Deposited research article

Osprey: A Network Visualization System

Bobby-Joe Breitkreutz*, Chris Stark* and Mike Tyers

Address: Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto M5G 1X5, Canada.

Correspondence: Mike Tyers. E-mail: tyers@mshri.on.ca

GB.c

Software report
Osprey: A Network Visualization System

Bobby-Joe Breitkreutz[*], **Chris Stark**[*], and **Mike Tyers**
Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Canada, M5G 1X5

*These authors contributed equally to this work.

Correspondence: Mike Tyers, Email tyers@mshri.on.ca

Abstract
We have developed a software platform called Osprey for visualization and manipulation of complex interaction networks. Osprey builds data-rich graphical representations that are color-coded for gene function and experimental interaction data. Mouse-over functions allow rapid elaboration and organization of network diagrams in a spoke model format. User-defined large-scale data sets can be readily combined with Osprey for comparison of different methods.

Rationale
The rapidly expanding biological datasets of physical, genetic and functional interactions present a daunting task for data visualization and evaluation [1]. Existing applications such as Pajek allow the user to visualize networks in a simple graphical format [2], but lack the necessary features needed for functional assessment and comparative analysis between datasets. Typically, interaction networks are viewed within a graphing application, but data is manipulated in other contexts, often manually.

To address these shortfalls, we developed a network visualization system called Osprey that not only represents interactions in a flexible and rapidly expandable graphical format, but also provides options for functional comparisons between datasets. Osprey was developed with the Sun Microsystems Java Standard Development Kit version 1.4.0_02 [3], which allows it to be used both in standalone form and as an add-on viewer for online interaction databases.

Network Visualization
Osprey represents genes as nodes and interactions as edges between nodes (Figure 1). Unlike other applications, Osprey is fully customizable and allows the user to define personal settings for generation of interaction networks, as described below. Any interaction dataset can be loaded into Osprey using one of several standard file formats, or by upload from an underlying interaction database. By default, Osprey uses The General Repository for Interaction Datasets as a database (The GRID; http://biodata.mshri.on.ca/grid), from which the user can rapidly build out interaction networks. User-defined interactions are added or subtracted from mouse-over pop-up windows that link to the database. Networks can be saved as tab delimited text files for future manipulation or exported as Joint Photographic Experts Group (JPEG or JPG) graphics, Portable Network Graphics (PNG), and Scalable Vector Graphics (SVG) [4]. The SVG image format allows the user to produce high quality images that can be opened into applications such as Adobe Illustrator [5] for further manipulation.

Searches and Filters
A drawback of current network visualization systems is the inability to search the network for an individual gene in the context of large graphs. To overcome this problem, Osprey allows text search queries by gene names. A further difficulty with visualization systems is the absence of functional information within the graphical interface. This problem is remedied by Osprey, which provides a one-click link to all database fields for all displayed nodes including Open Reading Frame (ORF) name, gene aliases, and a description of gene function. By default, this information is obtained from The GRID, which in turn compiles gene annotations provided by the *Saccharomyces* Genome Database (SGD, ref. 6). Various filters have been developed that allow the user to query the network. For example, an interaction network can be parsed for

interactions derived from a particular experimental method. Current Osprey filters include source, function, experimental system, and connectivity (Figure 2).

Network Layout

As network complexity increases, graphical representations become cluttered and difficult to interpret. Osprey simplifies network layouts through user implemented node relaxation, which disperses nodes and edges according to any one of a number of layout options. Any given node or set of nodes can be locked into place in order to anchor the network. Osprey also provides several default network layouts including circular, concentric circles, spoke, and dual ring (Figure 3). Finally, for comparison of large-scale datasets, Osprey can superimpose two or more datasets on top of each other in an additive manner. In conjunction with filter options, this feature allows interactions specific to any given approach to be identified.

Color Representations

Osprey allows user defined colors to indicate gene function, experimental systems and data sources. Genes are colored by their biological process as defined by standardized Gene Ontology (GO) annotations. Genes that have been assigned more than one process are represented as multicolored pie charts. Osprey currently recognizes 29 biological processes derived from the categories maintained by the GO Consortium [7]. Interactions are colored by experimental system along the entire length of the edge between two nodes. If a given interaction is supported by multiple experimental systems, the edges are segmented into multiple colors to reflect each system. Alternatively, interactions can be colored by data source, again as multiply colored if more than one source supports the interaction. When combined with filter options, a network can be rapidly visualized according to any number of experimental parameters.

