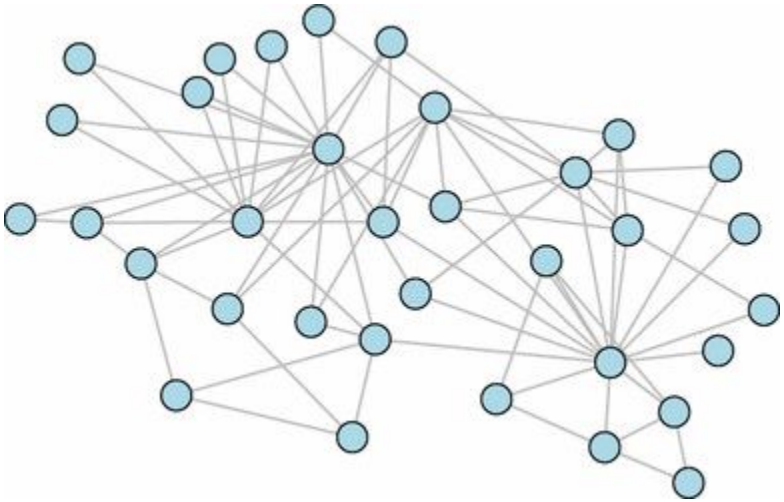
Network biology

Practical session 12.03.2018
Part II BBS Bioinformatics

Lecture summary

Network definition

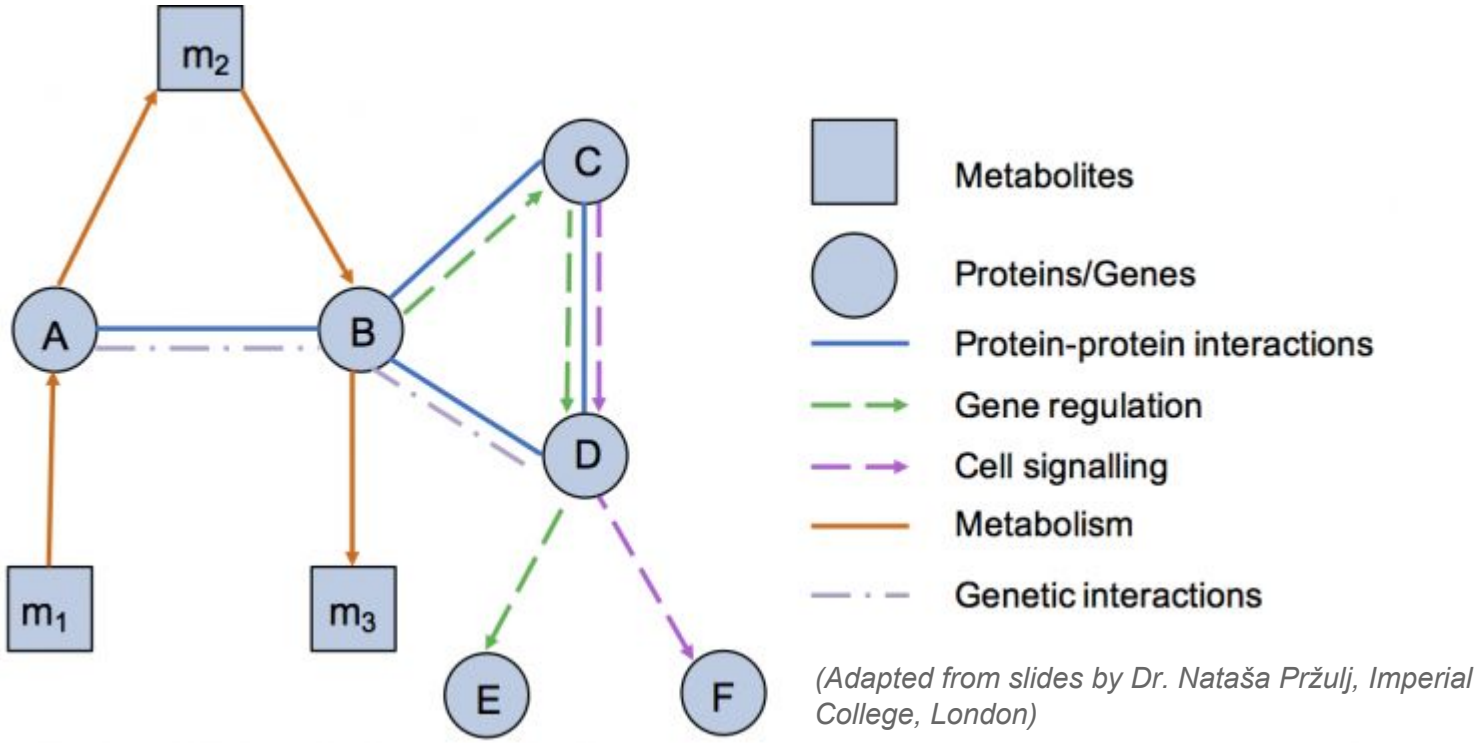
1. A system of interacting elements.
2. A map of interactions or relationships.



