

GraphExplore: a software tool for network visualization

Quanli Wang, Guang Yao, Joseph Nevins, Mike West and Adrian Dobra

Duke Center for Applied Genomics and Technology and Institute of Statistics and Decision Sciences,
114 Old Chemistry Building, Durham, NC 27708, USA

ABSTRACT

Summary: GraphExplore is a program for presenting, visualizing and interrogating large complex networks of interactions. Although the spectrum of applicability of our program is very broad, GraphExplore's numerous features are ideal for exploring interactions in literature, gene expression, protein interaction and promoter interaction profiling studies since it dynamically couples with biological annotation systems and databases to provide interactive facilities to interpret genes and proteins displayed.

Availability: <http://graphexplore.cgt.duke.edu>

Contact: quanli@stat.duke.edu

PROGRAM OVERVIEW

We have developed a comprehensive and easy-to-use point-and-click graphical environment called GraphExplore for dynamically exploring large graphical structures that possibly involve tens or hundreds of thousands of objects. It is a stand-alone application written in JAVA that runs on any operating system that has a JAVA virtual machine installed.

The main objective of the program is to generate displays of associations among a set of objects. The objects are vertices in a graph, while the associations among these objects are represented as arrows and lines. Three files can be loaded as part of a GraphExplore project: a description file with information about the objects in the network, a network file with information about the associations among the objects and a cluster file that identifies groups of objects. Symmetric associations are represented with lines while directed associations are represented with arrows.

GraphExplore can dynamically retrieve information about the objects in the network from the Internet. One can either specify a URL for each object or let GraphExplore query a relevant website with keywords identifying characteristics of interest. The resulting webpages can be displayed in a web browser or in a new graphical panel inside GraphExplore. The program creates the image of a network by employing several rendering engines such as the *neato* utility from AT&T's Graphviz open source graph drawing software (AT&T Labs—Research, 2000) or in-house modified versions of the layout libraries available from JUNG (White *et al.*, 2004).

Once a project is loaded, one can start building subgraphs of interest by interrogating the network. Each object is identified by one or more names that could be either numerical

identifiers or strings of characters. The objects whose names are consistent with the current user query are called *targets*. GraphExplore can connect the targets by identifying linker nodes located on shortest paths of various lengths between the targets (see Zhou *et al.* (2002)) or are k -steps away from the targets ($k=1, 2$ or 3) as in Figure 1. GraphExplore is also able to reveal the cluster a certain object belongs to.

GraphExplore can display only some of the edges in the network whose weights are between a minimum and a maximum threshold. This essentially creates another network with fewer connections that is further interrogated.

The objects in the current subgraph can be moved or deleted. GraphExplore dynamically displays information about the neighbors of the objects in the network as well as about their corresponding links. It also generates reports summarizing the information about the objects and the associations among these objects in a particular graphical display. The users of the system can choose to save their work in several graphical formats that can easily be embedded in other documents. Several queries can be active in the same time in different graphical panels.

A very important and attractive feature of GraphExplore is that it can display specific colors and shapes for each object in the network. The same shapes and/or colors can be assigned to various groups of objects. The shape of the objects can identify one clustering, the colors can identify another clustering and both these clusterings can be different than the clustering loaded with the project. This degree of flexibility is necessary to create meaningful displays of objects with various functions.

GraphExplore has many built-in utility programs such as a tool for generating the intersection or union of two networks whose images were previously saved in SVG format. The common nodes and edges are easily distinguished in the resulting image from the other elements by their shapes and/or colors.

Downloading the program requires only a couple of mouse clicks since it uses Java Web Start (<http://java.sun.com/products/javawebstart/>). This ensures that the users are always working with the latest released version of GraphExplore even after the initial installation. A comprehensive help file, a short tutorial as well as a number of examples are available from GraphExplore's website.

ACKNOWLEDGMENTS

The authors thank Joseph Nevins and Guang Yao for very helpful discussions. Research was partially supported by the NIH on Grant HL73042-01.

References

- AT&T Labs–Research (2000) Graphviz–open source graph drawing software. Available for download at <http://www.research.att.com/sw/tools/graphviz/>.
- Dobra, A., Jones, B., Hans, C., Nevins, J. and West, M. (2004) Sparse graphical models for exploring gene expression data. *Journal of Multivariate Analysis*, **90**, 196–212.
- Dobra, A. and West, M. (2004) Bayesian covariance selection. ISDS Discussion Paper #04-23.
- Huang, E., Cheng, S. H., Dressman, H., Pittman, J., Tsou, M. H., Horng, C. F., Bild, A., Iversen, E. S., Lio, M., Chen, C. M., West, M., Nevins, J. R. and Huang, A. T. (2003) Gene expression predictors of breast cancer outcomes. *The Lancet*, **361**, 1590–1596.
- Lacroix, M. and Leclercq, G. (2004) About GATA3, HNF3A, and XBP1, three genes co-expressed with the oestrogen receptor- α gene (ESR1) in breast cancer. *Molecular and Cellular Endocrinology*, **219**, 1–7.
- White, S., O’Madadhain, J., Fisher, D. and Boey, Y. B. (2004) JUNG–Java Universal Network/graph Framework. Available for download at <http://jung.sourceforge.net/>.
- Zhou, X., Kao, M. C. J. and Wong, W. H. (2002) Transitive functional annotation by shortest-path analysis of gene expression data. *Proc. Natl. Acad. Sci.*, **99**, 12783–12788.

