Network Biology

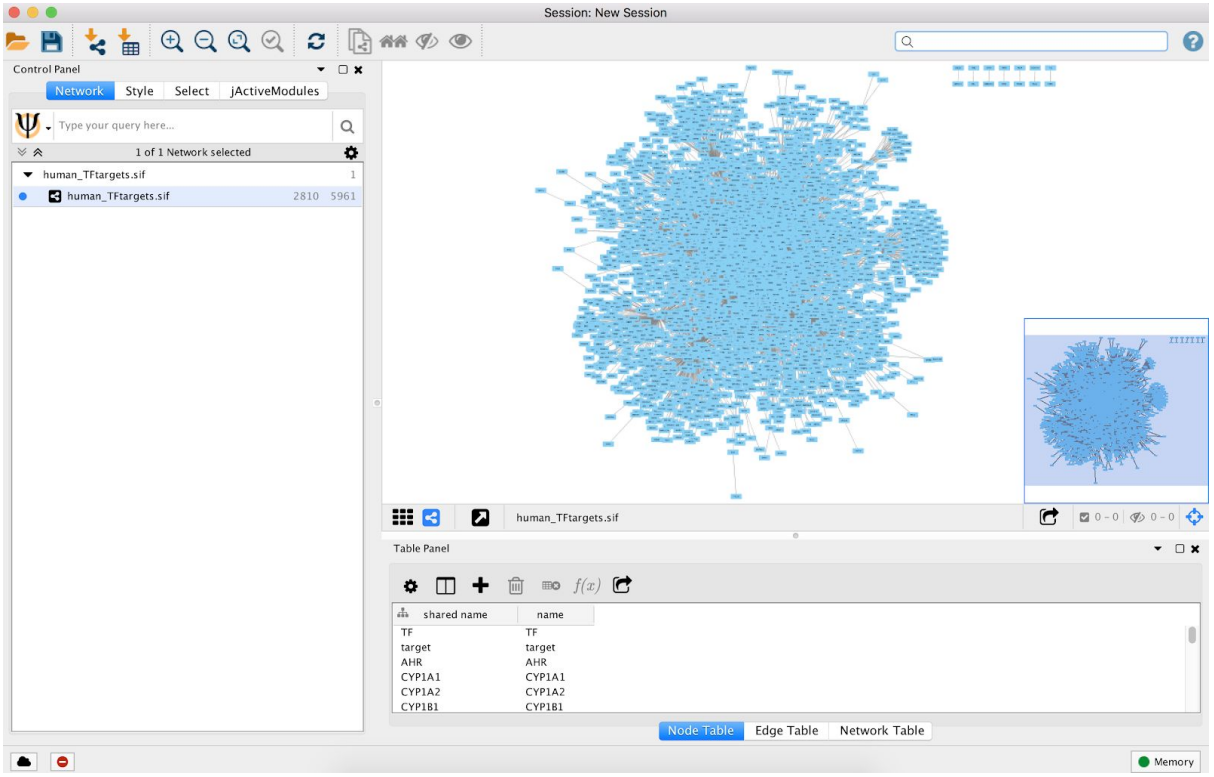
Practical session
12.03.2018

1. Getting started with Cytoscape

Cytoscape is a desktop open-source software for the visualization and analysis of networks. Is aimed to provide basic functionality to layout and query the network and to visually integrate the network with state data. The information below has been extracted from the Cytoscape tutorial (see section 4, Useful tutorials).

1.1. Overview and basic features

Cytoscape will look similar to the image below:

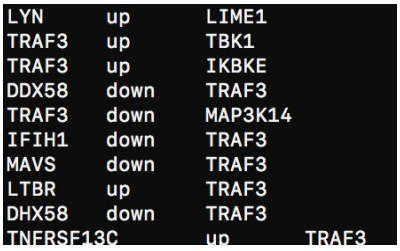


The main window has several components:

1. The Menu Bar at the top.
2. The Tool Bar, which contains icons for commonly used functions.
3. The Network Panel (in Control Panel). The network manager shows the networks that are loaded. Clicking on a network here will make that view active in the main window. Each network has a name and size (number of nodes and edges), which are shown in the network manager.
4. The main Network View Window, which displays the network.
5. The Table Panel (bottom right panel), which displays columns of selected nodes and edges and enables you to modify the values of column data.

1.2. Import and load networks

You can load a network from a local file you by selecting File → Import → Network → File. Network files can be specified in any of the formats described in the "Supported Network Formats" section in the Cytoscape tutorial (see section 4, Useful tutorials). Among these formats, there is the Simple Interactions File described in the Lecture and is the format that we will use in this practical session.



