Systems Biology

ToME: Topology Measurement Environment

Ajish D. George[1,2,*] and Thomas J. Begley[1,2]

[1]Department of Biomedical Sciences, University at Albany, State University of New York, Rensselaer, NY 12144

[2]Gen*NY*Sis Center for Excellence in Cancer Genomics, University at Albany, State University of New York,
Rensselaer, NY 12144-3456, USA

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Motivation: Summary: The Topology Measurement Environment (ToME) is a tool for filtering and exploring sub-networks within a larger interactome and measuring their small-world graph properties. It provides novel control generation methods to ensure both significance of selection and significance of connectivity for properties associated with the sub-network of interest.

Availability and Implementation: The ToME software and source code are available as a Java archive (JAR) file at http://ribonomics.albany.edu/tome/.

Contact: Ajish D. George (ajishg@gmail.com)

1 BACKGROUND

The Topology Measurement Environment (ToME) is a tool for subsetting protein-interaction networks and looking at their topological properties. These include clustering coefficients, average path lengths, and average degrees of individual nodes and entire subgraphs, and allow the study of small-world network properties (Watts and Strogatz 1998) as apply to the protein-network of interest. Interactome mapping, in general, has been shown to identify biologically important architectures and clustering coefficient analysis, specifically, can be used to identify signatures of protein pathways and complexes responding to damage (Rual et al. 2005) and can identify local areas of high connectivity in a mapped network (Jeong et al. 2000, Gunsalus et al. 2005).

Measures of graph properties include simple measures of the number of nodes and the number of edges in a graph and more complex measures of graph connectivity including clustering coefficient, average degree, and average path length (Alberts and Barabasi 2002). The clustering coefficient of a graph measures how likely any node in a graph is to be part of a clique or a complete graph where every node is directly connected to every other node (Watts and Strogatz 1998, Alberts and Barabasi 2002). The average degree of a graph is simply the average of the number of edges attached to each node while the average path length is the average of the lengths of the shortest paths between each pair of nodes (Watts and Strogatz 1998, Albert and Barabasi 2002).

Graphs with a higher than average clustering coefficient and a low average path length between nodes are called small-world networks (Watts and Strogatz 1998) and exhibit a few interesting characteristics. Small-world networks are a type of scale-free networks in which the degree distribution of the nodes (number of edges emanating from each node) follows a power law distribution i.e. there is a small number of nodes that are highly connected (Albert and Barabassi 2002). Scale-free networks are robust to damage in the sense that damage at a random node will likely not interfere with the general connectivity of the network or increase the average path length significantly (Albert and Barabassi 2002). It has been hypothesized that this property of small-world networks has made them evolutionarily survivable and is responsible for their prevalence in observed biological interactomes (Jeong et al. 2000).

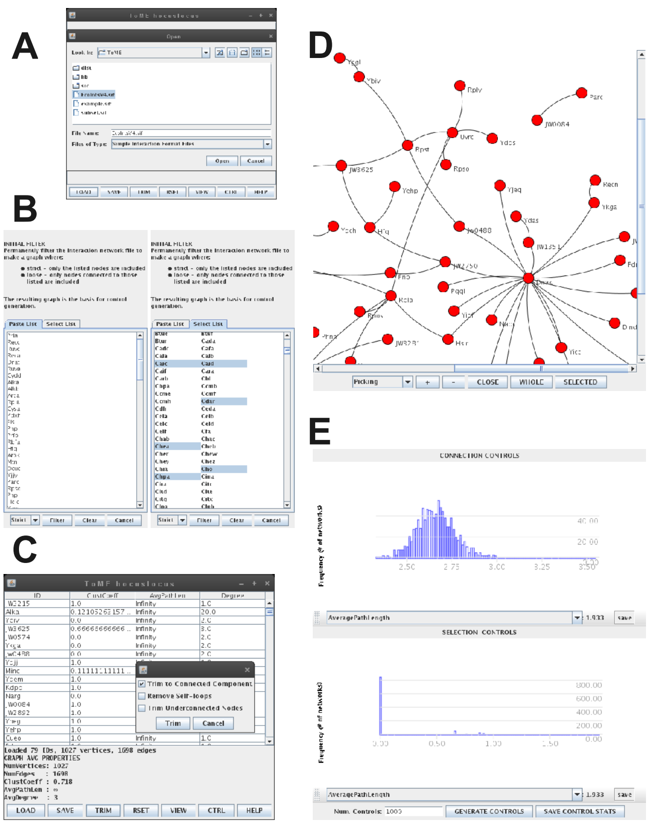
While many tools exist for measuring the graph theoretic properties of a given network, few tools exist for estimating the significance of any measured property. In practice, significance of graph theoretic properties are generally estimated by scrambling the edges of the network of interest and measuring the properties of each of these scrambled graphs (Said et al. 2004, Li et al. 2006, Rajarathinam et al. 2006) While this sort of control for connectivity (which we call a connection control) is a useful metric, it does not provide a complete measure of the space of graph topologies in a biological interactome.

When measuring the properties of a subnetwork of a biological interactome, usually the result of a some theoretical or experimental screening process, it is important to estimate the significance of the selection of the result nodes from the larger network. We do this by creating selection controls – random subgraphs of the interactome with the same number of nodes as the subgraph selected by the screen.

