

The tYNA Platform for Comparative Interactomics: A Web Tool for Managing, Comparing and Mining Multiple Networks

Kevin Y. Yip¹, Haiyuan Yu², Philip M. Kim³, Martin Schultz¹ and Mark Gerstein^{3, 1*}

¹Department of Computer Science, Yale University, 51 Prospect Street, New Haven, CT 06511, USA

²Department of Cancer Biology, Dana-Farber Cancer Institute, 1 Jimmy Fund Way, SM854, Boston, MA 02115, USA

³Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06520, USA

ABSTRACT

Biological processes involve complex networks of interactions between molecules. Various large-scale experiments and curation efforts have led to preliminary versions of complete cellular networks for a number of organisms. To grapple with these networks, we developed tYNA, a Web system for managing, comparing and mining multiple networks, both directed and undirected. tYNA efficiently implements methods that have proven useful in network analysis, including identifying defective cliques, finding small network motifs (such as feed-forward loops), calculating global statistics (such as the clustering coefficient and eccentricity), and identifying hubs and bottlenecks. It also allows one to manage a large number of private and public networks using a flexible tagging system, to filter them based on a variety of criteria, and to visualize them through an interactive graphical interface. A number of commonly used biological datasets have been pre-loaded into tYNA, standardized and grouped into different categories.

Availability: The tYNA system can be accessed at <http://networks.gersteinlab.org/tyna>. The source code, JavaDoc API, and WSDL can also be downloaded from the web site. tYNA can also be accessed from the Cytoscape software using a plugin.

Supplementary Information: Additional figures and tables can be found at <http://networks.gersteinlab.org/tyna/supp>.

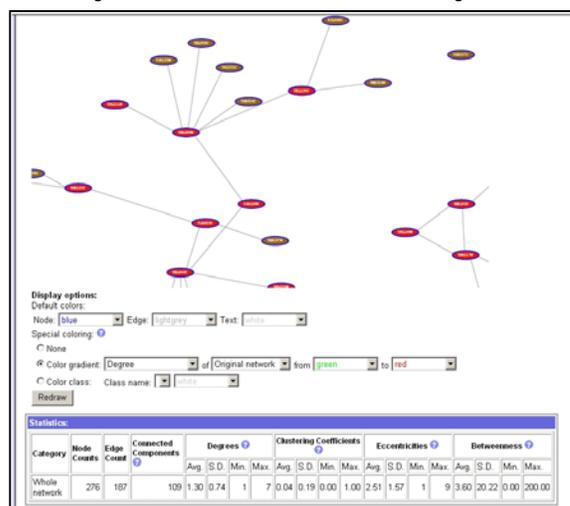
1 INTRODUCTION

In the era of systems biology, the focus on understanding complex organisms is shifting from studying individual genes and proteins towards the relationships between them (Sharan and Ideker, 2006). These relationships are usually expressed in terms of various kinds of biological networks. Recent developments of large-scale experiments such as mass spectrometry and array-based techniques (Lee *et al.*, 2002, Gavin *et al.*, 2006, Krogan *et al.*, 2006) have generated rough descriptions of the complete networks in the cells. Many studies have reported interesting biological findings from these networks, including the relationships between various statistical properties of a gene and its function and essentiality, and the elucidation of controls at the molecular level based on network motifs (Lee *et al.*, 2002, Milo *et al.*, 2002, Shen-Orr *et al.*, 2002, Han *et al.*, 2004, Yu *et al.*, 2004).

These studies require heavy computations on multiple networks. We have developed a Web system, tYNA (TopNet-like Yale Network Analyzer), to provide researchers with a set of tools to carry out such computations with great ease. The system provides five

main types of functionality: (1) Network management: storing, retrieving and categorizing networks. A comprehensive set of widely used network datasets is preloaded, put into standard form, and categorized with a set of tags. (2) Network visualization: displaying networks in an interactive graphical interface (Fig. 1). (3) Network comparison and manipulation: various kinds of filtering and multiple network operations. (4) Network analysis: computing various statistics for the whole network and subsets, and finding motifs and defective cliques. (5) Network Mining: predicting one network based on the information in another.

Fig. 1. The intersection of two yeast-two-hybrid datasets (Ito *et al.*, 2000, Uetz *et al.*, 2000) with all nodes having no edges in the intersection filtered by a statistics filter. The nodes are colored according to their degrees. Also shown in the figure are the various statistics of the resulting network.



