ALVIN: A System for Visualizing Large Networks

Davood Rafiei
Department of Computing Science
University of Alberta
drafiei@cs.ualberta.ca

Stephen Curial
Department of Computing Science
University of Alberta
curial@cs.ualberta.ca

Categories and Subject Descriptors: I.3.6 [Computer Graphics]: Methodology and Techniques–Interaction techniques; H3.3 [Information Storage and Retrieval]: Information Search and Retrieval

General Terms: Algorithms.

Keywords: visualizing the Web, network visualization, sampling.

1. INTRODUCTION

Visual display of networks can lead to both better understanding and clear presentation of patterns that can often be hidden [3]. However, effectively visualizing large networks has proven to be difficult, due to the limitations of the screen, the complexity of layout algorithms and the limitations of human visual perception. A good layout algorithm (eg. Spring layout) can easily take quadratic time assuming that the network fits in memory. The graph structure of the Web, for instance, is far too large to hold in the memory of most desktops let alone visualize it.

To gain insight into the complexity of the problem, consider the graph structure of the Web at the domain level as shown in Figure 1-a. This network is relatively small, having only 224 nodes, but it is still not easy to find any interesting patterns. Colouring the largest *Strongly Connected Component* [1](SCC), as is shown in Figure 1-b, singles out some of the domains that are not in the SCC. One such domain, for instance, is *Vatican City* (va), linked by a large number of domains in the SCC but is not linking back to any domain in the SCC. Even in this graph, it is not easy to see the connectivity structure of many of the domains in the SCC. Scaling up the visualization to a graph of the Web with millions of nodes at the site level or hundreds of millions of nodes at the page level is quite challenging if not impossible.

Our proposed alternative in ALVIN [2] is to refrain from visualizing the entire network. At the core of our methods is sampling. We sample the network and only visualize the sample. Even though the network can be quite large, the size of the sample can be adjusted to match the limitations of the visualization environment.

[1]A *strongly connected component* of a graph is a set of nodes such that for any pair of nodes u and v in the set, there is a path from u to v.

[2]The name *ALVIN* stands for *Alberta system for Visualizing Large Networks*.

Copyright is held by the author/owner.
WWW 2005, May 10–14, 2005, Chiba, Japan.
ACM 1-59593-051-5/05/0005.

A typical browsing session may start with a visualization of the result of a user-supplied query or its subset. More details of the interconnections may be added by expanding this set to include some of the immediately adjacent nodes and edges. Varying the degree of focus from immediately adjacent nodes to the rest of the network, one can see a more global picture of the network, including nodes and edges from the entire network and probably some of the major connections of these nodes from the entire network with nodes in the starting set.

There has been past work on layout and encoding schemes that can scale-up to large trees or more specific graphs (e.g. [4]). Our work is different from these in that we don't make any assumption on the structure of the network or the characteristics of the nodes. There has been also work on general multiscale abstraction methods that allow one to visualize either the global structure or the smaller components of a large network (e.g. [1]). Our work is orthogonal to these abstraction methods. Related to our work is also the more general work on analyzing social networks (e.g. [7]), mining graphs [5] and analyzing the graph structure of the Web [2].

2. SYSTEM OVERVIEW

ALVIN makes use of some sampling-based schemes for both focusing the search and visualizing networks which are too big to be fully visualized.

