Cytoscape User Manual

Table of Contents

Cytoscape User Manual

Introduction

Cytoscape is a project dedicated to building open-source network visualization and analysis software. A software "Core" provides basic functionality to layout and query the network and to visually integrate the network with state data. The Core is extensible through a plug-in architecture, allowing rapid development of additional computational analyses and features.

Cytoscape's roots are in Systems Biology, where it is used for integrating biomolecular interaction networks with high-throughput expression data and other molecular state information. Although applicable to any system of molecular components and interactions, Cytoscape is most powerful when used in conjunction with large databases of protein-protein, protein-DNA, and genetic interactions that are increasingly available for humans and model organisms. Cytoscape allows the visual integration of the network with expression profiles, phenotypes, and other molecular state information, and links the network to databases of functional annotations.

The central organizing metaphor of Cytoscape is a network (graph), with genes, proteins, and molecules represented as nodes and interactions represented as links, i.e. edges, between nodes.

Development

Cytoscape is a collaborative project between the Institute for Systems Biology (Leroy Hood lab), the University of California San Diego (Trey Ideker lab), Memorial Sloan-Kettering Cancer Center (Chris Sander lab), the Institut Pasteur (Benno Schwikowski lab), Agilent Technologies (Annette Adler lab) and the University of California, San Francisco (Bruce Conklin lab).

Visit http://www.cytoscape.org for more information.

License

Cytoscape is protected under the GNU LGPL (Lesser General Public License). The License is included as an appendix to this manual, but can also be found online: http://www.gnu.org/copyleft/lesser.txt. Cyto-

scape also includes a number of other open source libraries, which are detailed in the Cytoscape User Manual/Acknowledgements below.

What's the New

Cytoscape contains several new features, plus improvements to the performance and usability of the software. These include:

- New VizMapper user interface (UI)

  - More Intuitive UI.

  - Continuous mapping editors.

  - Visual editor for default view.

  - Visual Mapping Mapping Browser.

  - Improved visual legend generator.

  - Utilities to generate discrete values.

- New Features for Visual Style

  - Transparency (opacity) support.

  - Continuous edge width.

  - Color visual property is separated from Arrow and Edge.

- New Filter UI

  - Use intuitive UI widgets for basic filters

  - Suggest search values with indexing

  - Have options to save in session or globally

- Layout Manager/Broker

- Plugin Manager and new plugin website

  - Install/Update/Delete plugins from within Cytoscape.

  - Search for version compatible plugins from any host site.

  - Display list of installed plugins.

- Undo Manager

- Group API for plugin developers.

- Node stacking

- Tested on both Java SE 5 and 6.

- Many, many bug fixes!

Launching Cytoscape

Cytoscape is a Java application verified to run on Linux, Windows, and Mac OS X. Although not officially supported, other UNIX platforms such as Solaris or FreeBSD may run Cytoscape if Java version 5 or later is available for the platform.

System requirements

The system requirements for Cytoscape depend on the size of the networks the user wants to load, view and manipulate.

Table 1.

	Small Network Visualization	Large Network Analysis/Visualization
Processor	1GHz	As fast as possible
Memory	512MB	2GB+
Graphics Card	On board Video	Highend Graphics Card
Monitor	XGA (1024X768)	Wide or Dual Monitor

Getting Started

Install Java

If not already installed on your computer, download and install Java SE 5 or 6. *Cytoscape will no longer run with Java version 1.4.x or lower. You must install Java SE 5 or 6!!!*

These can be found at:

Java SE 5 Java SE 6

In general, Java SE 6 is faster than 5. If your machine is compatible with the 6 series, please try version 6.

Install Cytoscape

There are a number of options for downloading and installing Cytoscape. All options can be downloaded from the http://cytoscape.org website.

• Automatic installation packages exist for Windows, Mac OS X, and Linux platforms.

• You can install Cytoscape from a compressed archive distribution.

• You can build Cytoscape from the source code.

• You can check out the latest and greatest software from our Subversion repository.

Cytoscape installations (regardless of platform) containing the following files and directories:

Table 2.

File	Description
cytoscape.jar	Main Cytoscape application (Java archive)
cytoscape.sh	Script to run Cytoscape from command line (Linux, Mac OS X)
cytoscape.bat	Script to run Cytoscape (Windows)
LICENSE.txt/html	Cytoscape GNU LGPL License
lib/	library jar files needed to run Cytoscape.
docs/	Manuals in different formats. What you are reading now.
licenses/	Licence files for the various libraries distributed with Cytoscape.
plugins/	Directory containing cytoscape plugins, in .jar format.
sampleData/	
	galFiltered.gml -- Sample molecular interaction network file *
	galFiltered.sif -- Identical network in Simple Interaction Format *
	galExpData.pvals -- Sample gene expression matrix file *
	galFiltered.nodeAttrTable.xls -- Sample node attribute file in Microsoft Excel format
	galFiltered.cys -- Sample session file created from datasets above plus GO Annotations *
	BINDyeast.sif -- Network of all yeast protein-protein interactions in the BIND database **
	BINDhuman.sif -- Network of all human protein-protein interactions in the BIND database**
	yeastHighQuality.sif -- Sample molecular interaction network file ***
	interactome_merged.networkTable.gz -- Human interactome network file in tab-delimited format ****

* From Ideker et al., Science 292:929 (2001)

** Obtained from data hosted at http://www.blueprint.org/bind/bind_downloads.html

*** From von Mering et al., Nature, 417:399 (2002) and Lee et al, Science 298:799 (2002)

**** Created from Cytoscape tutorial web page. Original data sets are available at: http://www.cytoscape.orghttp://cytoscape.org/cgi-bin/moin.cgi/Data_Sets/ from "A merged human interactome" by Andrew Garrow, Yeyejide Adeleye and Guy Warner (Unilever, Safety and Environmental Assurance Center).

Launch the application

Double-click on the icon created by the installer or by running `cytoscape.sh` from the command line (Linux or Mac OS X) or double-clicking `cytoscape.bat` (Windows). Alternatively, you can pass the .jar file to Java directly using the command `java -Xmx512M -jar cytoscape.jar -p plugins`. The `-Xmx512M` flag tells java to allocate more memory for Cytoscape and the `-p plugins` option tells cytoscape to load all of the plugins in the plugins directory. Loading the plugins is important because many key features like layouts, filters and the attribute browser are included with Cytoscape as plugins in the `plugins` directory. See the Command Line chapter for more detail. In Windows, it is also possible to directly double-click the .jar file to launch it. However, this does not allow specification of command-line arguments (such as the location of the plugin directory).

