Computing topological parameters of biological networks

Yassen Assenov, Fidel Ramirez, Sven-Eric Schelhorn, Thomas Lengauer and Mario Albrecht*

Department of Computational Biology and Applied Algorithmics, Max Planck Institute for Informatics, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany

Received on June 16, 2007; revised on October 8, 2007; accepted on November 1, 2007
Advance Access publication November 15, 2007
Associate Editor: Martin Bishop

ABSTRACT
Summary: Rapidly increasing amounts of molecular interaction data are being produced by various experimental techniques and computational prediction methods. In order to gain insight into the organization and structure of the resultant large complex networks formed by the interacting molecules, we have developed the versatile Cytoscape plugin NetworkAnalyzer. It computes and displays a comprehensive set of topological parameters, which includes the number of nodes, edges, and connected components, the network diameter, radius, density, centralization, heterogeneity, and clustering coefficient, the characteristic path length, and the distributions of node degrees, neighborhood connectivities, average clustering coefficients, and shortest path lengths. NetworkAnalyzer can be applied to both directed and undirected networks and also contains extra functionality to construct the intersection or union of two networks. It is an interactive and highly customizable application that requires no expert knowledge in graph theory from the user.

Availability: NetworkAnalyzer can be downloaded via the Cytoscape web site: http://www.cytoscape.org
Contact: mario.albrecht@mpi-inf.mpg.de

Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION
In recent years, high-throughput experiments have produced large networks of interacting molecules, which are represented as nodes linked by edges in complex graphs (Albrecht et al., 2005; Ramirez et al., 2007; Zhu et al., 2007). In this context, the characterization of biological networks by means of graph-topological properties has become very popular for gaining insight into the global network structure (Albert, 2005; Almaas, 2007; Barabasi and Oltvai, 2004; Dong and Horvath, 2007; Zhu et al., 2007). However, general software libraries for graph analysis such as JUNG (http://jung.sourceforge.net/), LEDA (http://algorithmic-solutions.com/leda.htm), NetworkX (https://networkx.lanl.gov/) and yFiles (http://www.yworks.com/) are not easily applied by the biological user. Other applications like Pajek (Batagelj and Mrvar, 1998) require expert knowledge in graph theory on the user side. Specialized tools for the analysis of biological networks like CentiBiN (Junker et al., 2006), tYNA/TopNet (Yip et al., 2006; Yu et al., 2004) and VisANT (Hu et al., 2005) calculate only a limited set of topological parameters.

Therefore, we have developed NetworkAnalyzer, a user-friendly Java plugin for Cytoscape (Shannon et al., 2003), which is an established free open-source software platform for the visualization and analysis of molecular interaction networks (Shannon et al., 2003). An initial release of NetworkAnalyzer was made available in January 2006. In the following, we describe the basic functionality of NetworkAnalyzer and numerous extensions and improvements of the next major release.

2 PROGRAM OVERVIEW
NetworkAnalyzer efficiently computes a large number of topological network parameters for directed and undirected networks loaded into Cytoscape. The user can decide whether directed edges should be treated as undirected for the analysis. The computed simple and complex topology parameters are represented as single values and distributions, respectively. Simple parameters are the number of nodes, edges, self-loops, and connected components, the average number of neighbors, the network diameter, radius, density, centralization, heterogeneity, and clustering coefficient, the number of shortest paths, and the characteristic path length. Complex parameters are distributions of node degrees, neighborhood connectivities, average clustering coefficients, topological coefficients, shortest path lengths, and shared neighbors of two nodes. NetworkAnalyzer utilizes the free Java libraries JFreeChart (http://jfree.org/jfreechart/) and Batik (http://xmlgraphics.apache.org/batik/) to display the distributions as histograms or scatter plots (Fig. 1) and to export them as chart images in the formats JPEG/PNG/SVG or as tables in plain text files. Details on the formal definitions of all topological parameters are given in the online help page of the plugin. To ensure the validity of the calculations performed by NetworkAnalyzer, the computed parameters were compared with those obtained from Pajek, TopNet, and using the Python graph library NetworkX.

While the majority of the topological parameters included in NetworkAnalyzer is already well known and frequently used in the literature, our plugin additionally computes some novel network properties. In particular, we have extended the original definition of neighborhood connectivity (Maslov and Sneppen, 2004).
to directed networks by introducing three types of related connectivity parameters, see Supplementary Data for more details. NetworkAnalyzer is also capable of enumerating the shared neighbors of all node pairs in a network. As an application of that, the Supplementary Data describes the use of the shared neighbors distribution to detect bias in the topology of predicted human networks of protein–protein interactions in comparison to experimentally derived networks.

Further unique features of NetworkAnalyzer comprise various visual settings of the obtained diagrams (Fig. 1). The user has the option of switching between histogram or scatter plot of the computed distributions and between linear or logarithmic scales for any of the two displayed diagram axes. Gridlines can be enabled or disabled, and a power law can be fitted to resultant distributions. Additionally, the title of the chart diagram, the labels of the axes, and the colors of the scatter points and gridlines can be configured.

Topology parameters computed for network nodes are stored as node attributes in the Cytoscape data structure. Thus, users can easily apply the visual mapping settings of Cytoscape to highlight any parameter on the screen (see online tutorial). For example, the clustering coefficient may be visualized proportional to the node size, and the node color may be related to its degree. Another useful application of NetworkAnalyzer is the selection of nodes based on any of the calculated attributes. This enables Cytoscape users to examine, for instance, structural perturbations in a network caused by the removal of nodes with high degrees.

In NetworkAnalyzer, the complete set of simple and complex parameters is referred to as network statistics. Once calculated and displayed, the network statistics can be saved into and reloaded from a text file in order to avoid recomputation. The comparison of multiple network topologies can easily be achieved by the parallel inspection of the computed statistics for different networks. Optional user settings can be stored and reloaded. Users can customize the appearance of the results by choosing between two alternative dialog interfaces, the compact one shown in Fig. 1 and an expandable interface.

Aside from parameter computations, NetworkAnalyzer offers a useful set of network modifications and supports the construction of the intersection, union, and difference of two networks, the extraction of connected components as new separate networks, and the removal of self-loops.

3 CONCLUSIONS

NetworkAnalyzer is a versatile and user-friendly tool for the analysis of biological and other networks. This plugin is well integrated into Cytoscape and computes a comprehensive list of simple and complex topology parameters using efficient graph algorithms. It incorporates useful visualization settings to display and export the resulting distributions and adds node attributes for the results.

ACKNOWLEDGEMENTS

Part of this work has been financially supported by the German National Genome Research Network (NGFN) and the German Research Foundation (DFG), contract number KFO 129/1-1. The research has been conducted in the context of the BioSapiens Network of Excellence funded by the European Commission under grant number LSHG-CT-2003-503265.

Conflict of Interest: none declared.

REFERENCES

*BMC Syst. Biol.*, 1, 24.