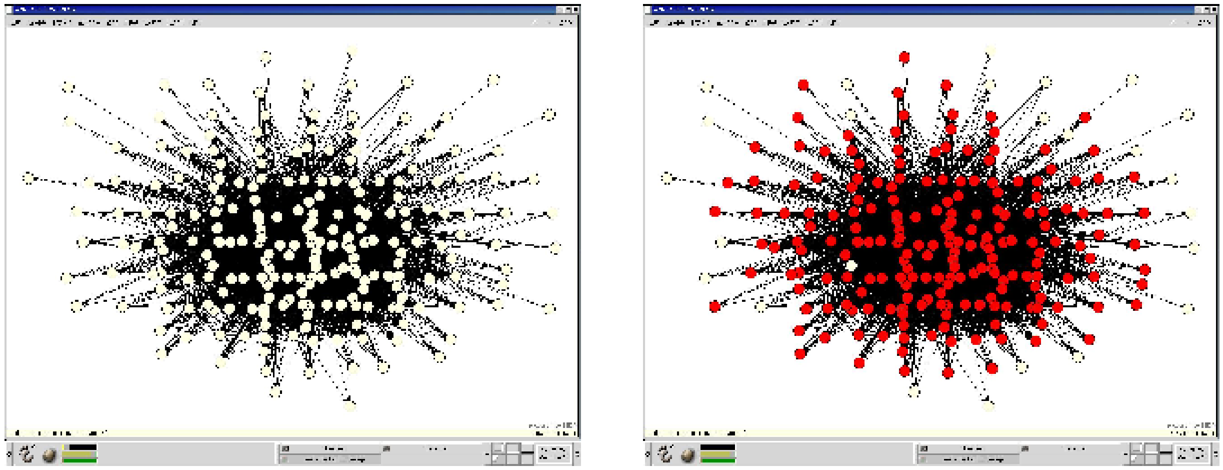
2.1 Sampling a Network

There are several ways of sampling a network, and a few of the characteristics of the original network can be preserved in the sample. Given a network, any subgraph of the network can be treated as a sample of the network. Clearly, there are different ways of taking a subgraph and as a result there are many different sampling strategies. We use three methods for obtaining a simple random sample of a network: (1) sampling from nodes, (2) sampling from edges and (3) sampling from edges followed by adding the edges with both endpoints in the sample.

There is a caveat when sampling from nodes; unless the network is very "well-connected," the resulting sample would be quite sparse. In this case, sampling from edges instead may be more desirable.

2.2 A Model of Browsing for Large Networks

The original network can be large, and visualizing a small sample that can preserve some of the desired topological properties of the network may not be feasible. To address this problem and to provide a navigation scheme, we develop



(a) (b)

Figure 1: *(a) The connectivity network of the World Wide Web at the domain level, and (b) the same network with the SCC coloured red.*

several growth processes, collectively referred to as *network growth*, that allows one to interactively visualize a network.

- *Global growth*: Visualize the network without emphasizing one specific part of it.
- *Local growth*: Narrow down the visualization to a part of the network where the user is interested in.
- Mixed growth: Narrow down the visualization to a part of the network where the user is interested in but visualize this part in the context of the larger network.
- Wiring: Show more interconnections of the network that is being visualized.

In an interactive fashion to some degree similar to Web browsers, our visualization starts with a small subset of the network which may include a set of hand-picked nodes and edges or the result of a query. The visualization may proceed towards the goal by iteratively growing the initial set. This is useful for narrowing down the visualization to some of the interesting elements when the network is too large to be fully visualized. A novelty of our method is the way the network, currently displayed on canvas, is expanded. Our method uses user-controllable parameters to describe how and to what degree the network must be expanded. The expanded network often has more detail about the elements being studied yet is small enough to be visualized and internalized. After a few layers of extension, the network may become too large; this may be an indication that the browsing should switch to another small subset before it can continue.

3. EVALUATION AND CONCLUSIONS

ALVIN has been implemented and makes use of the DB2 relational database as its back-end data storage and querying engine and the LEDA class library for drawing and layout tools. To evaluate our prototype, we ran experiments on a few different networks, including the movie database, the linkage structure of a snapshot of the Web, and other synthetic networks. Due to the lack of space, we cannot report our experiments here. The result shows that indeed sampling can preserve some of the topological properties of a network and this can be useful in visualizing very large networks.

Given the limitations of the screen and the size of a sample, our proposed scheme allows the search to be localized, thus increasing the ratio of sample size to the size of the desired network and removing possible biases due to the sample size.

More details about our analysis and evaluations can be found elsewhere [6].

Acknowledgment

This research was supported by Natural Sciences and Engineering Research Council and the Alberta Ingenuity Centre for Machine Learning.

4. REFERENCES

[1] D. Auber, Y. Chiricota, F. Jourdan, and G. Melancon. Multiscale visualization of small world networks. In *Proc. of IEEE Symposium on Information Visualization*, pages 75–81, 2003.

[2] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the Web. In *Proc. of the World Wide Web Conference*, pages 309–320, May 2000.

[3] A. Klovdahl. A note on images of networks. *Social Networks*, 3:197–214, 1981.

[4] T. Munzner. *Interactive visualization of large graphs and networks.* PhD thesis, Stanford University, 2000.

[5] C. Palmer, P. Gibbons, and C. Faloutsos. Data mining on large graphs. In *Proc. ACM Intl. Conf. on SIGKDD*, pages 81–90, 2002.

[6] D. Rafiei and S. Curial. Effectively visualizing large networks through sampling. Technical Report TR05-08, University of Alberta, 2005.

[7] D. Watts. *Small worlds: the dynamics of networks between order and randomness.* Princeton University Press, 1999.