Our system shares some elements with some other network analysis and visualization systems, such as Cytoscape (Shannon *et al.*, 2003), JUNG¹, N-Browse², and Osprey³, but also offers some additional features such as defective clique finding. Besides, being a Web-based system, tYNA also has some unique advantages:

- Users can share networks through a centralized database.
- Computationally intensive tasks such as motif finding and statistics calculations can be performed on powerful servers.

¹<http://jung.sourceforge.net/>

²<http://nematoda.bio.nyu.edu:8080/NBrowse/N-Browse.jsp?last=false>

³<http://biodata.mshri.on.ca/osprey/>

*To whom correspondence should be addressed.

- The system can be linked from/to other online resources.
- Users can incorporate some functions of tYNA into their own programs using the SOAP-based web service interface.

Table 1 summarizes some major differences between the systems. A more detailed description is given on the supplementary web site.

Table 1. A brief comparison of several existing systems

	Cytoscape	JUNG	N-Browse	Osprey	tYNA
Main purpose	Visualization	Library	Visual.	Visual.	Analysis
System	Standalone	Standalone	Web	Standalone	Web
Link from external resources	Indirect: Java Web Start	No	Direct	No	Direct
Stat. calculation	Plugin	Built-in	No	No	Built-in
Motif finding	Plugin	No	No	No	Built-in

2 USING tYNA

tYNA provides a simple view with some basic features, and an advanced view for more complex analyses.

2.1 Uploading networks and categories

The first step of analysis is to upload networks. tYNA accepts various file formats, including the SIF format of Cytoscape. One may also enter additional attributes to organize the networks into groups, such as network type (e.g. protein-protein interaction), organism (e.g. yeast) and experimental method (e.g. yeast-two-hybrid) (Fig. S1). Furthermore, tYNA allows users to analyze subsets of the networks (e.g., active parts in a dynamic network (Han et al., 2004, Luscombe et al., 2004)) by using category files.

2.2 Loading networks into workspaces

After uploading a network, one may view its statistics and visualize it graphically by loading it into a workspace. A workspace is a working area for a single network (Fig. 1). Various statistics are computed, such as the clustering coefficient, eccentricity and betweenness (Yu et al., 2004). Networks are visualized in Scalable Vector Graphics (SVG) using the aiSee package⁴, which facilitates an interactive interface: one may change the appearance of the network in real time (Fig. S2, S3).

2.3 Single-network operations (advanced view)

Filtering allows one to retain a portion of the network, based on a statistics cutoff (e.g. the 5% of nodes with the highest out-degrees) or node names. It will easily allow one to identify the hubs and bottlenecks in a graph. Motif finding identifies various regular patterns in the network, including chains, cycles, feed-forward loops and complete two-layers. They generalize the motifs discussed in previous studies (Lee et al., 2002, Milo et al., 2002, Shen-Orr et al., 2002). Currently all occurrences of a specified motif pattern are returned. We will study the feasibility of returning only statistically over-represented motifs in future work. tYNA also identifies defective cliques (Yu et al., 2006) that suggest potential missing edges in a network (Fig. S4, S5).

2.4 Multiple-network operations (advanced view)

Multiple-network operations allow one to select multiple networks, perform some operations on each of them, and merge them into a single network. For example, the intersection of multiple high-throughput protein-protein interaction networks offers a high-confidence set of potential interactions. The relationships between different kinds of networks, such as gene regulation and co-expression, can also be studied (Fig. S6).

2.5 Mining and Edge overlap (advanced view)

The edge overlap feature allows the comparison of the edges in two networks. It can test, using some prediction functions, how well one network predicts another. Some prediction functions are predefined, such as identity, sibling and couple (Table S2).

2.6 Saving and downloading analyzed networks

Finally, one may save a working network into the database, or send it to another workspace (as a temporary backup). Likewise, one may download a working network in various network and graphics formats, including SIF, bitmap, postscript and PDF (Fig. S3, S4).

3 IMPLEMENTATION

All source codes were written in Java using standard J2EE architecture. A detailed JavaDoc API is available for users who want to use the classes in their own codes. The tYNA database can also be accessed through standard SOAP-based web services, and we have developed a plugin (available on the tYNA website) to interface tYNA with the network visualization system Cytoscape (Fig. S7).