When you succeed in launching Cytoscape, a window will appear that looks like this (captured on Mac OS 10.4):
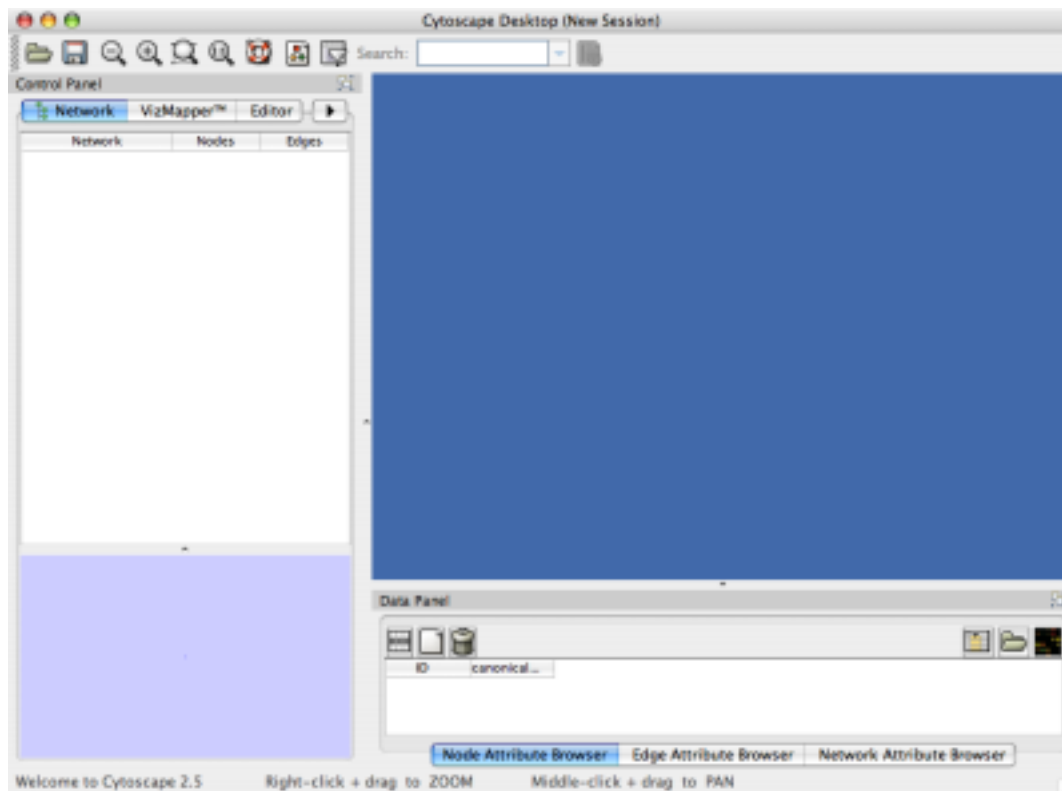
Note on Memory Consumption

For users interested in loading large networks, the amount of memory needed by Cytoscape will increase. Memory usage depends on both number of network objects (nodes+edges) and the number of attributes. Here are some *rough* suggestions for memory allocation:

*Suggested Memory Size **Without** View*

Table 3.

Number of Objects (nodes + edges)	Suggested Memory Size
0 - 70,000	512M (default)
70,000 - 150,000	800M

*Suggested Memory Size **With** View*

Table 4.

Number of Objects (nodes + edges)	Suggested Memory Size
0 - 20,000	512M (default)
20,000 - 70,000	800M
70,000 - 150,000	1G

To increase the maximum memory size for Cytoscape, you can specify it from command line. For example, if you want to assign 1GB of memory, type:

*java -Xmx1GB -jar cytoscape.jar -p plugins*

from the command line.

For more details, see How to increase memory for Cytoscape.

Note on Directory Location

For the application to work properly, all files should be left in the directory in which they were unpacked. The core Cytoscape application assumes this directory structure when looking for the various libraries needed to run the application. If you are adventurous, you can get creative with the $CLASSPATH and/or the cytoscape.jar manifest file and run Cytoscape from any location you want.

Quick Tour of Cytoscape

When a network is loaded, Cytoscape will look something like the image below:
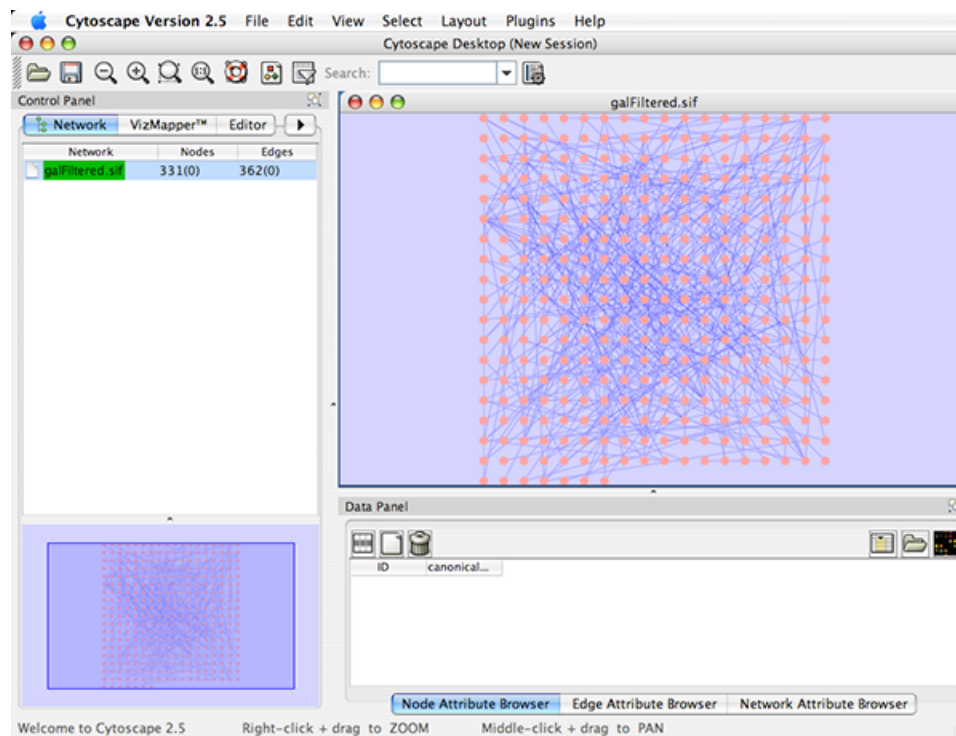
The main window here has several components:

- The menu bar at the top (see below for more information about each menu).

- The toolbar, which contains icons for commonly used functions. These functions are also available via the menus. Hover the mouse pointer over an icon and wait momentarily for a description to appear as a tooltip.

- The network management panel (top left panel). This contains an optional network overview pane (shown at the bottom left).

- The main network view window, which displays the network.

- The attribute browser panel (bottom panel), which displays attributes of selected nodes and edges and enables you to modify the values of attributes.

The network management and attribute browser panels are dockable tabbed panels known as CytoPanels. You can undock any of these panels by clicking on the Float Window control


in the upper-right corner of the CytoPanel.

If you select this control, e.g. on the attribute browser panel, you will now have two Cytoscape windows, the main window, and a new window labeled CytoPanel 2, similar to the one shown below.

-
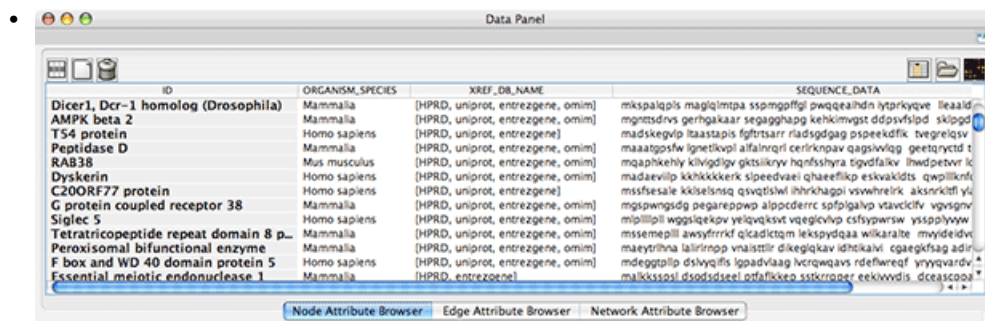
Note that CytoPanel 2 now has a Dock Window control. If you select this control, the window will dock onto the main window.