2 DESCRIPTION OF APPROACH

ToME allows system biologists to filter large scale protein-interaction networks (represented as SIF files) (Figure 1A) and filter them using a list of nodes (Figure 1B) of interest. ToME allows two levels of stringency when filtering a network with a list of nodes. Strict filtering will leave only the nodes in the filter list, while loose filtering will also leave in nodes connected to those in the filter list.

The initial filtering usually results in graph with multiple disconnected subcomponents. This is problematic since many of the graph-theoretic properties are only defined for fully connected



☑ Trim to Connected Component
□ Remove Self-loops
□ Trim Underconnected Nodes

```
Loaded 79 IDs, 1027 vertices, 1698 edges
GRAPH AVG PROPERTIES
NumVertices: 1027
NumEdges   : 1698
ClustCoeff : 0.718
AvgPathLen : ∞
AvgDegree  : 3
```

Num. Controls: 1000

Figure 1: Overview of ToME Features

A. Loading an Interactome: An initial interactome from which all sub-graphs are derived must be provided as a SIF file. B. Selecting a Sub-Graph: Paste the list of nodes to be included or select the nodes from a list and choose strict or loose filtering. C. Viewing Node Properties & Trimming: Name, clustering coefficient, average path length, and degree for each node in the sub-graph is shown in a sortable table. Users can trim a graph to include only the largest connected component, to remove edges connecting nodes to themselves, or to remove nodes connected to only one other node. D. Visualizing the Network: Select one or more nodes from the table and click View to display the network around the selected nodes. E. Generating Controls: Enter the number of controls of each type to be generated. Distributions of graph properties for each set of control graphs are shown. Distribution values and images can be exported,

networks. To facilitate this we allow the user to trim the graph to its largest connected component. Also depending on the context of the study, the user may not be interested in self-loops, since they inflate metrics such as the clustering coefficient. A final consideration arises when the user wishes to inspect particularly highly connected clusters (possibly representing complexes). Here nodes which are on the periphery of the main network, connected to only one other node, are less interesting. Connected chains of these types of nodes are recursively removed by the trim underconnected filter (Figure 1C).

For any trimmed graph, users will see a table listing each node along with its graph properties (Figure 1C). The table can be sorted according to any of the columns. If some rows are of particular interest the user can select those and view the graph around them (Figure 1D). The full subgraph is also available for viewing there.

Estimating the statistical significance of any of the properties associated with our graph is a problem that is best appoached empirically. We sample the distributions of the graph theoretic properties of interest in control graphs to estimate the likelihood of a given value for a graph property. We use the two definitions of control graphs discussed previously -- connection controls and selection controls. In both of these cases, all trimming functions run on the query graph are also run on the control graphs. ToME provides facilities to generate any desired number of connection and selection controls, to examine distributions of the graph theoretic properties in these control networks and to export these distributions for further analyses. The control generation interface is pictured in Figure 1E.

The ToME utility and novel control-generation methods described here have been recently used in characterizing protein-interaction networks responding to a variety of DNA damage stimulus in a yeast system. A highly-connected subnetwork of proteins coordinating response to MMS toxicity by modulating mRNA translation was identified (Rooney et al. 2009) using the clustering-coefficient analysis tools described here.

ACKNOWLEDGEMENTS

We would like to thank God.

Funding: NIH cashola.

REFERENCES

Albert, Reka, and Albert Barabasi. "Statistical mechanics of complex networks." Reviews of Modern Physics 74.1 (2002): 97, 47.

Gong, Yunchen, and Zhaolei Zhang. "Global robustness and identifiability of random, scale-free, and small-world networks." Annals of the New York Academy of Sciences 1158 (2009): 82-92.

Gunsalus, Kristin C. et al. "Predictive models of molecular machines involved in Caenorhabditis elegans early embryogenesis." Nature 436.7052 (2005): 861-865.

Jeong, H et al. "The large-scale organization of metabolic networks." Nature 407.6804 (2000): 651-654.

Li, Dong et al. "Protein interaction networks of Saccharomyces cerevisiae, Caenorhabditis elegans and Drosophila melanogaster: large-scale organization and robustness." Proteomics 6.2 (2006): 456-461.

Rajarathinam, Thanigaimani, and Yen Han Lin. "Topological properties of protein-protein and metabolic interaction networks of Drosophila melanogaster." Genomics, Proteomics & Bioinformatics / Beijing Genomics Institute 4.2 (2006): 80-89.

Rooney, John P et al. "Systems based mapping demonstrates that recovery from alkylation damage requires DNA repair, RNA processing, and translation associated networks." Genomics 93.1 (2009): 42-51.

Rual, Jean-Francois et al. "Towards a proteome-scale map of the human protein-protein interaction network." Nature 437.7062 (2005): 1173-1178.

Said, Maya R. et al. "Global network analysis of phenotypic effects: Protein networks and toxicity modulation in Saccharomyces cerevisiae." Proceedings of the National Academy of Sciences of the United States of America 101.52 (2004): 18006-18011.

Watts, D J, and S H Strogatz. "Collective dynamics of 'small-world' networks." Nature 393.6684 (1998): 440-442.