualizing large hierarchies. *Journal of Visual Languages and Computing*, 6, 1999.
- [24] G. M. Namata, L. Getoor, and C. Diehl. Inferring formal titles in organizational email archives. In *Proceedings of the ICML Workshop on Statistical Network Analysis*, 2006.
- [25] C. North and B. Shneiderman. Snap-together visualization: a user interface for coordinating visualizations via relational schemata. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, 2000.
- [26] A. Perer and B. Shneiderman. Balancing systematic and flexible exploration of social networks. In *Proceedings of the IEEE Symposium on Information Visualization*, 2006.
- [27] C. Plaisant, J. Grosjean, and B. B. Bederson. Spacetree: Supporting exploration in large node link tree, design evolution and empirical evaluation. In *Proceedings of the IEEE Symposium on Information Visualization*, 2002.
- [28] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, 1996.
- [29] Verisign. Verisign internet security intelligence briefing. <http://www.verisign.com/static/030910.pdf>, June 2005. [Online; accessed 20-May-2007].
- [30] G. J. Wills. Nicheworks - interactive visualization of very large graphs. In *Proceedings of the 5th International Symposium on Graph Drawing*, 1997.
- [31] K.-P. Yee, D. Fisher, R. Dhamija, and M. A. Hearst. Animated exploration of dynamic graphs with radial layout. In *Proceedings of the IEEE Symposium on Information Visualization*, 2001.