4 DISCUSSION

As biological network analysis is a vigorous research area, new statistics, motifs, and mining algorithms are expected to emerge continuously. tYNA was thus designed in a modular fashion so that new features can be readily added. Being a Web system, the new features are immediately made accessible to users.

We plan on connecting tYNA with dataset management systems such as YeastHub (Cheung et al., 2005), and Bind (Alfarano et al., 2005), DIP (Xenarios et al., 2002), MINT (Zanzoni et al., 2002), Reactome (Joshi-Tope et al., 2005) and annotation databases to provide a unified platform for performing complex analyses. We think that the combination of the analysis features provided by tYNA and the advanced visualization facilities of Cytoscape can prove particularly powerful. We also plan on interfacing tYNA with analysis and visualization tools such as bioPIXIE (Myers et al., 2005) and Pajek (Batagelj and Mrvar, 2003), which would allow researchers to combine the distinct features of each tool.

REFERENCES

- Alfarano C. et al. (2005) The Biomolecular Interaction Network Database and Related Tools 2005 Update. *Nucleic Acids Research* **33** D418-D424.
- Batagelj V. and Mrvar A. (2003) Pajek – Analysis and Visualization of Large Networks. Jünger, M., Mutzel, P., (Eds). *Graph Drawing Software*, Springer, 77-103.
- Cheung K. H. et al. (2005) YeastHub: A Semantic Web Use Case for Integrating Data in the Life Sciences Domain. *Bioinformatics* **21** Suppl1 i85-196.
- Gavin A-C. et al. (2006) Proteome Survey Reveals Modularity of the Yeast Cell Machinery. *Nature* **440** 631-636.
- Han J-D H. et al. (2004) Evidence for Dynamically Organized Modularity in the Yeast Protein-protein Interaction Network. *Nature* **430** 88-93.
- Ito T. et al. (2000) Toward a Protein-Protein Interaction Map of the Budding Yeast: A Comprehensive System to Examine Two-hybrid Interactions in all Possible Combinations between the Yeast Proteins. *Proc. Natl. Acad. Sci. USA* **98** 4569-4574.
- Joshi-Tope G. et al. (2005) Reactome: A Knowledgebase of Biological Pathways. *Nucleic Acids Research* **33** D428-D432.
- Krogan N. J. et al. (2006) Global Landscape of Protein Complexes in the Yeast *Saccharomyces cerevisiae*. *Nature* advance online publication.
- Lee T. I. et al. (2002) Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*. *Science* **298** 799-804.
- Luscombe N. M. et al. (2004) Genomic Analysis of Regulatory Network Dynamics Reveals Large Topological Changes. *Nature* **431** 308-312.
- Milo R. et al. (2002) Network Motifs: Simple Building Blocks of Complex Networks. *Science* **298** 824-827.
- Myers et al. (2005) Discovery of Biological Networks from Diverse Functional Genomic Data. *Genome Biology* **6**, R114.
- Shannon P. et al. (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13** 2498-2504.
- Sharan R. and Ideker T. (2006) Modeling Cellular Machinery through Biological Network Comparison. *Nature Biotechnology* **24** 427-433.
- Shen-Orr S. S. et al. (2002) Network Motifs in the Transcriptional Regulation Network of *Escherichia coli*. *Nature Genetics* **31** 64-68.
- Uetz P. et al. (2000) A Comprehensive Analysis of Protein-protein Interactions in *Saccharomyces cerevisiae*. *Nature* **403** 623-627.
- Xenarios I. et al. (2002) DIP, the Database of Interacting Proteins: A Research Tool for Studying Cellular Networks of Protein Interactions. *Nucleic Acids Research* **30** 303-305.
- Yu H. et al. (2004) TopNet: A Tool for Comparing Biological Subnetworks, Correlating Protein Properties with Topological Statistics. *Nucleic Acids Research* **32** 328-337.
- Yu H. et al. (2006) Predicting Interactions in Protein Networks by Completing Defective Cliques. *Bioinformatics* **22** 823-829.
- Zanzoni A. et al. (2002) MINT: A Molecular Interaction Database. *FEBS Letters* **513** 135-140.

⁴<http://www.aisee.com/>