Osprey Download

A personal copy of the Osprey Network Visualization System Version 0.9.9 for use in not-for profit organizations can be downloaded from the Osprey web page at http://biodata.mshri.on.ca/osprey. Registration is required for the sole purpose of enabling notification of software fixes and updates. A limited version of Osprey used for online interaction viewing can be used at The GRID website [http://biodata.mshri.on.ca/grid]. For implementation of Osprey as an online viewer for other online interaction databases please contact the authors.

Acknowledgment

References (add numbers)

1. Vidal, M: **A biological atlas of functional maps.** *Cell* 2001, 104: 333-339.
2. Batagelj V, Mrvar A: **Pajek – Program for Large Network Analysis.** *Connections* 1998, 21:47-57.
3. Sun Microsystems Java Standard Development Kit 1.4.0_02 [http://java.sun.com].
4. **Batik SVG Toolkit** [http://xml.apache.org/batik/]
5. **Adobe Illustrator 10** [http://www.adobe.com/products/illustrator/main.html]
6. Cherry, J. M., Ball, C., Dolinski, K., Dwight, S., Harris, M., Matese, J. C., Sherlock, G., Binkley, G., Jin, H., Weng, S., and Botstein, D. **Saccharomyces Genome Database** [ftp://genome-ftp.stanford.edu/pub/yeast/SacchDB/], June 2002.
7. The Gene Ontology Consortium: **Gene Ontology: tool for the unification of biology**. *Nature Genetics* 2000, 25: 25-2
8. Gavin, A.-C. et al. **Functional organization of the yeast proteome by systematic analysis of protein complexes.** Nature 2002;415:141-147.
9. Ho, Y. et al: **Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry.** Nature 2002;415:180-3.

