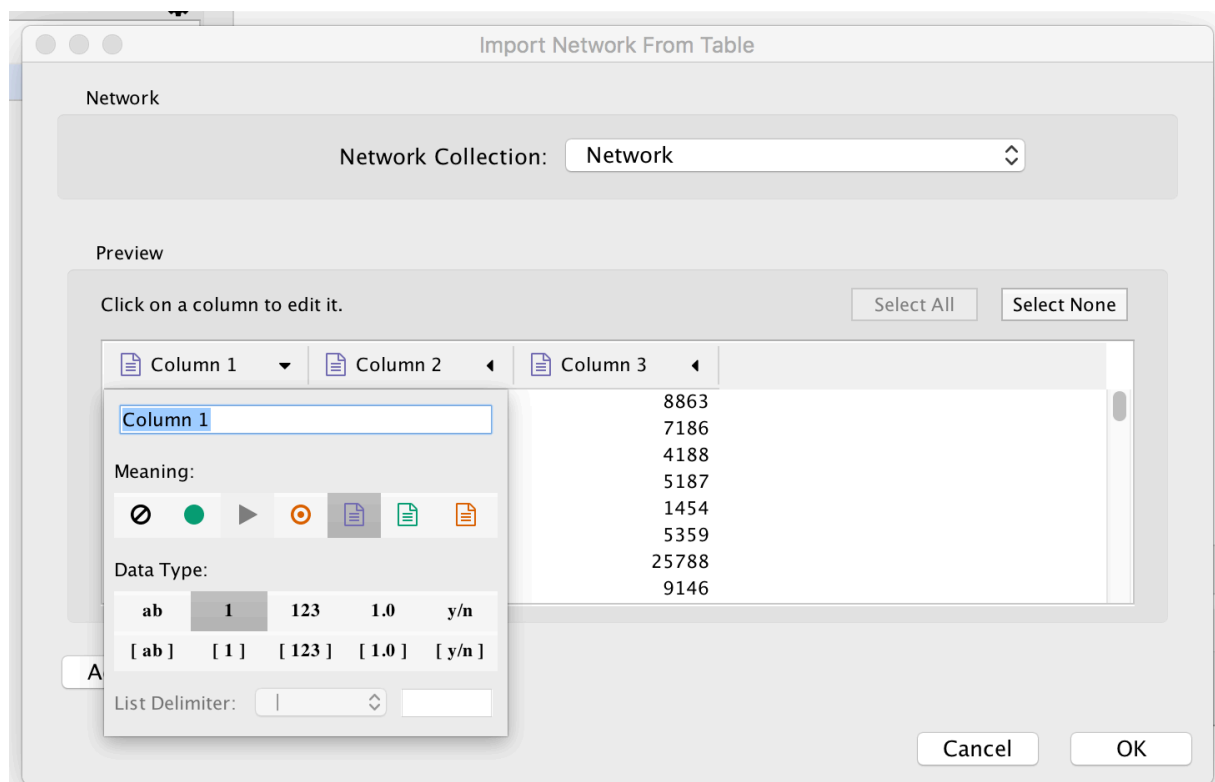


Cytoscape : Network topology and statistics

The purpose of the exercise is to:

1. Get familiar with Cytoscape
 2. Demonstrate how graph properties can be analyzed.
-

Part I. Getting started



Network import dialog

In this exercise, we will use a subset of the human interaction dataset by Rual et al. (Nature.2005 Oct 20;437(7062):1173-8).

