Orderly Analysis of Social Visualizations

Adam Perer
Human-Computer Interaction Lab &
Department of Computer Science
University of Maryland
College Park, MD 20742 USA
adamp@cs.umd.edu

Ben Shneiderman
Human-Computer Interaction Lab &
Department of Computer Science
University of Maryland
College Park, MD 20742 USA
ben@cs.umd.edu

CHI 2006, April 22–27, 2006, Montreal, Canada.
ACM 1-xxxxxxxxxxxxxxxxxx.

Abstract

Recent advances in technology have allowed people to shift many of their communication practices online. These practices, such as email, IM, VoIP, and video conferencing, can be archived to serve as digital artifacts from which social information, such as social networks, can be extracted. Social network analysis has emerged as a powerful method for understanding the importance of relationships among people.

Visualizing social networks can allow analysts to make structural discoveries about the networks and individuals within. However, interpreting these visualizations is challenging because: (1) it is difficult to comprehend the characteristics and structure of networks when there are many edges and nodes, and (2) current systems are often a medley of statistical methods and overwhelming visual output which leaves many analysts uncertain about how to explore in an orderly manner. This results in exploration that is largely opportunistic. The contribution of our work is an interface that supports systematic analysis of social networks. Our approach uses a rank-by-feature framework that enables users to better understand the structure of networks and the social groups within.

1D Rankings

Betweenness Centrality

The number of shortest paths between pairs of nodes that pass through a given node.

Rank	Node	Degree	Type
350.0	Colombia	11	Country
182.0	Mexico	8	Country
176.0	PM	2	Terrorist Group
133.5	FARC	3	Terrorist Group
112.0	Venezuela	6	Country
81.0	Tupac Amaru Movement	2	Terrorist Group
57.0	Peru	3	Country
57.0	Panama	3	Country
52.5	ELN	2	Terrorist Group
0.0	PRD Activists	1	Terrorist Group
0.0	EPL	1	Terrorist Group
0.0	Varela Antiguerrilla Group	1	Terrorist Group
0.0	ERP	1	Terrorist Group
0.0	Peasant Self-Force PM	1	Terrorist Group
0.0	PRI	1	Terrorist Group
0.0	PRI Activists	1	Terrorist Group
0.0	ACCU	1	Terrorist Group
0.0	Che Guevara Guerrillas	1	Terrorist Group
0.0	Colombian Guerrillas	1	Terrorist Group
0.0	National Police	1	Terrorist Group
0.0	EZLN	1	Terrorist Group
0.0	Sendero Luminoso	1	Terrorist Group
0.0	PRI PM	1	Terrorist Group
0.0	M-19	1	Terrorist Group
0.0	Peasant Self Defense	1	Terrorist Group
0.0	Justice Army	1	Terrorist Group
0.0	Soldiers	1	Terrorist Group
0.0	Carlos Toledo Plata	1	Terrorist Group
0.0	Death to Rustlers	1	Terrorist Group
0.0	Columbia PM	1	Terrorist Group
0.0	EPR	1	Terrorist Group

Figure 1. A ranked list of nodes, ordered by betweenness centrality.

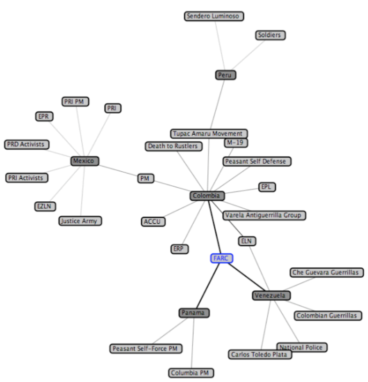

Figure 2. The selected node above is highlighted and its edges are colored in decreasing opacity to illustrate the feature.

Keywords

social networks, information visualization, rank-by-feature framework

ACM Classification Keywords

H.1.2 [Models and Principles]: User/Machine Systems – human factors; H.5.3 [Information Interfaces and Presentation]: Group and Organizational Interfaces – theory and models.

Introduction

Social network analysis has emerged as an important research perspective because it focuses on the importance of relationships among interacting individuals. The social network analysis model allows researchers to conceptualize structure as patterns of relationships, witness "flow" of information between individuals and their relational ties, and understand how an individual is influenced by a structural environment [5].

Social networks can be inferred from many of our daily communication habits. For instance, networks can be constructed based on to whom we send emails, which people we list on our Blogrolls, or who we publicly articulate as our friends on networking sites like Friendster and MySpace.com.

