

Visualization of Information Flows in a Very Large Social Network

Shin-gyu Kim, Hyuck Han, Kyungho Jeon, Hyungsoo Jung, Heon Y. Yeom
School of Computer Science and Engineering
Seoul National University
Seoul 151-742, Korea

{sgkim, hhyuck, khjeon, jhs, yeom}@dcslab.snu.ac.kr

ABSTRACT

In this paper, we present our research result that enables users to navigate a very large social network and to take a look at information flows on the network. To this end, we devise two techniques; (i) mapping a very large social network to a 2-dimensional graph layout, and (ii) exploring the graph to all directions with zooming it in/out. With these methods, we can also look through information flows over social networks.

Categories and Subject Descriptors

H.5.2 [Information Interfaces]: [User Interfaces]; I.3.6 [Methodology and Techniques]: [Interaction Techniques]

Keywords

Social Network, Information Visualization

1. INTRODUCTION

Recently, many research efforts [2, 4] have suggested diverse methods to visualize graphs, especially in the area of studying social networks. Most relevant research results focus on displaying better patterns of graphs that have been discovered recently. However, we think that displaying *information flows* over weblike human relationships is more important and valuable than displaying patterns and motifs of graphs in the context of online social network services such as Facebook.

Laying out social networks is a key component to displaying information flows on a computer screen. Especially, in case of a large network (e.g., one million nodes), it is unfeasible to display it on a computer screen, even if we depict one node as one pixel. Hence, we suggest three design principles of social network visualization. Visualization systems should (i) display social networks while preserving the small world property [3], (ii) provide access to social networks in an user-friendly way, and (iii) provide quick responses to users' requests. In this paper, we present a new method that

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.

SAC'09 March 8-12, 2009, Honolulu, Hawaii, U.S.A.

Copyright 2009 ACM 978-1-60558-166-8/09/03 ...\$5.00.

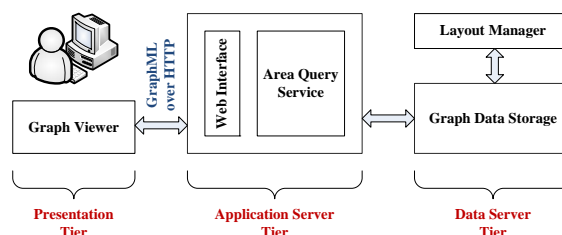


Figure 1: Overall architecture of our system

visualizes social networks and information flows or propagation over social networks in compliance with these design principles.

2. ARCHITECTURE

Figure 1 shows the three-tier architecture of our system. The data server tier consists of two components: a graph DB and a layout manager. The graph DB stores an entire social network graph; in the meantime the layout manager groups given nodes into disjoint sets of clusters and updates the pagerank of each node periodically. The area query service at the application server tier provides the presentation tier with the subgraph of a given social network graph according to a specified area and a zoom level. Users can navigate and monitor information flows on social networks through the graph viewer in the presentation tier.

Figure 2 shows the overall process in the layout manager. In phase I, the layout manager prepares input graph data for the next procedures. In phase II, the layout manager creates a set of clusters from the graph and computes the pagerank of all nodes. The clustering and the ranking modules can run in parallel, and this improves overall efficiency. In phase III, the outputs of two modules are fed into the mapping module that relocates nodes in a structured way.

Clustering has an important role in the visualization of information flows. Researchers observed that information flows between nodes were vividly spreading within the boundary of their own cluster [5]. Without clustering, we can hardly analyze this phenomenon. We use Metis for our clustering module.

Ranking is also important in our system. It enables us to select top-k nodes to draw in a specified area. The dimension of the area is determined by a screen size and a zoom level. Using the rank of a node, we can select important nodes from the entire graph when drawing a subgraph satisfying

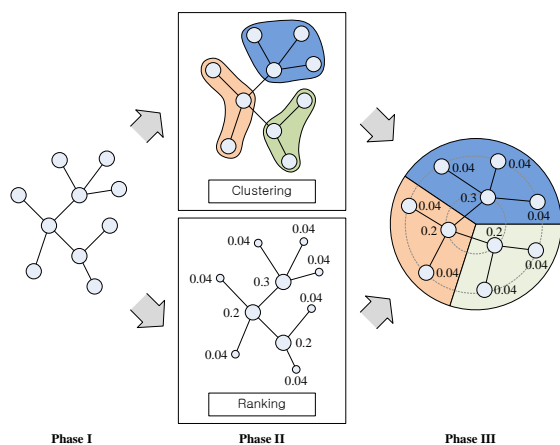


Figure 2: Layout managing process

system’s constraints. Our layout manager adopts the well-known PageRank algorithm for ranking.