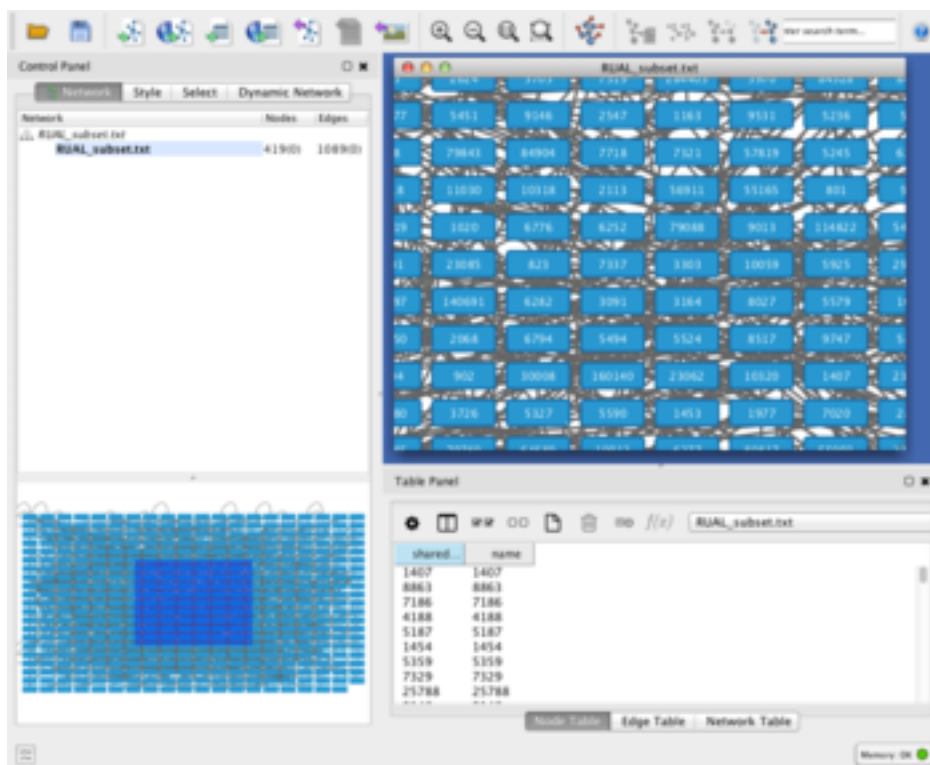
TASK: Download data files and check file names and content

- Download the following data files (please make a note as to which directory you have downloaded them to):
 - [RUAL_subset.txt](#)
 - [RUAL_subset_names.txt](#)

TASK: Load the network into Cytoscape

- **STEP 1:** Import the network "RUAL_subset.txt" into Cytoscape by selecting from the menu: **File -> Import -> Network -> File** then find and **Open** the file (maybe on your Desktop). You can also use "Start new session ... From Network File" from the start-up panel.
- **STEP 2:** Select Column 1 under **Source Node** (ex. 1407), Column 2 under **Interaction Type** (ex. Y2H), and Column 3 under **Target Node** (ex. 8863) so that the dialog looks like the figure at right.

This network consists of 1089 interactions observed between 419 human proteins, and is a small subset of a larger human interaction dataset. This subset consists of proteins that interact with the transcription factor TP53. Note that Cytoscape will only create an automatic view if there are less than 500 nodes.



The network before applying a layout

Part II. Network layout & Selecting nodes

TASK: Explore network layouts

- **STEP 3:** Try some of the different layouts (circular, organic, hierarchical and random) by selecting a method from the **yFiles Layouts** under **Layout**. Note: the Orthogonal layout can take a very long time to render, and to abort it you'll need to force quit Cytoscape and restart.

- By default, Cytoscape generates a grid layout which is efficient but not very useful for visualization. One of the most useful layouts for network biology is the spring layout (similar to the organic layout). Try the spring embedded layout: **Layout -> Perforce Force Directed Layout**

TASK: Explore proteins in network

- **STEP 4:** In the Cytoscape canvas (the blue window with the network view) you can select and move nodes by clicking and dragging with the left mouse button. Select a few nodes, and move them around the screen.
- **STEP 5:** The nodes in this network are labeled by numeric Entrez IDs, which is one of the IDs used by NCBI (www.ncbi.nlm.nih.gov). The node representing TP53 has Entrez ID 7157. Since these database IDs are not human readable, we typically import more meaningful node names to display or use in searches.
- **STEP 6: File -> Import -> Table -> File** and select "RUAL_subset_names.txt" that you have downloaded from above. This will add a "Symbol" node attribute. Any number of node attributes can be added in this way.

Select TP53 using the search box (far right of top menu bar) by entering either the node ID "7157" or Symbol "TP53" and <Return>, which should highlight TP53 (TP53 should now appear yellow in the network).

- You can unselect any selection by clicking on the canvas. You can also select nodes that interact with a specific node (e.g. TP53) by STEP 7.
- **STEP 7:** With TP53 selected, select the first neighbors of this node: **Select -> Nodes -> First Neighbors of Selected Nodes -> Undirected**. You should see a network with many highlighted yellow nodes.