Cytoscape also has an editor that enables you to build and modify networks interactively by dragging and dropping nodes and edges from a palette onto the main network view window. The Node shapes and Edge arrows on the palette are defined by the currently used Visual Style. To edit a network, just select the Editor tab on CytoPanel 1. An example of an editor, with the palette contained in CytoPanel 1 and defined by the BioMoleculeEditor Visual Style, is shown below.
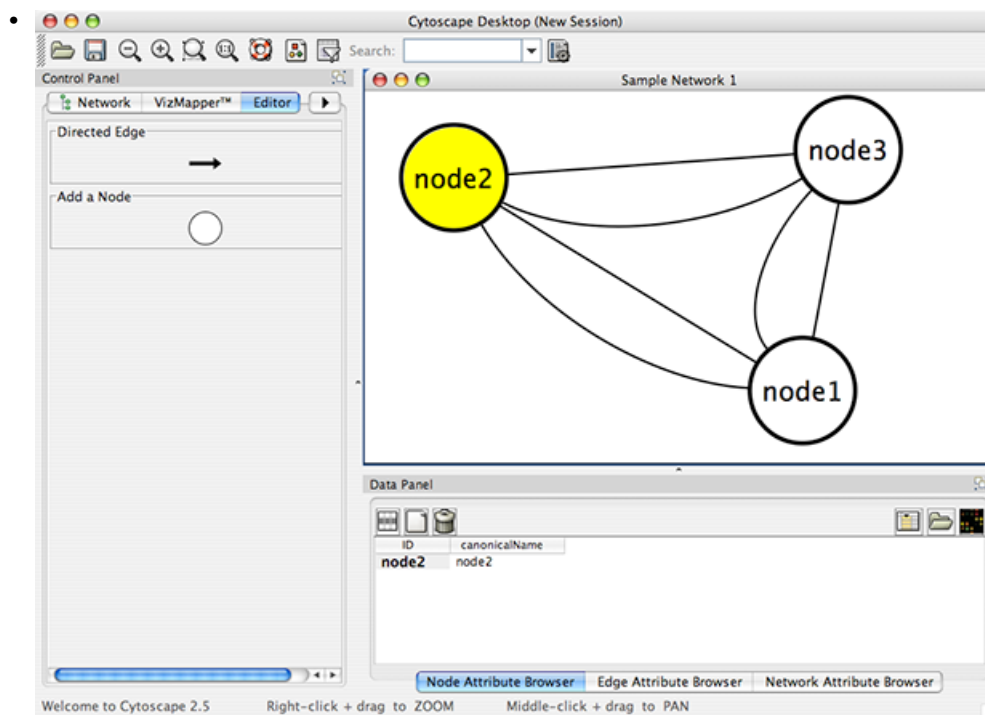
The Menus

File

The File menu contains most basic file functionality: File → Open for opening a Cytoscape session file; File → New for creating a new network, either blank for editing, or from an existing network; File → Save for saving a session file; File → Import for importing data such as networks and attributes; and File → Export for exporting data and images. Also, File → Print allows printing, while File → Quit closes all windows of Cytoscape and exits the program.
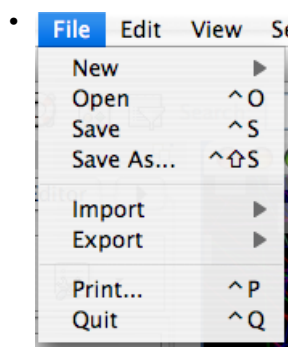

Edit

The Edit menu contains Undo and Redo functions which undo and redo edits made in the Attribute Browser, the Network Editor and to layout.

There are also options for creating and destroying views (graphical representations of a network) and networks (the raw network data – not yet visualized), as well as an option for deleting selected nodes and

edges from the current network. All deleted nodes and edges can be restored to the network via Edit → Undo. Editing preferences for properties and plugins is found under Edit → Preferences → Properties... .

-
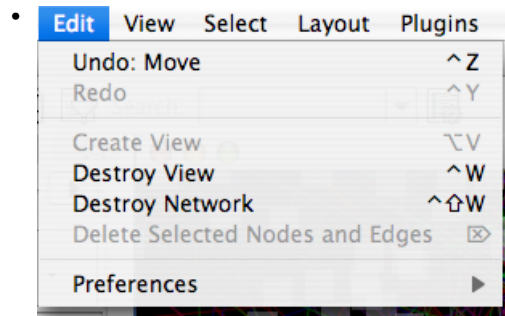
View

The View menu allows you to display or hide the network management panel (CytoPanel 1), the attribute browser (CytoPanel 2), the Network Overview (in CytoPanel 1), and the VizMapper.

-
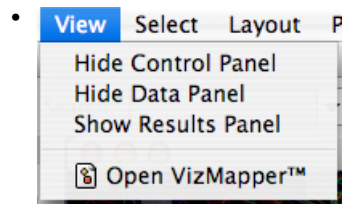
Select

The Select menu contains different options for selecting nodes and edges. It also contains the Select → Use Filters option, which allows filters to be created for automatic selection of portions of a network whose node or edge attributes meet a filtering criterion.

-
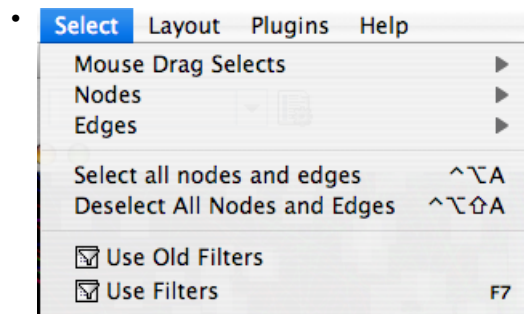
Layout

The Layout menu has an array of features for visually organizing the network. The features in the top portion of the network (Rotate, Scale, Align and Distribute) are tools for manipulating the network visualization. The bottom section of the menu lists a variety of layout algorithms which automatically lay a network out.

-
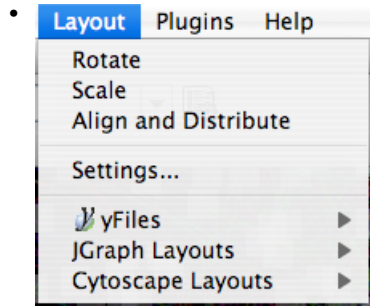
Plugins

The Plugins menu contains options for managing (install/update/delete) your plugins and may have options added by plugins that have been installed, such as the Agilent Literature Search or Merge Networks. Depending on which plugins are loaded, the plugins that you see may be different than what appear here.

-
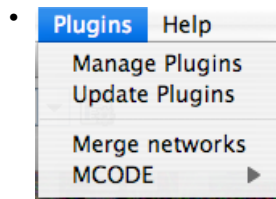
Table 5.

Note: A list of available Cytoscape plugins with descriptions is available online at: http://cytoscape.org/plugins2.php

Help

The Help menu allows you to launch the online help viewer and browse the table of contents for this manual. The "About…" option displays information about the running version of Cytoscape.

-
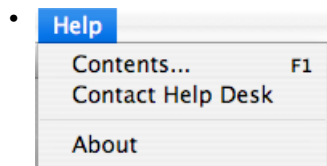
Network Management

A network stores all the nodes and edges that are loaded by the user and a view displays them. You can have many views of the same network. Networks (and their optionally associated views) can be organized hierarchically.