The unit of analysis in social networks is not the individual, but instead a collection of individuals and the ties between them. As the number of individuals and ties increase, so does the difficulty in interpreting the network. The networks can be measured quantitatively using techniques from both sociology and graph theory. However, current systems that provide such measurements are often a patchwork of statistical methods with overwhelming visual output. The interaction between these two is often not interlinked and impedes exploratory data analysis.

Our contribution is a tool that supports orderly analysis of social networks so analysts can make measurable and systematic progress when exploring the networks. When we use the term "orderly analysis", we mean an approach that allows different analysts to arrive at similar discoveries.

Analyzing Social Networks

Constructing visual images of social networks has provided analysts with insights about the structure of a network, as well as being an aid for communicating network phenomena [1]. However, as network complexity increases, illegibility of networks increases as well. Much attention has been devoted to drawing networks in an aesthetic manner so that humans can cognitively perceive the structure. However, even with the most advanced graph layout algorithms involving force-directed or spring-embedded models to reduce edge crossings, the networks become impossible to discern without interactive operations on the nodes, such as filtering or manual placement.

Our approach to improve the ease of analysis is to employ a rank-by-feature framework [3]. In a rank-by-feature framework, users can select an interesting ranking criterion (also called a "feature") and all nodes and relationships are ranked according to that criterion. Seo and Shneiderman followed this framework to create a successful tool for exploring multidimensional data called Hierarchical Cluster Explorer (HCE), which biologists have used to understand gene activity from microarray data [3]. Data analysts were able to

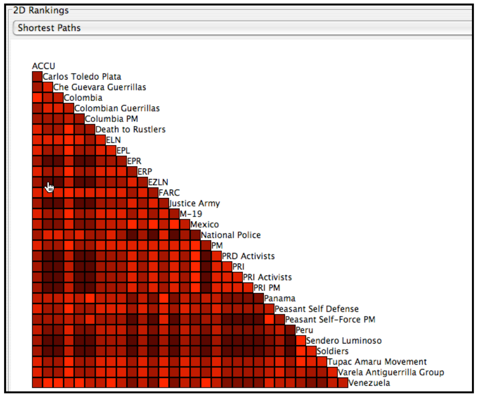

Figure 4. A matrix colored based on length of the shortest path for each dyadic relationship in the network. A half-matrix is displayed here because the graph is undirected.

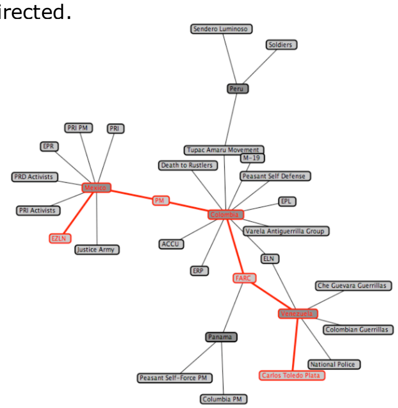

Figure 3. When a cell is selected in the relationship matrix, the feature is visually explained. In this case, the shortest path between two nodes is displayed.

interpret the many dimensions by selecting a feature that interested them, such as correlation, gaps, or outliers, and finding important data points or clusters.

We have built a brand new tool inspired by the success of HCE to improve analysis for social networks. We illustrate several scenarios here:

Entity Rankings

Analysts can quickly apply a ranking on every node or edge in the network to see the distribution across the network. For example, Figure 1 shows a ranking of all nodes by their betweenness centrality. Analysts might choose this ranking feature if they were interested in finding gatekeepers in the network. This measure is based on the number of shortest paths between nodes that pass through it. Features such as this are complicated statistical measures, so the system can highlight the network to provide a visual explanation. For instance, in Figure 2, a selected node is highlighted in blue. All of the minimum paths that pass through this node are colored according to the length of the route, with the edge's opacity decreasing for each level of distance. This coloring scheme explains the betweenness calculation without need for a formula. Even for analysts with intimate knowledge of these features, the visualization allows users to find nodes that fit the patterns they are looking for by quickly iterating through the list.

Relationship Rankings

Since the primary units of analysis in social networks are relationships, and not individual entities, our system displays each dyadic relationship in a two-dimensional matrix. In Figure 4, a matrix is shown with its dimensions equal to the number of nodes present in the network. Since this network is undirected, only half is shown. In this example, the "shortest path" feature is selected, so each cell in the matrix is colored according to the length of the shortest path between each pair of nodes. The coloring ranges from red for short paths to black for long paths, and white if no such path exists. Visual explanations are also featured here, so when the user selects a cell, the shortest path is highlighted (e.g. Figure 3). Just as a large network becomes visually overwhelming, so will a matrix, as its size increases quadratically with each new node. To address this issue, we provide the user with the ability to analyze subgroups of the network.

