Visualization and Analysis of Small-World Email Networks

Xiaoyan Fu*
National ICT Australia

Seok-Hee Hong†
National ICT Australia
School of IT,
University of Sydney,
Australia

Nikola S. Nikolov‡
Department of CSIS,
University of Limerick,
Ireland

Xiaobin Shen§
Department of Civil and
Environmental Engineering,
University of Melbourne,
Australia

Yingxin Wu¶
National ICT Australia
School of IT,
University of Sydney,
Australia

Kai Xu‖
National ICT Australia

ABSTRACT

This poster presents various methods for visualization and analysis of small-world email networks with various perspectives: visualization on the surface of a sphere to reveal the relationships between different groups, a 2.5D hierarchical visualization method combined with the centrality value of nodes to analyze important people, a 2.5D visualization method for temporal email networks to analyze the evolution of email relationships between people changing over time, and an ambient display method for finding social circles derived from the email network. These methods are applied to various email networks inside a research organization.

Keywords: Visualization, Email network, SOM, Centrality, Ambient Display

Index Terms: H.5.2 [INFORMATION INTERFACES AND PRESENTATION]: User Interfaces—Theory and methods

1 INTRODUCTION

Visualization of email networks has been widely applied to assist users in understanding email data and analyzing the social networks they reflect. For example, an email exchange network existing in email conversation has been studied [5]. In [3], an email network was used to study information seeking and workplace collaboration. Another interesting development of email visualization is the application of ambient display. An example is the "Info-Lotus" [7].

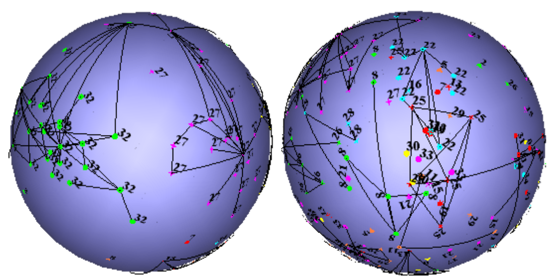
In this poster, we consider *small-world* email networks. In general, visualizing small world networks is challenging due to their short diameter [4]. We present various methods for the visualization and analysis of small-world email networks with different perspectives: visualization on the surface of a sphere to reveal the relationships between different groups, a 2.5D hierarchical visualization method combined with the centrality value of nodes to identify important people, a 2.5D visualization method for temporal email networks to analyze the evolution of email relationships over time, and an ambient display method for finding social circles that may reflect informal collaboration.

These methods are applied to various data sets from a research organization. The data used here was collected from July to August 2004. The original email network contains 604 nodes and 8605 edges. The diameter of the network is 5, and the average path length is merely 2.2, which means that the email network is an *ultra small-world* network with small diameter and short distance between any pair of nodes.

*e-mail: xiaoyan.fu@nicta.com.au
†e-mail:shhong@it.usyd.edu.au
‡e-mail:nikola.nikolov@ul.ie
§e-mail:xrshen@unimelb.edu.au
¶e-mail:chwu@it.usyd.edu.au
‖e-mail:kai.xu@nicta.com.au

2 SPHERICAL DRAWING OF AN EMAIL NETWORK

The first method uses a self-organizing map to visualize an email network on the surface of a sphere. The self-organizing map [2] is an unsupervised competitive artificial neural network. It projects high-dimensional data onto a low-dimensional space. The projection preserves the topological relationships of the original data. The spherical surface provides a natural fisheye effect which shows the focus point with details and shows other points of the image with less detail.



(a) Two research groups. (b) The management group

Figure 1: Spherical drawing

In Figure 1, only edges representing intra-group communications are shown to reduce visual complexity. Inter-group communications can be observed by the closeness between groups: the closer the groups, the more communication between them. Some communication patterns can be seen. People in the same research group, such as 27 and 32, tend to form the same communication pattern. Their intra-group communication edges almost form cliques. See Figure 1(a). However, people in different management groups such as 8 (The CEO office), 22 (Finance) and 25 (Human Resource) are mixed together. See Figure 1(b).

3 INTEGRATION OF CENTRALITY ANALYSIS WITH VISUALIZATION

Centrality, which measures the *importance* of a node in the network, is a useful concept to analyze network structure. Different centrality methods, such as degree, closeness, betweenness and eigenvector, can indicate the importance of the nodes with different aspects [6]. Therefore, it is interesting to visualize different centrality measures and combine them with visualization.