An example where a number of networks have been loaded and arranged hierarchically is shown below:
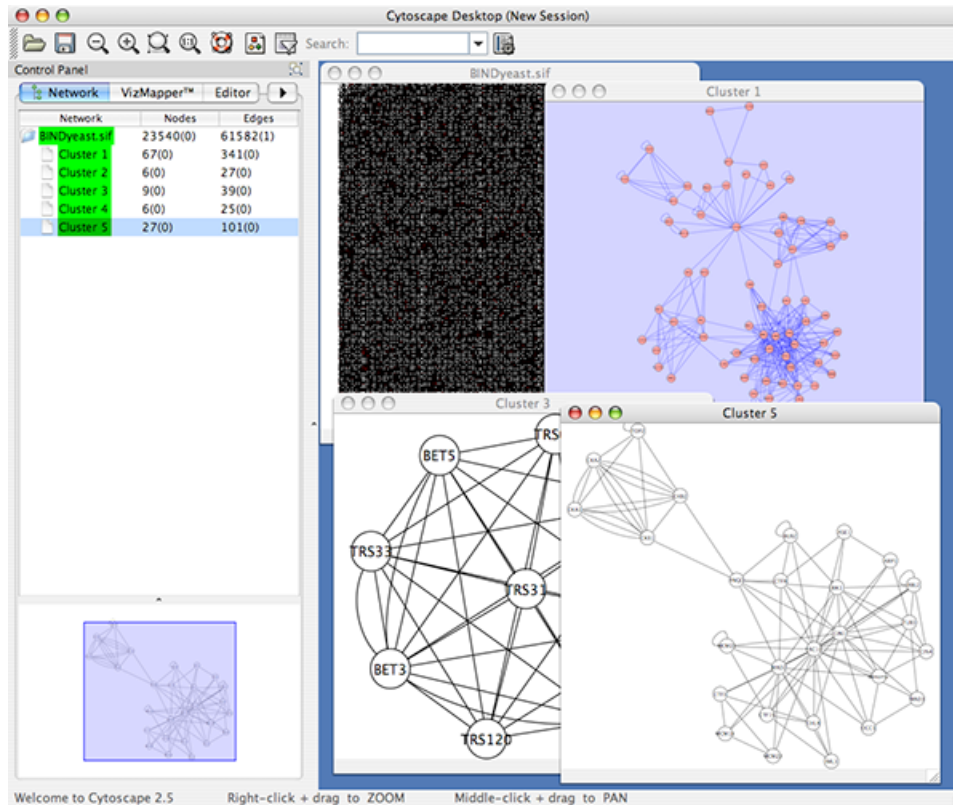
The network manager (top-right tree view in CytoPanel 1) shows the networks that are loaded. Clicking on a network here will make that view active in the main window, if the view exists (green highlighted networks only). Each network has a name and size (number of nodes and edges), which are shown in the network manager. If a network is loaded from a file, the network name is the name of the file.

Some networks are very large (thousands of nodes and edges) and can take a long time to display. For this reason, a network in Cytoscape may not contain a 'view'. Networks that have a view are highlighted in green and networks that don't have a view are highlighted in red. You can create or destroy a view for a network by right-clicking the network name in the network manager or by choosing the appropriate option in the Edit menu. You can also destroy previously loaded networks this way. In the picture above, seven networks are loaded, six green ones with views and one red one without a view.

Certain operations in Cytoscape will create new networks. If a new network is created from an old network, for example by selecting a set of nodes in one network and copying these nodes to a new network (via the File → New → Network option), it will be shown as a child of the network that it was derived from. In this way, the relationships between networks that are loaded in Cytoscape can be seen at a glance. Networks in the top part of the tree in the figure above were generated in this manner.

The available network views are also arranged as multiple overlapping windows in the network view window. You can maximize, minimize, and destroy network views by using the normal window controls for your operating system.

The Network Overview Window

The network overview window shows an overview (or 'bird's eye view') of the network. It can be used to navigate around a large network view. The blue rectangle indicates the portion of the network currently

displayed in the network view window, and it can be dragged with the mouse to view other portions of the network. Zooming in will cause the rectangle to appear smaller and vice versa.

-

-
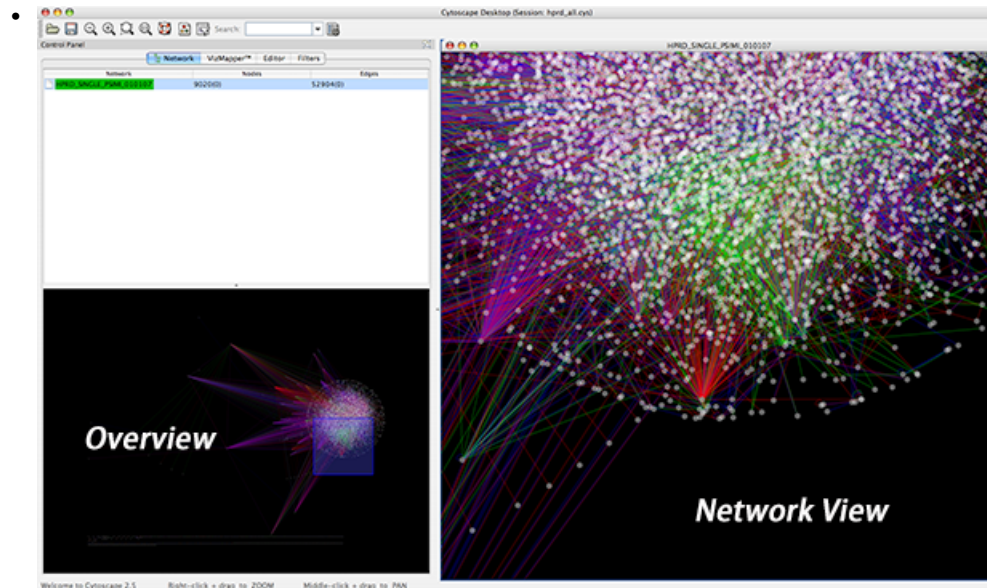
Command Line Arguments

Cytoscape recognizes a number of optional command line arguments, including run-time specification of network files, attribute files, and session files. This is the output generated when the cytoscape is executed with the "-h" or "--help" flag:

```
usage: java -Xmx512M -jar cytoscape.jar [OPTIONS]
 -h,--help                     Print this message.
 -v,--version                  Print the version number.
 -s,--session <file>        Load a cytoscape session (.cys) file.
 -N,--network <file>     Load a network file (any format).
 -e,--edge-attrs <file>    Load an edge attributes file (edge attribute format).
 -n,--node-attrs <file>   Load a node attributes file (node attribute format).
 -m,--matrix <file>        Load a node attribute matrix file (table).
 -p,--plugin <file>         Load a plugin jar file, directory of jar files,
                                  plugin class name, or plugin jar URL.
 -P,--props <file>         Load cytoscape properties file (Java properties
                                  format) or individual property: -P name=value.
 -V,--vizmap <file>       Load vizmap properties file (Java properties format).
```

Any file specified for an option may be specified as either a path or as a URL. For example you can specify a network as a file (assuming that myNet.sif exists in the current working directory): `cytoscape.sh -N myNet.sif`. Or you can specify a network as a URL: `cytoscape.sh -N http://example.com/myNet.sif`.

Table 6.