Cohesive Subgroups

We provide interactive procedures for merging nodes into cohesive subgroups. Users can select a feature, such as components in an unconnected network, a community algorithm based on link structure, or attributes of the node. In Figure 5, a community algorithm is selected and users can interactively adjust the parameter by dragging a slider and seeing the subgroups update in real time, an approach inspired by [2] and [3]. Once users are comfortable with the subgroups displayed, the 1D and 2D ranking displays can be reduced to only show the nodes and edges of a selected subgroup, or shrink an entire subgroup to be seen as one entity. This allows analysts to both understand the relationships inside a group of interest and compare subgroups with one another.

Ego-Centered Exploration

Analysts can also explore a network by focusing on one node, the node's neighbors, and the ties among them. An ego network of depth 2 is shown for a selected node in Figure 6. Users can interactively increase the depth

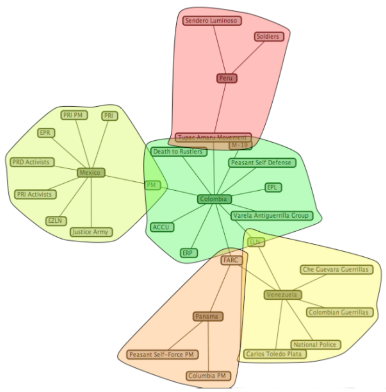

Figure 5. Colored bubbles surround the subgroups of the network. For a community feature, the user can interactively expand and shrink the groups by moving a slider bar. This allows the explorer to interactively set the parameter of the algorithm that fits the network best.

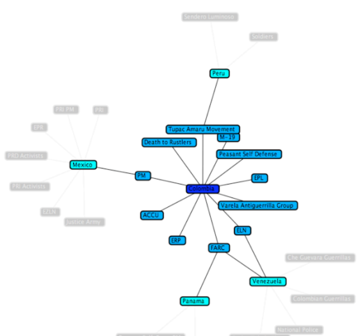

Figure 6. The ego network of a selected node with depth 2. The nodes' saturation decreases as distance increases from the ego. A user can expand or shrink the depth of the network by dragging a slider bar. All other nodes fade to gray and can be removed.

of the neighborhood by dragging a slider bar. The 1D and 2D ranking displays can be reduced to show only the nodes and edges of the ego network.

Social network visualizations were designed to uncover two kinds of patterns: social positions (nodes linked in the network in similar ways) and social groups (nodes closely linked to one another) [1]. We believe our rank-by-feature approach allows analysts to discover both of these patterns in a systematic way, making our system more effective than a standard node-link visualization system.

Future Work

We are working with sociologists to develop and append this tool to ensure it supplies useful measures of social networks. By making sociologists an integral part of an iterative design process, we will keep our focus on supporting analysts to uncover the important information in networks. These partners are also experts in the most popular systems used for social network analysis today, which can act as a barometer to measure our system against.

Social network analysis is a serious research effort and cannot be replicated easily in the form of small user studies. We intend to run a series of longitudinal case studies with sociologists to evaluate the effectiveness of our system.

Conclusion

By visualizing social networks, analysts gain the opportunity the opportunity to discover new structural insights and communicate their findings [1]. However, social network visualizations with many nodes and edges can be very challenging to interpret. In this paper, we suggest that a rank-by-feature framework improves the ability for exploratory data analysis on social networks. By providing an orderly process for navigating social networks, we believe we improve the opportunities for discovering insights and allow different analysts to reach the same conclusions. Furthermore, we believe that the rank-by-feature framework will provide similar benefits for other types of social visualizations.

Acknowledgements

We would like to thank Dr. Alan Neustadtl for many influential ideas. This work is supported by the Joint Institute for Knowledge Discovery at the University of Maryland.

References

[1] Freeman, L.C., Visualizing Social Networks. Journal of Social Structure, 2000.

[2] Heer, J. and boyd, d., Vizster: Visualizing Online Social Networks. IEEE Information Visualization 2005.

[3] Seo, J. and Shneiderman, B., "A Rank-by-Feature Framework for Interactive Exploration of Multidimensional Data," Information Visualization, Vol. 4, No. 2, pp. 99-113, 2005

[4] Viegas, F. boyd, d., Nguyen, D., Potter, J., and Donath, J. Digital Artifacts for Remembering and Storytelling: PostHistory and Social Network Fragments. HICSS 2004.

[5] Wasserman, S., and Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press.