In the last step, the layout manager creates a virtual continent in a structured way. From a distance, the continent looks like a big pie (the phase III in Figure 2). Each cluster occupies a certain sector of the pie, which is proportional to the number of nodes the cluster has, and the nodes are positioned by their pagerank values. Higher-ranked nodes (i.e., hub nodes) are located near the center of the pie, and lower-ranked nodes are placed on the outskirts. This layout strategy considers ranking and clustering information effectively. This strategy indeed has two advantages. Firstly, we do not need to explore all areas to visit each cluster, when we want to observe the status of each cluster. Secondly, by looking at the center region where most hub nodes are placed, it is easy to keep observing the behaviors of the hub nodes. These benefits come from the fact that the center of the pie is a start point of each cluster and a rendezvous point of hub nodes.

3. VISUALIZATION

The goal of our system is to build a visualization system through which end-users of social networking services learn information about their community and watch the information flow in it. We also aim at displaying large-scale social networks. Because it was not sufficient for saving system resources to use a static layout method, we omit links between nodes to provide an interactive environment and maintain good readability in this condition. Moreover, we found that dense links hinder users from understanding the overall aspect. For better understanding, we draw only essential links (e.g., links between first-order nodes). In drawing nodes, we also omit imagery when the node is not considered to have any interest. It greatly helps in highlighting nodes and links. For evaluation, we generated a synthetic social network which has 1M nodes and 26M links. Additionally, we also used real data [1], which are from Flickr and Orkut.

Our system represents the information flow with different colors that are chosen by the distance from the source. Figure 3 shows the flow of personal advertisement in social networks. First-order neighbors are colored with yellow, second order with red, and third order with blue. Distribution ratio is shown in the right column. The ratio is also visually rec-

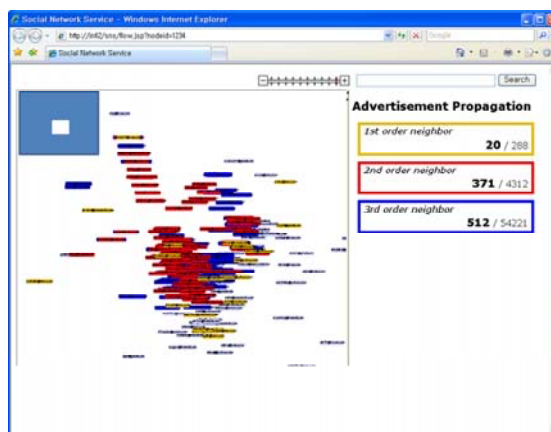


Figure 3: Screenshot of the user-interface showing personal advertisement propagation status

ognized in the right map; unreached nodes are distinguished vividly by its stale color.

We omit links in this visualization. At first, we tried to represent the information flow by the color of links. If the flow is considered as a route, this option is better than coloring nodes. However, the option to color links is not helpful for representing the distribution of reached nodes. Next, we also tried to color both nodes and links. In this case, readability is lessened. This result partly came from the large size of social networks.

4. CONCLUSION

We described a new method to visualize a large-scale social network and the information flows within it. To this end, we introduced the 2-dimensional layout technique and the efficient query processing technique.

5. ACKNOWLEDGEMENTS

This work was supported by the development program of Service Innovation and New Industry Model, the Ministry of Knowledge Economy, Korea. The ICT at Seoul National University provides research facilities for this study.

6. REFERENCES

- [1] IMC 2007 Data Sets, <http://socialnetworks.mpi-sws.org/data-imc2007.html>.
- [2] D. Archambault, T. Munzner, and D. Auber. Grouseflocks: Steerable exploration of graph hierarchy space. *IEEE Transactions on Visualization and Computer Graphics*, 14(4):900–913, 2008.
- [3] Travers, Jeffrey, and S. Milgram. An experimental study of the small world problem. *Sociometry*, 32(4):425–443, 1969.
- [4] F. van Ham and N. Krishnan. Ask-graphview: A large scale graph visualization system. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):669–676, 2006. Member-James Abello.
- [5] F. Wu, B. A. Huberman, L. A. Adamic, and J. R. Tyler. Information flow in social groups. *Physica A Statistical Mechanics and its Applications*, 337:327–335, June 2004.