Argument	Description
-h,--help	This flag generates the help output you see above and exits.
-v,--version	This flag prints the version number of Cytoscape and exits.
-s,--session <file>	This option specifies a session file to be loaded. Since only one session file can be loaded at a given time, this option may only specified once on a given command line. The option expects a .cys Cytoscape session file. It is customary, although not necessary, for session file names to contain the .cys extension.
-N,--network <file>	This option is used to load all types of network files. SIF, GML, and XGMML files can all be loaded using the -N option. You can specify as many networks as desired on a single command line.
-e,--edge-attrs <file>	This option specifies an edge attributes file. You may specify as many edge attribute files as desired on a single command line.
-n,--node-attrs <file>	This option specifies a node attributes file. You may specify as many node attribute files as desired on a single command line.
-m,--matrix <file>	This option specifies a data matrix file. In a biological context, the data matrix consists of expression data. All data matrix files are read into node attributes. You may specify as many data matrix files as desired on a single command line.
-p,--plugin <file>	This option specifies a cytoscape plugin (.jar) file to be loaded by Cytoscape. This option also subsumes the previous "resource plugin option". You may specify a class name that identifies your plugin and the plugin will be loaded if the plugin is in Cytoscape's CLASSPATH. For example, assuming that the class MyPlugin can be found in the CLASSPATH, you could specify the plugin like this: `cytoscape.sh -p MyPlugin.class`. A final means of specifying plugins is to specify a file name whose contents contain a list of plugin jar files.
-P,--props <file>	This option specifies Cytoscape properties. Properties can be specified either as a properties file (in Java's standard properties format), or as individual properties. To specify individual properties, you must specify the property name followed by the property value where the name and value are separated by the '=' sign. For example to specify the defaultSpeciesName: `cytoscape.sh -P defaultSpeciesName=Human`. If you would like to include spaces in your property, simply enclose the name and value in quotation marks: `cytoscape.sh -P "defaultSpeciesName=Homo Sapiens"`. The property option subsumes previous options -noCanonicalization, -species, and -bioDataServer. Now it would look like: `cytoscape.sh -P defaultSpeciesName=Human -P noCanonicalization=true -P bioDataServer=myServer`.
-V,--vizmap <file>	This option specifies a visual properties file.

All options described above (including plugins) can be loaded from the GUI once Cytoscape is running.

-

Cytoscape Preferences

Managing Properties

Table 7.

> **Important!** If you have used previous versions of Cytoscape, you will notice that handling of properties has changed. The most important change is that properties are no longer saved by default to the current directory or to your home `.cytoscape` directory. Properties are stored by default in Cytoscape session files (`.cys` extension). The `cytoscape.props` file still exists in the `.cytoscape` directory but is only written to when the user explicitly requests that the current settings be made the defaults for all future sessions of Cytoscape. Unless you have something important in your `.cytoscape/cytoscape.props` file, your best bet will be to delete the file and use the defaults.

The Cytoscape Properties editor, accessed via Edit → Preferences → Properties…, is used to specify general and default properties. Properties are now stored in Cytoscape session files, so changes to general properties will be saved as part of the current session, but will only carry over to subsequent sessions if they are set as defaults or exported using the File → Export function.

Cytoscape properties are configurable via Add, Modify and Delete operations.

-
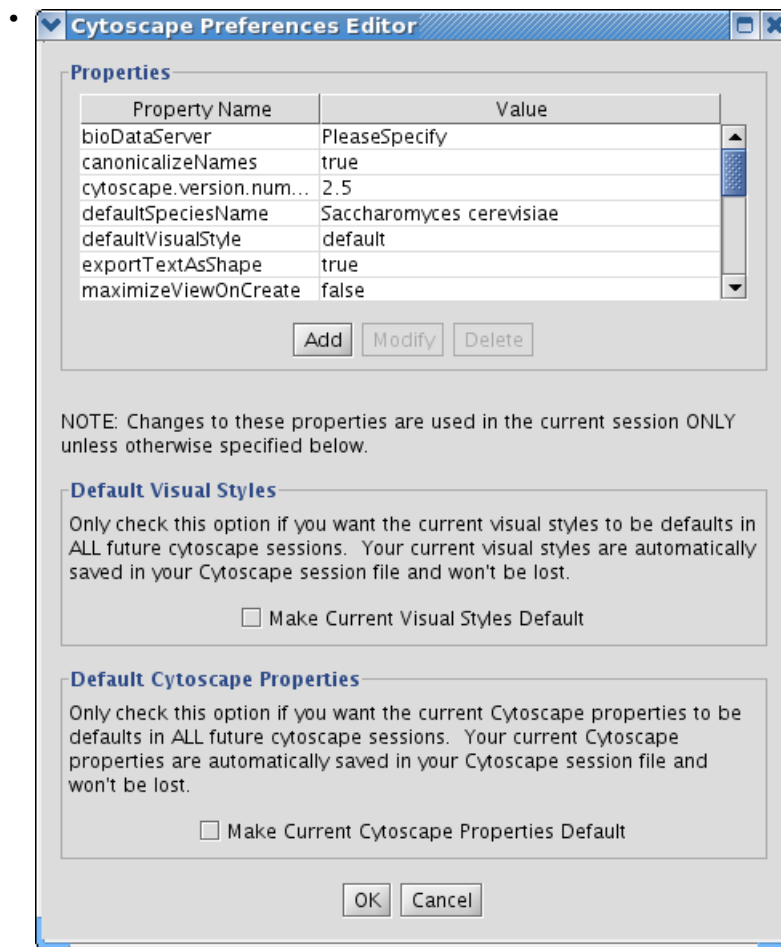
Some common properties are described below.

Table 8.

| Property name	Default value	Valid values
viewThreshold	10000	integer > 0
secondaryViewThreshold	30000	integer > 0
viewType	tabbed	tabbed
defaultWebBrowser		A path to the web browser on your system. This only needs to be specified if Cytoscape can't find the web browser on your system.

Setting Default Properties

It is possible to alter the default properties for Cytoscape.

Edit the properties via Edit → Preferences → Properties... and check the Make Current Cytoscape Properties Default checkbox. This will save the current properties to the `.cytoscape` directory, where they will then be applied to all of your Cytoscape sessions from that point on. Otherwise, Cytoscape will automatically save the properties used in a particular session inside its `.cys` session file, while the default properties will be applied at the beginning of subsequent sessions.

Managing Bookmarks

Cytoscape contains a pre-defined list of bookmarks, which point to sample network files located on the Cytoscape web server. Users may add, modify, and delete bookmarks through the Bookmark manager, accessed by going to Edit → Preferences → Bookmarks… .

-
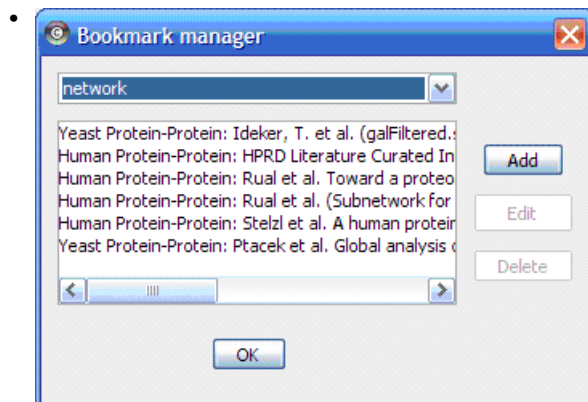
There are currently two types of bookmarks: network and annotation. Network bookmarks are URLs pointing to network files available on the Internet. These are nomal networks that can be loaded into Cytoscape. The annotation bookmarks are URLs pointing to ontology annotation files. The annotation bookmarks are only used when importing an ontology.

Managing Proxy Servers

You can define and configure a proxy server for Cytoscape by going to Edit → Preferences → Proxies… .