To do so, we firstly adopted a 2.5D hierarchical layout method [1] for drawing directed graphs. In the 2.5D hierarchical layout, each layer is further divided into two parallel *walls* as an

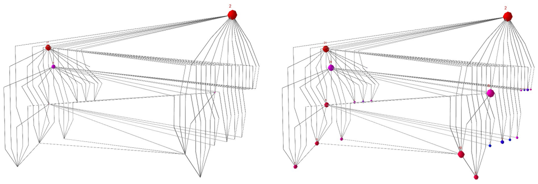


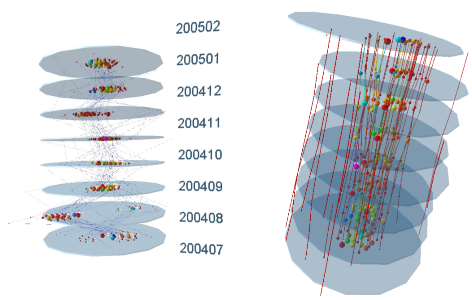
Figure 2: Edge directions display degree centrality. Node sizes display betweenness (left) and eigenvector (right) centrality.

effective way of using the third dimension for reducing the visual complexity and minimizing occlusion. Once having a hierarchy with edge directions related to the degree centrality values we can map other centrality values to the node size.

We have produced two visualizations of the same network with mappings of the eigenvector and betweenness centrality values to the node size respectively. They are shown in Figure 2. The drawings demonstrate how 2.5D hierarchical drawings in combination with visual properties of the nodes can be used for efficient visualization of several centrality values in the same drawing.

4 Visualization of a Temporal Email Network

A 2.5D visualization method can show the entire history of network changes without introducing significant visual complexity. More specifically, nodes are placed into plates; nodes in same plate are connected by edges representing communication; plates of different times are stacked in order. A force-directed layout is applied in each plate to draw each temporal network. Further, degree centrality and betweenness centrality measures are also applied in order to provide analysis. Finally, edges are added between the same nodes in different time plates so that the evolution can be highlighted.



(a) Before adjustment. (b) After adjustment.

Figure 3: Temporal network

Note that the force-directed method implies a degree of randomness. This may result in a drawing as shown in Figure 3(a). Here, inter-plane edges are drawn as long edges, resulting in occlusion, thus hiding the real evolution. We implemented two methods to solve this problem. The first method is to draw each plate separately, but initializing the location of the same vertex in the next plate with the location in the previous plate. The other solution is to define a *supergraph* that consists of the whole temporal network, plus inter-plate edges. Then we run the force-directed algorithm for the supergraph. Thus, inter-plate edges are considered as part of the graph and are assigned corresponding edge weights. Both methods can provide drawings similar to Figure3(b), which has less occlusion.

5 Ambient Display of an Email Network

The aim of ambient display is to provide useful information that blends in aesthetically with the surroundings. Email communications, as a method of human collaboration, have become an integral part of our lives.

Figure 4: Ambient display for collaboration

To meet aesthetic requirements, we use a watercolor image as our output picture. We model the email network as weighted graphs. For the layout, we used a modification of the spring algorithm, such that the distance between the stars depends on the weight of the edges in email network: if two people exchange emails frequently, the stars correspond to the people will be drawn closely. For display, we made a special monitor whose border was covered by an old-fashioned picture frame and it is presented as a traditional wall-hung painting.

6 Current Work

There are a few research challenges require further investigation. The first one is the evaluation and comparison of these visualization methods. The second challenge is interaction. The last challenge is to extend these methods for visual analysis of larger and more complex temporal email networks, for example, the Enron email data set.

References

[1] S. Hong and N. S. Nikolov. Layered drawings of directed graphs in three dimensions. In S. Hong, editor, *Information Visualisation 2005, Asia-Pacific Symposium on Information Visualisation*, volume 45, pages 69–74. CRPIT, 2005.

[2] Teuvo Kohonen. *Self-Organizing Maps*. Springer-Verlag, Berlin Heidelberg, 3rd. edition, 2003.

[3] David W. McDonald. Recommending collaboration with social networks: a comparative evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 593–600, New York, NY, USA, 2003. ACM Press.

[4] Van Wijk J. J. Van Ham F. Interactive visualization of small world graphs. In *IEEE Symposium on Information Visualization*, pages 199–206, Washington, DC, USA, 2004, 2004. IEEE Computer Society.

[5] Gina Danielle Venolia and Carman Neustaedter. Understanding sequence and reply relationships within email conversations: a mixed-model visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 361–368, New York, NY, USA, 2003. ACM Press.

[6] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applicaitons*. Cambridge University Press, 40 West 20th Street, New York, NY 10011-4211, USA, 1st. edition, 1995.

[7] Leizhong Zhang, Nan Tu, and Dave Vronay. Info-lotus: a peripheral visualization for email notification. In *CHI '05 extended abstracts on Human factors in computing systems*, pages 1901–1904, New York, NY, USA, 2005. ACM Press.