Can be represented mathematically as a **graph**, where vertices denote the **nodes** and **edges** (or links) denote the interactions.

Cell = network of networks

Different types of information can be represented as networks in order to model the cell.

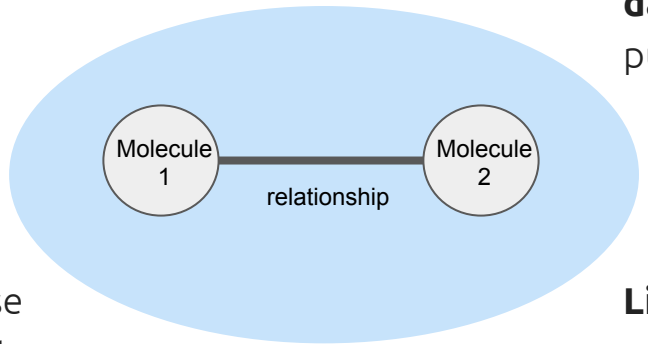


(Adapted from slides by Dr. Nataša Pržulj, Imperial College, London)

Sources of molecular interactions

Expert curation that evaluate existing published evidence.

Experimental: High-throughput datasets (e.g. PPIs via affinity purification + mass spectrometry).



Computational predictions, use existing experimental evidence to predict unexplored relationships

Literature text-mining, computationally extract molecular relationships from the published literature.

All of them characterized by different levels of

Coverage, bias & noise

Applications of network analysis in biology

1. Characterization of molecular networks at the systems level
    - Identify emergent patterns (structure, organization, relevant sub-networks) that can not be observed if the individual molecules were examined individually

2. Use of biological networks as a comprehensive scaffold of molecular interactions
    - To functionally characterize experimental data (high-throughput)
    - To make phenotypical predictions

3. Prediction of new interactions and networks

Applications of network analysis in biology

1. Characterization of molecular networks at the systems level
   - Identify emergent patterns (structure, organization, relevant sub-networks) that can not be observed if the individual molecules were examined individually

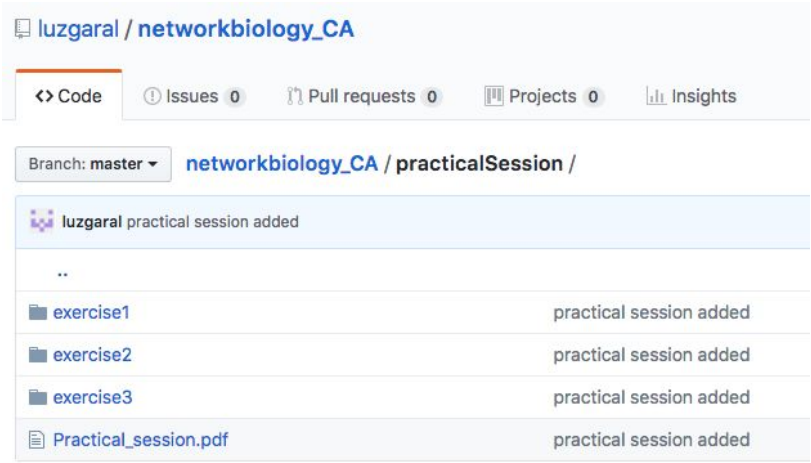
2. Use of biological networks as a comprehensive scaffold of molecular interactions
   - To functionally characterize experimental data (high-throughput)
   - To make phenotypical predictions