Figure 1. Representative Osprey network with genes colored by GO process and interactions colored by experimental system.
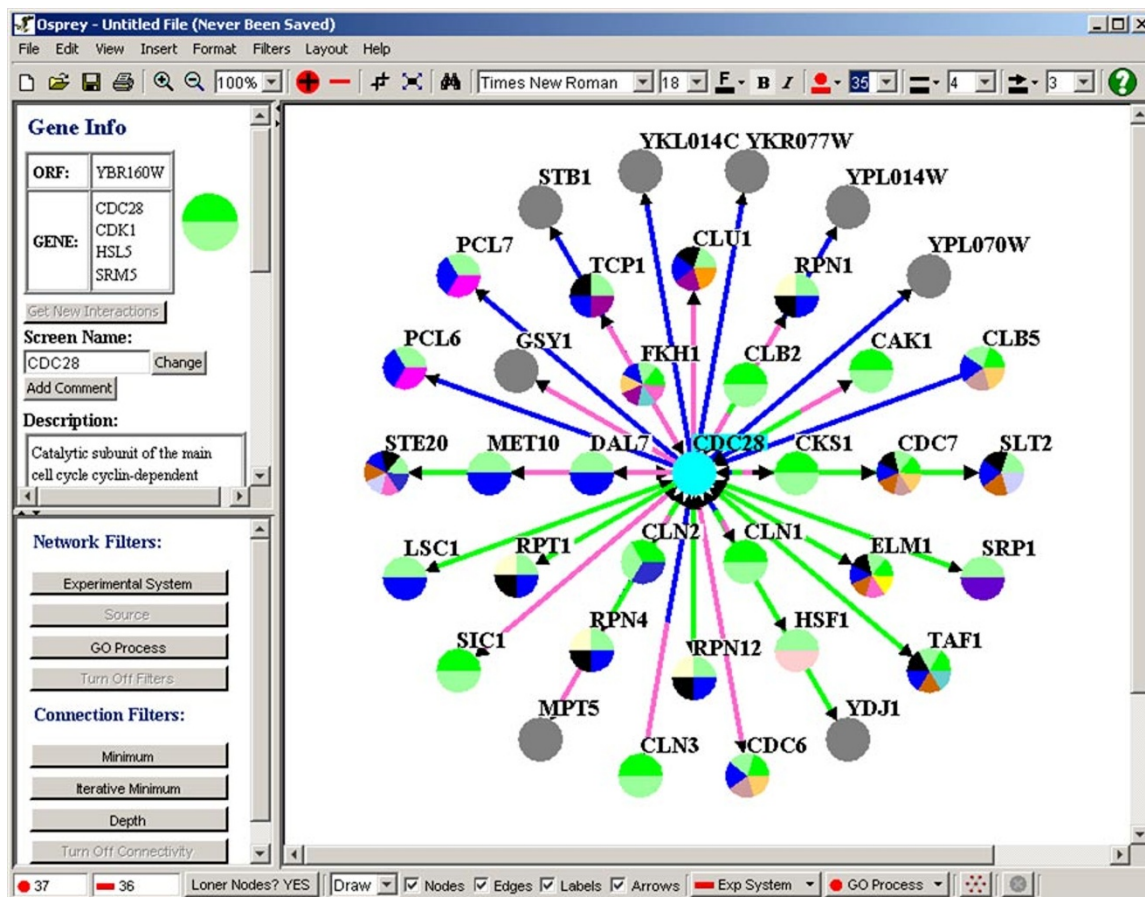
Figure 2. A. Network containing 2245 vertices and 6426 edges from combined datasets of Gavin et al [8], shown in red, and Ho et al [9], shown in yellow. B. A source filter reveals only those interactions shared by both datasets, namely 212 vertices and 188 edges.
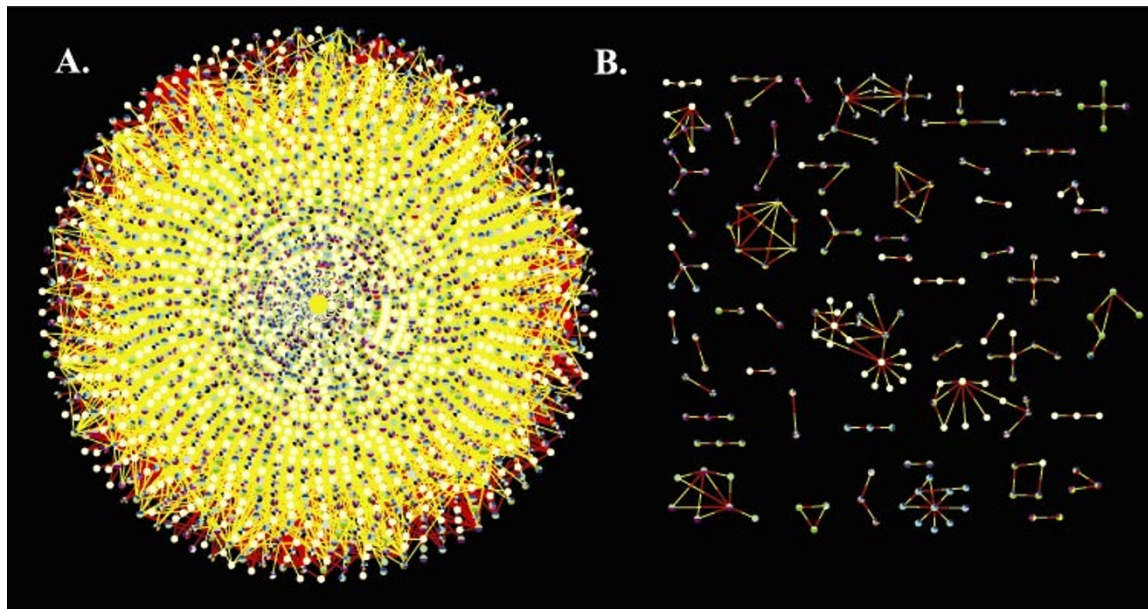
Figure 3. Layout options in Osprey. A. Circular B. Concentric circle with 5 rings C. Dual ring with highly connected nodes on the inside. D. Dual ring with highly connected nodes outside. E. Spoked dual ring.