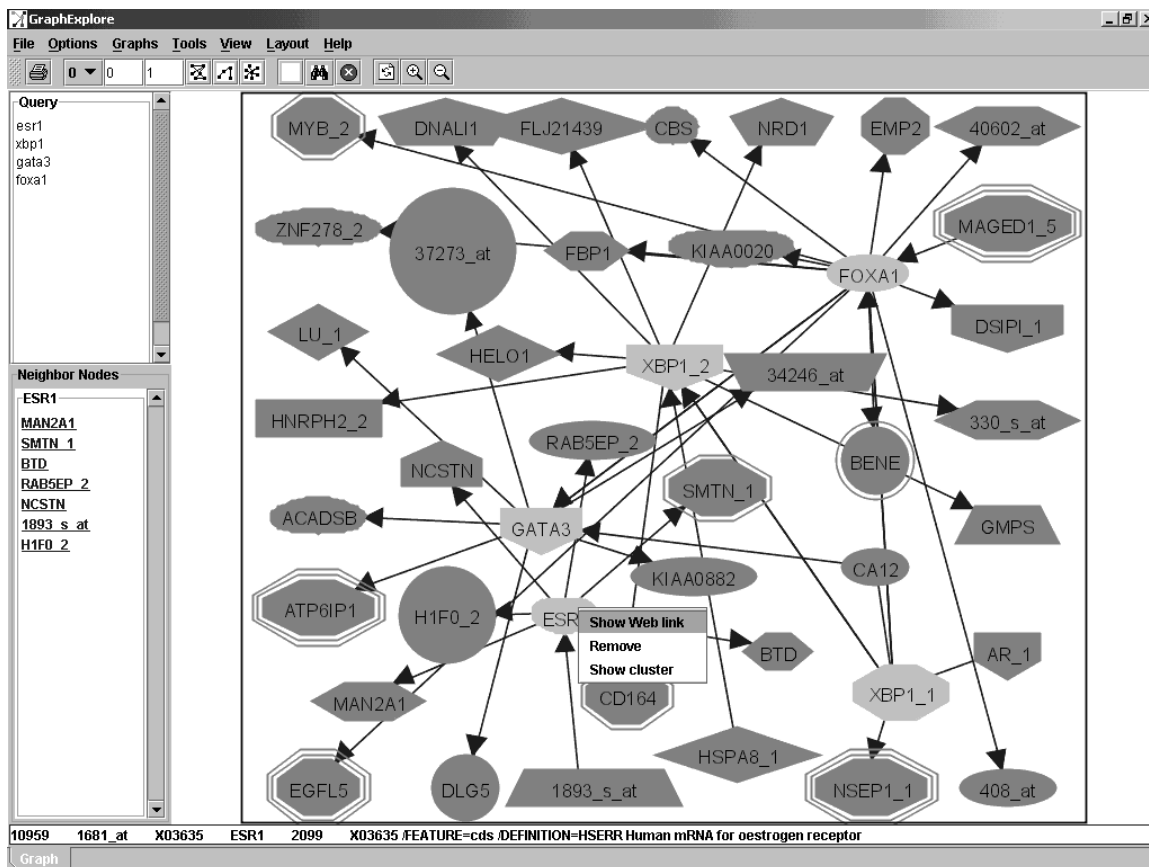


Figure 1: User interface of GraphExplore. The objects in this network are genes while the associations displayed are derived from sparse graphical Gaussian models (Dobra and West, 2004; Dobra *et al.*, 2004) constructed from a publicly available breast cancer gene expression dataset analyzed in Huang *et al.* (2003). The targets are the oestrogen receptor- α gene (ESR1) together with three other genes (GATA3, FOXA1/HNF3A, XBP1) that are known to have expression levels strongly correlated with ESR1 (Lacroix and Leclercq, 2004). The linker genes are displayed in a darker color and represent the one-step neighbors of the target genes. Here the shapes of the nodes were chosen at random to illustrate GraphExplore's features. The upper-left panel gives the current user query while the lower left panel shows the direct neighbors of ESR1. Description of objects, edge weights and system messages are displayed in the status bar located at the bottom of the window.