3. Prediction of new interactions and networks

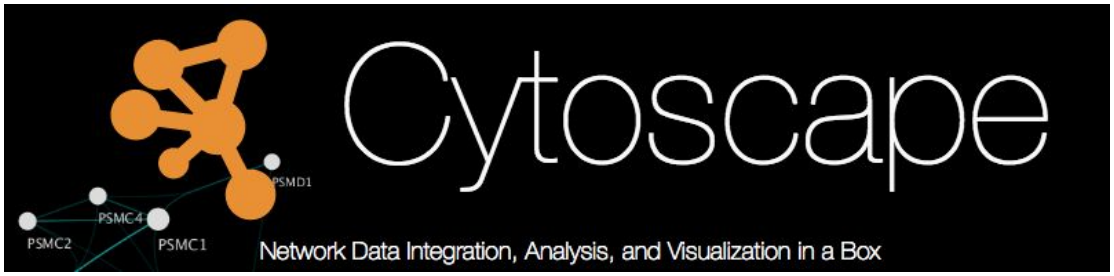
Hands-on

Practical session

DATA: https://github.com/luzgaral/networkbiology_CA



SOFTWARE: Cytoscape 3 http://www.cytoscape.org/



Cytoscape

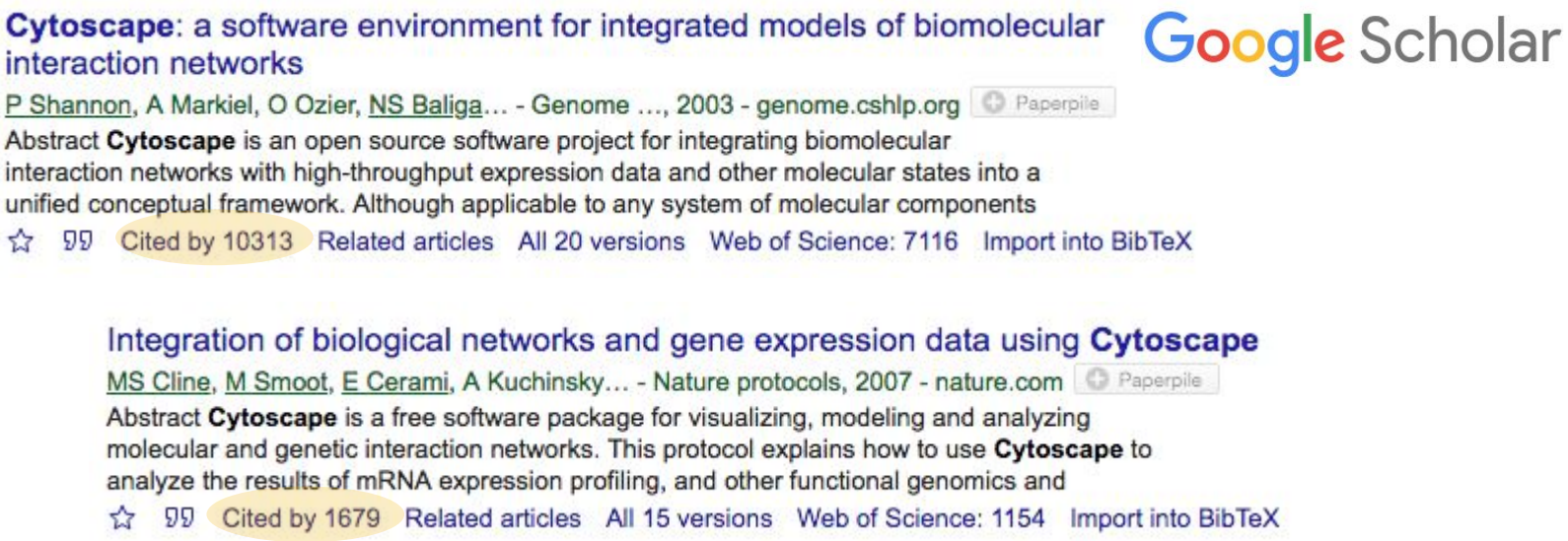
www.cytoscape.org

Software for Visualization, Integration & Analysis of Molecular Networks.

Free & open source software application (LGPL license)

