

Worksheet 2

Objective of the Worksheet: In this exercise, you are to familiarize yourself with Cytoscape and its visualization abilities. Then you will input a network of PPI from a file and do a statistical analysis with a network analyzer plugin. You will also find the possible modules in the network with the MCODE plugin.

A) Running Cytoscape:

STEP 1: Launch Cytoscape 3.

STEP 2: Load the network RUAL_subset.sif into Cytoscape by selecting 'Import' under the 'File' menu, then selecting 'Network', and specifying the location of the file.

This network consists of 1089 interactions observed between 419 human proteins, and is a small subset of a large human interaction dataset. This subset of interactions consists of proteins that interact with the transcription factor protein TP53.

STEP 3: Try some of the different layouts (circular, organic, hierarchical and random) by selecting the appropriate layout in the 'yFiles Layouts' under the 'Layout'. One of the most useful layouts for network biology is the Organic, which is similar to the spring embedded layout. Try the organic layout.

STEP 4: In the Cytoscape canvas, you can select nodes by clicking on them with the mouse, or dragging with the left mouse button. You can zoom in the network canvas by clicking the magnifier icon or use your mouse wheel button.

The nodes in this network are labeled by numeric Entrez IDs, which are the IDs employed by NCBI (www.ncbi.nlm.nih.gov). The node representing TP53 is numbered 7157.

STEP 5: Select the TP53 node by entering the node id (7157) into the search text box (the text box beside the binocular icon) and hit return. The selected node will be highlighted by yellow. You can unselect any selection by clicking on the canvas.

STEP 6: Under the 'Select' menu, select 'Nodes' and 'First neighbors of selected nodes'. You should see a network with several yellow nodes in the center. This can be useful when looking for genes that have a common regulator or a common regulatory target.

Q1: How many proteins interact with TP53? Ans: Undirected 64, outgoing (22), ingoing (44).

B) Network Statistics: In this part of the exercise, you will work with the Network Analyzer plugin.

STEP 7: First, deselect all nodes in the network. Then apply the Network Analyzer plugin to the network by selecting 'Network Analyzer' from the 'plugins' menu (ignore the Cytoscape warning that the network contains both directed and undirected edges).

As you can see, the NetworkAnalyzer plugin calculates various network parameters that may be useful for describing the graphs you see. Browse through the various network statistics/parameters and try to answer the following questions:

Q2: What is the average degree (connectivity) of the network? Ans: 4.59.

Q3: What is the most likely degree of a random selected node in the network? Ans: 1.

Q4: Use the node degree distribution and the distribution of average cluster coefficient ($C(k)$) to determine whether the network structure is random, scale free or hierarchical?

C) Identification of Clusters (Complexes): As the average cluster coefficient (0.12) is relatively high it is to be expected that there will exist some clusters in the network. Next we will try to identify these using the MCODE algorithm.

STEP 8: Select 'MCODE' under the Apps menu and 'run MCODE on current network'. This will identify several complexes.

STEP 9: Try clicking on a complex (a complex in the 'MCODE Results summary'). This will highlight the complex (yellow nodes) in the network. You can zoom in the view by clicking the "Zoom selected region" icon. You can also examine the selected cluster in detail by clicking "New Network from Selection" icon. Try to look at the clusters with scores larger than 2. Have you found these clusters by manual inspection of the large network?

STEP 10: Have a look at the shortest path length distribution for the entire network using the Network Analyzer plugin.

Q5: What is the highest number of edges that you need to connect any two nodes in the network? Ans: 4.

This phenomenon is known as 'small-world-network' and can be found in many real life networks.