-
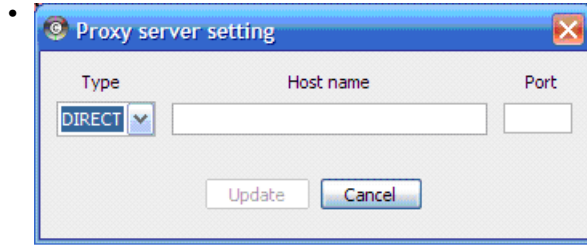
After the proxy server is set, all network traffic related to loading a network from URL will pass through the proxy server. Other plugins use this capability as well.

Creating Networks

There are 3 different ways of creating networks in Cytoscape:

1. Importing pre-existing, formatted network files.

2. Importing pre-existing, unformatted text or Excel files.

3. Creating an empty network and manually adding nodes and edges.

# Import Fixed-Format Network Files

Network files can be specified in any of the formats described in the Supported Network Formats chapter. Networks are imported into Cytoscape through the "Import Network" window, which can be accessed by going to File → Import → Network (multiple file types). The network file can either be located directly on the local computer, or found on a remote computer (in which case it will be referenced with a URL).

Load Networks from Local Computer

By default, Cytoscape loads networks from the local computer.

The Import Networks dialog shows a default setting of "Data Source Type: Local," meaning that network files from the local computer will be available for importing. Choose the correct file by clicking on the Select button (only file types that Cytoscape recognizes will be shown), and then load the network by clicking on the Import button. Some sample network files of different types have been included in the sampleData folder in Cytoscape.

Network files in SIF, GML, and XGMML formats may also be loaded directly from the command line using the –N option.
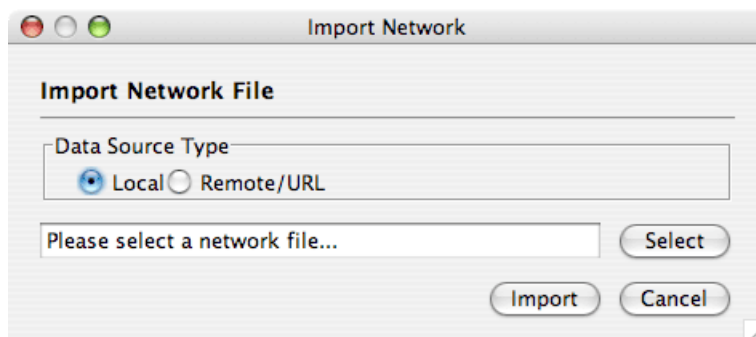
Load Networks from a Remote Computer (URL import)

The Import Networks dialog is also capable of importing network files using a URL. To do this, set the Data Source Type to Remote and insert the appropriate URL, either manually or using URL bookmarks. Bookmarked URLs can be accessed by clicking on the arrow to the right of the text field (see the Bookmark Manager in Preferences for more details on bookmarks). Also, you can drag and drop links from web browser to the URL text box. Once a URL has been specified, click on the Import button to load the network.
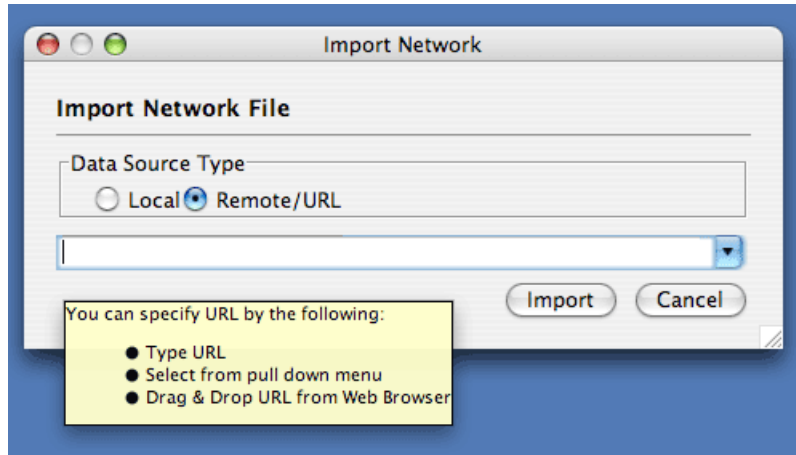

Importing networks from URL addresses has an important caveat. Because Cytoscape determines file type primarily (not exclusively) by file extension, it can have trouble importing networks with URLs that don't end in a human readable file name. If Cytoscape does not recognize a meaningful file name and extension in the URL, it will attempt to guess the type of file based on MIME type. If the MIME type is not recognizable to any of our import handlers, then the import will fail.

Another issue for network import is the presence of firewalls, which can affect which files are accessible to a computer. To work around this problem, Cytoscape supports the use of proxy servers. To configure the proxy server, go to Edit → Preferences→ Proxy Server... . This is further described in the Preferences chapter.

Import Free-Format Table Files

Cytoscape now supports the import of networks from delimited text files and Excel workbooks using Edit → Import → Network from Table (Text/MS Excel)... . An interactive GUI allows users to specify parsing options for specified files. The screen provides a preview that shows how the file will be parsed given the current configuration. As the configuration changes, the preview updates automatically. In addition to specifying how the file will be parsed, the user must also choose the columns that represent the Source nodes, the Target nodes, and an optional edge interaction type.
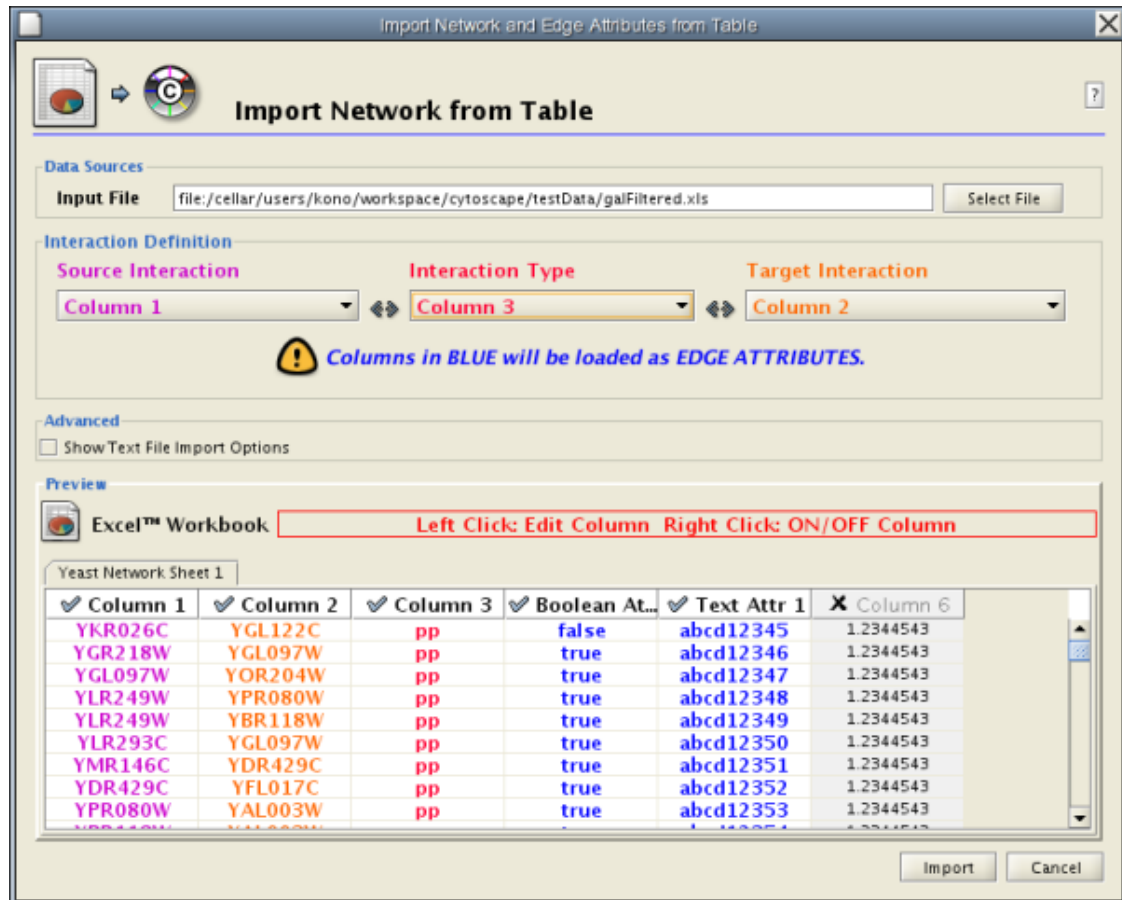
Supported Files