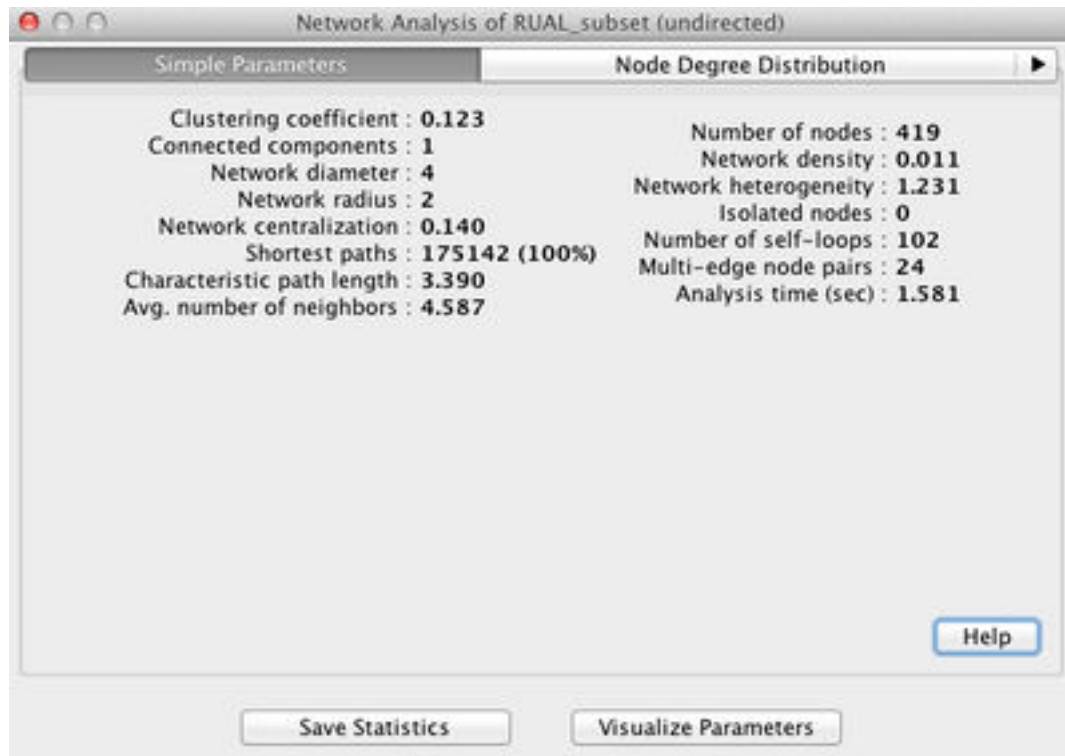


Question 1: How many proteins interact with TP53?

- Hint: The number of selected nodes are visible in the Network Control Panel in '()

Part III. Network statistics

TASK: Use the NetworkAnalyzer to calculate statistics on the network



Network statistics in Cytoscape

One of Cytoscape's strengths is the ability to write "Apps" (plugins) that can be run in Cytoscape. There is a large community of developers that have contribute with plugins (prior to v3.x.x) most of which are being ported to the current Cytoscape version.

- **STEP 8:** First, deselect all nodes in the network. Then apply the NetworkAnalyzer tool (under **Tools**) to the network by selecting **Network Analysis -> Analyze Network** (and "Treat the network as undirected"). This should produce a window with results somewhat similar to the windows below (depending on which operating system, the window may appear different from the illustrations):

As you may be able to see, the NetworkAnalyzer tool calculates various network parameters. Browse through the various network statistics/parameters and try to answer the following questions.



Question 2: What is the average node degree of the network? Hint: It is listed on the "Simple Parameters" tab of the plugin window.



Question 3: What is the most likely node degree of a randomly selected node in the network?

- Hint: Imagine that you are to select 1 node at random, and you must bet money on which node degree the selected node will have, what node degree would you bet on?



Question 4: Where is TP53 in the Node Degree Distribution plot?



Question 5: Use the node degree versus average cluster coefficient (Avg. Cluster Coefficient Distribution plot) to determine whether the network structure appears to be random, scale free or scale free and hierarchical?

Part IV. Network connectivity

TASK: Explore the connectivity of the network

- **STEP 9:** Have a look at the shortest path length distribution for the entire network using the NetworkAnalyzer.



Question 6: What is the highest number of edges that you need to connect any two nodes in the network? and is it smaller or larger than you thought?

This phenomenon is known as ‘small-world-network’ and can be found in many real life networks, e.g. the network that connects actors who have appeared in the same movie.

- **STEP 10:** You can connect any two actors on <http://oracleofbacon.org/>. Try, just for fun, with a few actors and see how many edges (movies) that are required to connect them.