LYN	up	LIME1	
TRAF3	up	TBK1	
TRAF3	up	IKBKE	
DDX58	down	TRAF3	
TRAF3	down	MAP3K14	
IFIH1	down	TRAF3	
MAVS	down	TRAF3	
LTBR	up	TRAF3	
DHX58	down	TRAF3	
TNFRSF13C	up	TRAF3	

1.3. Node and Edge Data

Networks are most powerful for answering scientific questions when integrated with additional information. Cytoscape allows adding an arbitrary node, edge and network information as node/edge/network data columns (for example, annotation data on a gene or confidence values in a protein-protein interaction). These column data can then be visualized in a user-defined way by setting up a mapping from columns to network properties (colors, shapes, and so on).

1.4. Styles

Cytoscape allows the users to link any table data (name, type, degree, weight, expression data, etc.) as a property of the network (such as color, size of node, transparency, or font type). Style can be created or edited in the Style panel at the Control Panel.

1.5. Installing plugins with the App Manager

Cytoscape's capabilities can be expanded with apps. You can install apps within Cytoscape: go to the menu bar and choose Apps → App Manager.
Search and install the following modules: jActiveModules, BiNGO, MCODE.

2. Characterizing networks

In this section we will explore different strategies to characterize the structure, organization and function of a molecular interaction network.

2.1. Explore topological properties with NetworkAnalyzer

The way in which the nodes and edges are arranged within a network is called topology. Some topological properties apply to the network as a whole while others apply to individual nodes and edges. Description of the network-level topological properties can give us insights about the organization of the network and provide a framework for understanding biological function and evolution at the systems-level. Description of the node-level topological properties can give us information about the role and relevance of a molecule in the organization of the network.

Exercise 1. Load and study the topological properties of the 3 networks in the folder "exercise 1".

- Upload networks (Menu: File → Import → File). Note: Create a new collection for each network. Use the Control Panel to move from one network to another.
- Customize the visualization style (Left Control panel: Style)
- Change layout (Menu: Layout → *Prefuse Force Directed Layout* (prefered))
- Could you guess the *degree distribution* of each network by looking at the network display? Would you be able to identify which network follows a random distribution of the degree (e.g. such as the Erdös–Rényi (ER) model).
- Analyze Topology (Menu: Tools → NetworkAnalyzer → Network Analysis → Analyze Network (undirected)).
    - Explore network-level properties (Menu: View → Show Results Panel):
        - **Q-1.1:** What degree distribution follows every network? Which network model describes better each network?
        - **Q-1.2:** Which is the diameter, number of components and the average path length of every network?
    - Explore node-level properties (Table panel (bottom) → Node table):
        - **Q-1.3:** Which network would be more damaged (i.e. impacting a larger proportion of shortest paths) after removing the node 24?
- Save the session as "exercise 1" and close it.

2.2. Explore network function and modularity

An early study on the S. cerevisiae Protein-Protein Interaction (PPI) network revealed that proteins with shared functions were highly connected between them forming clusters, suggesting modularity [Schwikowski et al. 2000]. This property has been observed in Gene Regulatory Networks, co-expression network, and metabolic networks of several organisms.

Exercise 2. In this section we are going to analyse 2 collections of human PPIs (*human_ppis_t1_c.sif* and *human_ppis_t2_ph.sif*) that have been identified by 2 new experimental techniques. Unfortunately, these techniques are able to identify only specific types of PPIs. Here, we are going to use GO enrichment approaches (via BiNGO app) to characterize the coverage and bias in these datasets. Also, we are going to study their organization and modularity (via MCODE).

- Load and explore the networks (Tip: go to View → Show Graphic Details).
- **Q-2.1.** Which % of human proteins are covered in the PPIs dataset? Do you think that any of the techniques can be considered a universal detection method for PPIs in humans? Explain why.
- **Q-2.2.** Describe the network topology. What differences can you tell between both networks?
- Now let's focus on the network *human_ppis_t2_ph.sif*. Select the proteins with degree → 9 (Nodes Table → sort according to the column Degree → select proteins → click right button → Select nodes from selected rows). Now, let's investigate the biological functions enriched among these proteins with BiNGO (Apps → BiNGO). BiNGO is an analysis tool to determine which Gene Ontology (GO) categories are statistically over- or underrepresented in a set of genes or a subgraph of a biological network. NOTE: Select GO Molecular Function, *Homo Sapiens* as organism and customize the reference list of proteins with the "human_protein_census.txt" file. Save settings as default.
- **Q-2.3.** Explore BiNGO output. What can you say about these highly connected proteins? Repeat the analysis using selecting all the proteins in this network. What type of bias do you observe in this network? Can you guess what the type of PPIs is represented here?
- Select all the proteins in the *human_ppis_t1_c.sif* network. Now let's run 2 GO enrichment analysis: one to study overrepresented and another to study underrepresented GO Molecular Function terms.
- **Q-2.4.** Explore BiNGO GO networks. Use the Layout → Node Layout Tools to improve the visualization. What type of bias do you observe in this network? Can you guess what the type of PPIs are represented here using the information that you have regarding the functional enrichment and the organization of the network?
- Now let's explore the modularity of the *human_ppis_t1_c.sif* network. Open MCODE (Apps → MCODE) and analyze the modules/clusters of the "Whole Network".
- **Q-2.5.** Select the nodes in the first module and run an enrichment analysis with BiNGO. Repeat the same process for the 2nd and 3rd modules. Can you identify any relationship between the modules and protein functions?
- **Q-2.6.** Now, if you are asked to investigate a disease that is expected to be related to dysfunctional ribosomes, which of the two networks would you use as a biological model?
- Save the session as "exercise 2" and close it.