The "Import Network from Table" function supports delimited text files and single-sheet Microsoft Excel Workbooks. The following is a sample table file:

```
source  target  interaction  boolean attribute  string attribute      floating point attribute
YJR022W YNR053C pp           TRUE      abcd12371  1.2344543
YER116C YDL013W pp           TRUE      abcd12372  1.2344543
YNL307C YAL038W pp           FALSE     abcd12373  1.2344543
YNL216W YCR012W pd           TRUE      abcd12374  1.2344543
YNL216W YGR254W pd           TRUE      abcd12375  1.2344543
```

The network table files should contain at least two columns: source nodes and target nodes. The interaction type is optional in this format. Therefore, a minimal network table looks like the following:

```
YJR022W YNR053C
YER116C YDL013W
YNL307C YAL038W
YNL216W YCR012W
YNL216W YGR254W
```

One row in a network table file represents an edge and its edge attributes. This means that a network file is considered a combination of network data and edge attributes. A table may contain columns that aren't meant to be edge attributes. In this case, you can choose not to import those columns by clicking on the column header in the preview window. This function is useful when importing a data table like the following (1):

```
Unique ID A   Unique ID B   Alternative ID A   Alternative ID B   Aliases A   Aliases B   Interaction detection methods   First author surnames   Pubmed IDs   species A

7205   5747   TRIP6   PTK2   Q15654   Q05397-1   vv|HPRD Currently not available 14688263|15892868(Marcotte)   Mammalia   Homo sapiens   protein|protein HPRD|Marcotte   0   Thyro
```

```
4174    7311    MCM5    UBA52   P33992  P62987  neighbouring_reaction   Currently not available 15608231(Reactome)       Homo sapiens    Homo sapiens    protein|protein Reactome        1       P3399

7040    7040    TGFB1   TGFB1   P01137  P01137  nmr: nuclear magnetic resonance Currently not available 8679613 Homo sapiens    Homo sapiens    protein|protein BIND    2       TGFB1-TGFB1-    72085
```

This data file is a tab-delimited text and contains network data (interactions), edge attributes, and node attributes. To import network and edge attributes from this table, you need to choose Unique ID A as source, Unique ID B as target, and Interactor types as interaction type. Then you need to turn off columns used for node attributes (Alternative ID A, species B, etc.). Other columns can be imported as edge attributes.

The network import function cannot import node attributes - only edge attributes. To import node attributes from this table, please see the Attributes section of this manual.

Note (1): This data is taken from the *A merged human interactome* datasets by Andrew Garrow, Yeyejide Adeleye and Guy Warner (Unilever, Safety and Environmental Assurance Center, 12 October 2006). Actual data files are available at http://www.cytoscape.orghttp://cytoscape.org/cgi-bin/moin.cgi/Data_Sets/

Basic Operations

To import network text/Excel tables, please follow these steps:

1. Select File → Import → Network from Table (Text/MS Excel)...

2. Select a table file by clicking on the Select File button.

3. Define the interaction parameters by specifying which columns of data contain the Source Interaction, Target Interaction, and Interaction Type. Setting the Interaction Type as Default Interaction will result in all interactions being given the value pp; this value can be modified in Advanced Options (below).

4. (Optional) Define edge attribute columns, if applicable. Network table files can have edge attribute columns in addition to network data.

   - Enable/Disable Attribute Column - By *left*-clicking on a column header in the preview table, you can enable/disable edge attributes. If the header is checked and entries are blue, the column will be imported as an edge attribute. For example, the table below shows that columns 1 through 3 will be used as network data, column 4 will not be imported, and columns 5 and 6 will be imported as edge attributes.

   -

   - Change Attribute Name and Data Types - If you *right*-click on a column header in the preview table, you can modify the attribute name and data type. For more detail, see "Modify Attribute Name/Type" below.

5. Click the Import button.

Advanced Options

You can select several options by checking the Show Text File Import Options checkbox.

- Delimiter: You can select multiple delimiters for text tables. By default, Tab and Space are selected as delimiters.

- Preview Options: When you select a network table file, the first 100 entries will be displayed in the Preview panel. To display more entries, change the value for this option. If you want to show all entries in the file, select "Show all entries in the file". You will need to click the Reload button to update the Preview panel.

- Attribute Names

  - Transfer first line as attribute names: Selecting this option will cause all edge attribute columns to be named according to the first data entry in that column.

  - Start Import Row: Set which row of the table to begin importing data from. For example, if you want to skip the first 3 rows in the file, set 4 for this option.

  - Comment Line: Rows starting with this character will not be imported. This option can be used to skip comment lines in text files.

- Network Import Options: If the Interaction Type is set to Default Interaction, the value here will be used as the interaction type for all edges.

Modify Attribute Name/Type

Attribute names and data types can be modified here.

- Modify Attribute Name - just enter a new attribute name and click OK.

- Modify Attribute Data Type - The following attribute data types are supported:

  - String

  - Boolean (True/False)

  - Integer

  - Floating Point

  - List of (one of) String/Boolean/Integer/Floating Point

Cytoscape has a basic data type detection function that automatically suggests the attribute data type of a column according to its entries. This can be overridden by selecting the appropriate data type from the radio buttons provided. For lists, a global delimiter must be specified (i.e., all cells in the table must use the same delimiter).

Edit a New Network

A new, empty network can also be created and nodes and edges manually added. To create an empty network, go to File → New → Network → Empty Network, and then manually add network components using the Editor in CytoPanel 1 (see the Editor chapter for more details).

Supported Network File Formats

Cytoscape can read network/pathway files written in the following formats:

- Simple interaction file (SIF or .sif format)

- Graph Markup Language (GML or .gml format)

- XGMML (extensible graph markup and modelling language).

- SBML

- BioPAX

- PSI-MI Level 1 and 2.5

- Delimited text

- Excel Workbook (.xls)

The SIF format specifies nodes and interactions only, while other formats store additional information about network layout and allow network data exchange with a variety of other network programs and data sources. Typically, SIF files are used to import interactions when building a network for the first time, since they are easy to create in a text editor or spreadsheet. Once the interactions have been loaded and network layout has been performed, the network may be saved to GML or XGMML format for interaction with other systems. All file types listed (except Excel) are text files and you can edit and view them in a regular text editor.

SIF Format

The simple interaction format is convenient for building a graph from a list of interactions. It also makes it easy to combine different interaction sets into a larger network, or add new interactions to an existing data set. The main disadvantage is that this format does not include any layout information, forcing Cytoscape to re-compute a new layout of the network each time it is loaded.

Lines in the SIF file specify a source node, a relationship type (or edge type), and one or more target nodes:

```
nodeA <relationship type> nodeB
nodeC <relationship type> nodeA
nodeD <relationship type> nodeE nodeF nodeB
nodeG
...
nodeY <relationship type> nodeZ
```

A more specific example is:

```
node1 typeA node2
node2 typeB node3 node4 node5
node0
```

The first line identifies two nodes, called node1 and node2, and a single relationship between node1 and node2 of type typeA. The second line specifies three new nodes, node3, node4, and node5; here "node2" refers to the same node as in the first line. The second line also specifies three relationships, all of type typeB and with node2 as the source, with node3, node4, and node5 as the targets. This second form is simply shorthand for specifying multiple relationships of the same type with the same source node. The third line indicates how to specify a node that has no relationships with other nodes. This form is not needed for nodes that do have relationships, since the specification of the relationship implicitly identifies the nodes as well.

Duplicate entries are ignored. Multiple edges between the same nodes must have different edge types. For example, the following specifies two edges between the same pair of nodes, one of type xx and one of type yy:

```
node1 xx node2
node1 xx node2
node1 yy node2
```

Edges connecting a node to itself (self-edges) are also allowed:

```
node1 xx node1
```

Every node and edge in Cytoscape has an identifying name, most commonly used with the node and edge data attribute structures. Node names must be unique, as identically named nodes will be treated as identical nodes. The name of each node will be the name in this file by default (unless another string is mapped to display on the node using the visual mapper). This is discussed in the section on visual styles. The name of each edge will be formed from the name of the source and target nodes plus the interaction type: for example, sourceName (edgeType) targetName.

The tag <relationship type> can be any string. Whole words or concatenated words may be used to define types of relationships, e.g. geneFusion, cogInference, pullsDown, activates, degrades, inactivates, inhibits, phosphorylates, upRegulates, etc.

Some common interaction types used in the Systems Biology community are as follows:

```
pp ................. protein - protein interaction
pd ................. protein -> DNA
 (e.g. transcription factor binding upstream of a regulating gene.)
```

Some less common interaction types used are:

```
pr ................ protein -> reaction
rc ................ reaction -> compound
cr ................ compound -> reaction
gl ................ genetic lethal relationship
pm ................ protein-metabolite interaction
mp ................ metabolite-protein interaction
```

Delimiters

Whitespace (space or tab) is used to delimit the names in the simple interaction file format. However, in some cases spaces are desired in a node name or edge type. The standard is that, if the file contains any tab characters, then tabs are used to delimit the fields and spaces are considered part of the name. If the file contains no tabs, then any spaces are delimiters that separate names (and names cannot contain spaces).

If your network unexpectedly contains no edges and node names that look like edge names, it probably means your file contains a stray tab that's fooling the parser. On the other hand, if your network has nodes whose names are half of a full name, then you probably meant to use tabs to separate node names with spaces.

Networks in simple interactions format are often stored in files with a `.sif` extension, and Cytoscape recognizes this extension when browsing a directory for files of this type.

GML Format

In contrast to SIF, GML is a rich graph format language supported by many other network visualization packages. The GML file format specification is available at:

http://www.infosun.fmi.uni-passau.de/Graphlet/GML/

It is generally not necessary to modify the content of a GML file directly. Once a network is built in SIF format and then laid out, the layout is preserved by saving to and loading from GML. Visual attributes specified in a GML file will result in a new visual style named `Filename.style` when that GML file is loaded.

XGMML Format

XGMML is the XML evolution of GML and is based on the GML definition. In addition to network data, XGMML contains node/edge/network attributes. The XGMML file format specification is available at:

http://www.cs.rpi.edu/~puninj/XGMML/

XGMML is now preferred to GML because it offers the flexibility associated with all XML document types. If you're unsure about which to use, choose XGMML.

SBML (Systems Biology Markup Language) Format

The Systems Biology Markup Language (SBML) is an XML format to describe biochemical networks. SBML file format specification is available at:

http://sbml.org/documents/

BioPAX (Biological PAthways eXchange) Format

BioPAX is an OWL (Web Ontology Language) document designed to exchange biological pathways data. The complete set of documents for this format is available at:

http://www.biopax.org/index.html

PSI-MI Format

The PSI-MI format is a data exchange format for protein-protein interactions. It is an XML format used to describe PPI and associated data. PSI-MI XML format specification is available at:

http://psidev.sourceforge.net/mi/xml/doc/user/

Delimited Text Table and Excel Workbook

Cytoscape has native support for Microsoft Excel files (.xls) and delimited text files. The tables in these files can have network data and edge attributes. Users can specify columns containg source nodes, target nodes, interaction types, and edge attributes during file import. For more detail, please read the Import Free-Format Tables section section of the Creating Networks chapter.

Node Naming Issues in Cytoscape

Typically, genes are represented by nodes, and interactions (or other biological relationships) are represented by edges between nodes. For compactness, a gene also represents its corresponding protein. Nodes may also be used to represent compounds and reactions (or anything else) instead of genes.

If a network of genes or proteins is to be integrated with Gene Ontology (GO) annotation or gene expression data, the gene names must exactly match the names specified in the other data files. We strongly encourage naming genes and proteins by their systematic ORF name or standard accession number; common names may be displayed on the screen for ease of interpretation, so long as these are available to the program in the annotation directory or in a node attribute file. Cytoscape ships with all yeast ORF-to-common name mappings in a synonym table within the `annotation/` directory. Other organisms will be supported in the future.

Why do we recommend using standard gene names? All of the external data formats recognized by Cytoscape provide data associated with particular names of particular objects. For example, a network of protein-protein interactions would list the names of the proteins, and the attribute and expression data would likewise be indexed by the name of the object.

The problem is in connecting data from different data sources that don't necessarily use the same name for the same object. For example, genes are commonly referred to by different names, including a formal "location on the chromosome" identifier and one or more common names that are used by ordinary researchers when talking about that gene. Additionally, database identifiers from every database where the gene is stored may be used to refer to a gene (e.g. protein accession numbers from Swiss-Prot). If one data source uses the formal name while a different data source used a common name or identifier, then Cytoscape must figure out that these two different names really refer to the same biological entity.

Cytoscape has two strategies for dealing with this naming issue, one simple and one more complex. The simple strategy is to assume that every data source uses the same set of names for every object. If this is the case, then Cytoscape can easily connect all of the different data sources.

To handle data sources with different sets of names, as is usually the case when manually integrating gene information from different sources, Cytoscape needs a data server that provides synonym information (see the chapter on Annotation). A synonym table gives a canonical name for each object in a given organism and one or more recognized synonyms for that object. Note that the synonym table itself defines which set of names are the "canonical" names. For example, in budding yeast, the ORF names are commonly used as the canonical names.

If a synonym server is available, then by default Cytoscape will convert every name that appears in a data file to the associated canonical name. Unrecognized names will not be changed. This conversion of names to a common set allows Cytoscape to connect the genes present in different data sources, even if they have different names – as long as those names are recognized by the synonym server.

For this to work, Cytoscape must also be provided with the species to which the objects belong, since the data server requires the species in order to uniquely identify the object referred to by a particular name. This is usually done in Cytoscape by specifying the species name on the command line with the –P option (`cytoscape.sh -P "defaultSpeciesName=Saccharomyces cerevisiae"`) or by editing the properties (under Edit → Preferences → Properties...).

The automatic canonicalization of names can be turned off using the -P option (`cytoscape.sh -P canonicalizeName=false"`) or by editing the properties (under Edit → Preferences → Properties...). This canonicalization of names currently does not apply to expression data. Expression data should use the same names as the other data sources or use the canonical names as defined by the synonym table.