3. Network-based characterization of Differential Gene Expression data

Combining expression signatures with molecular networks (such as PPIs, Signalling Pathways or Gene Regulatory Networks), can help us to explain the molecular mechanisms underlying the phenotype under study. For example we can identify signalling cascade or transcription factors known to regulate the altered genes. The major benefit of integrating expression and network data is that the identified mechanisms (hypothesis) are supported by more than one data types and, therefore, are less likely to contain false positives.

Exercise 3. Imagine that you are conducting a research on Medulloblastoma (MB), a common malignant brain tumor in children, where MYC overexpression or amplification has been associated with poor clinical outcome. MYC is a well known transcriptional regulator and you would like to characterize the molecular alterations associated with aberrant MYC activity to design a therapeutic strategy. For this purpose, you conducted an *in vitro* experiment where you overexpress MYC in an MB-derived cell line and compared the gene expression profiles before and after the perturbation. Here, you are going to use two types of human biological networks (gene regulatory network: *human_TFtargets.sif* and signalling network: *human_signalling.sif*) to characterize the results from the differential expression analysis (file: *MYC_GSE22139_differential_expression.csv*). NOTE: The expression data has been derived from Giulio et al 2011 and downloaded from GEO (accession GSE22139).

- Open the expression file *MYC_GSE22139_differential_expression.csv* with Excel and check the content. It contains 5 columns:
    1. Gene.
    2. Expression fold change between controls and MYC overexpressing cells (positive values indicate overexpression; negative values indicate underexpression).
    3. 2-tailed adjusted p-value.
    4. Adjusted p-value 1 (1-tailed; from testing gene overexpression).
    5. Adjusted p-value 2 (1-tailed, from testing gene underexpression).
- Load the provided networks as independent collections and explore them.
- For each network, load the expression file as nodes-related data (File → Import → Table → File → *MYC_GSE22139_differential_expression.csv*). Select Import data as Node Table Columns and be sure that the KEY column are the gene names.
- Configure the nodes color (Control Panel → Style → Nodes → Fill Color) so that the color is proportional to the expression fold change (node data column). Apply continuous mapping.
- Now we are going to extract, from the signalling network, the subnetworks (or modules) associated with our overexpressed genes. Open jActiveModules (Apps → jActiveModules). Select the attribute *adj.pvalue-over* (0 indicates high confidence for

gene overexpression, while 1 indicates the opposite) and the signalling network as target network. Click on the search button. In the Control Panel → Networks, you will find the top 5 jActiveModules associated with the genes with higher overexpression (tip: change style to the default to see the node color as a function of the expression fold change). Use BinGO to functionally characterize the top 2 (tip: use the file human_signalling_proteinlist.txt as reference set and Biological Process as GO terms). Repeat the jActiveModules analysis using the *adj.pvalue-under* (0 indicates high confidence for gene underexpression, while 1 indicates the opposite).

- **Q-3.1.** What signalling processes is MYC activating? And inhibiting?
- Use NetworkAnalyzer to calculate the topological parameters of the identified module 1.
- **Q-3.2.** Which proteins seem more relevant for the connectivity of module 1?
- Repeat the same process using the gene regulatory network: *human_TFtargets.sif*.
- **Q-3.2.** What TFs seem upregulated in MYC-overexpressing cells? And downregulated?
- Save the session as “exercise 3” and close it.

4. Useful tutorials

With the information described in this document, you should be able to perform all the requested tasks. However, if you have any question and would like to learn more about the functionalities of Cytoscapes and the Plugins, you can check the following links:

- Cytoscape User Manual: http://manual.cytoscape.org/en/stable/Introduction.html
- Cytoscape Nature Protocols: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3685583
- BiNGO: https://www.psb.ugent.be/cbd/papers/BiNGO/User_Guide.html
- jActiveModules: https://github.com/idekerlab/jActiveModules/wiki/Help
- MCODE: http://opentutorials.cgl.ucsf.edu/index.php/Tutorial:MCODE_3