> 10,000s users, 3000 downloads/month

Extensible through plugins/apps developed by third parties

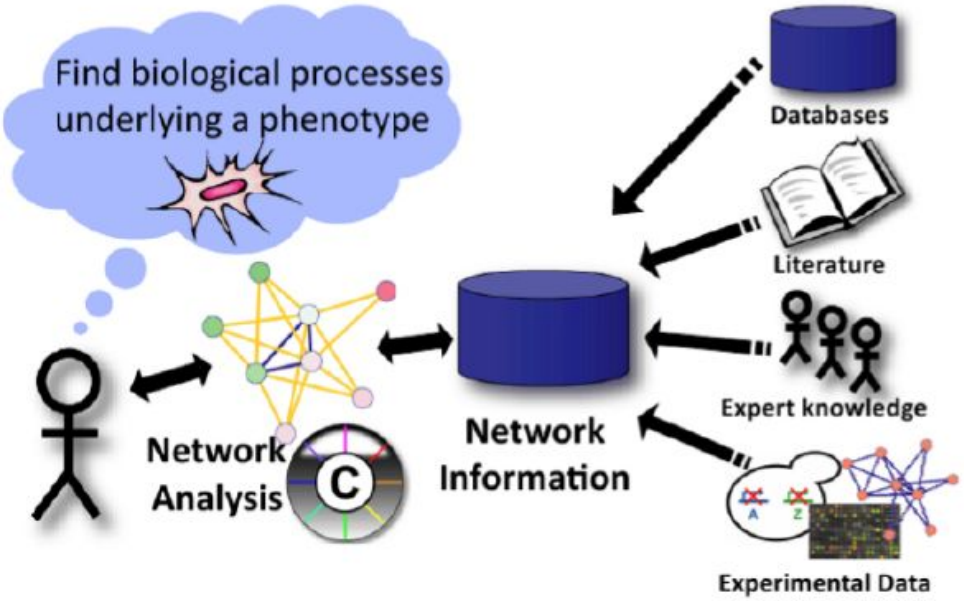


Google Scholar

Cytoscape

A Typical Cytoscape Workflow

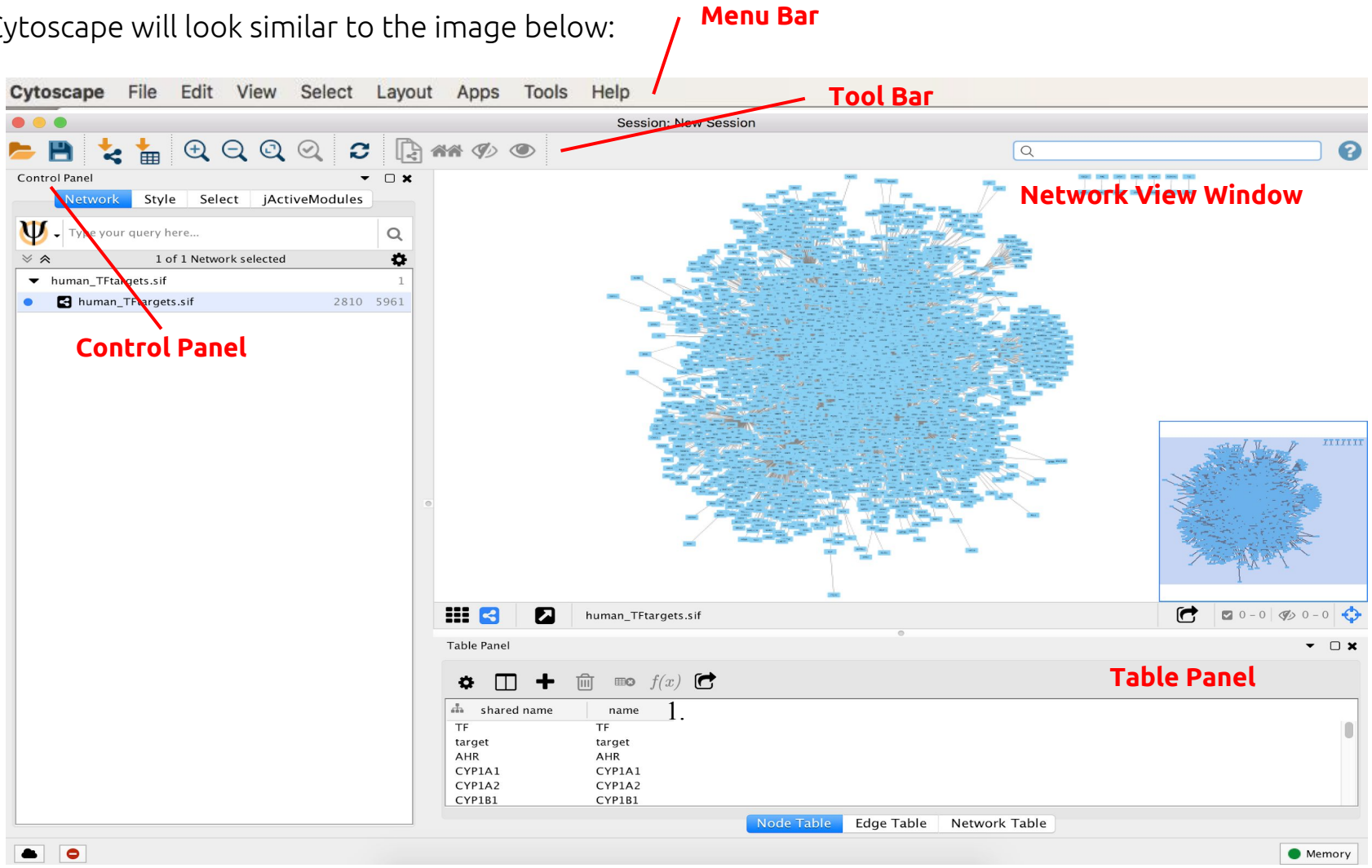
1. Load Networks (Import network data into Cytoscape)
2. Load node/edges related data or attributes (optional)
3. Analyse and Visualise Networks
4. Prepare for Publication



Cytoscape

Panels overview

Cytoscape will look similar to the image below:

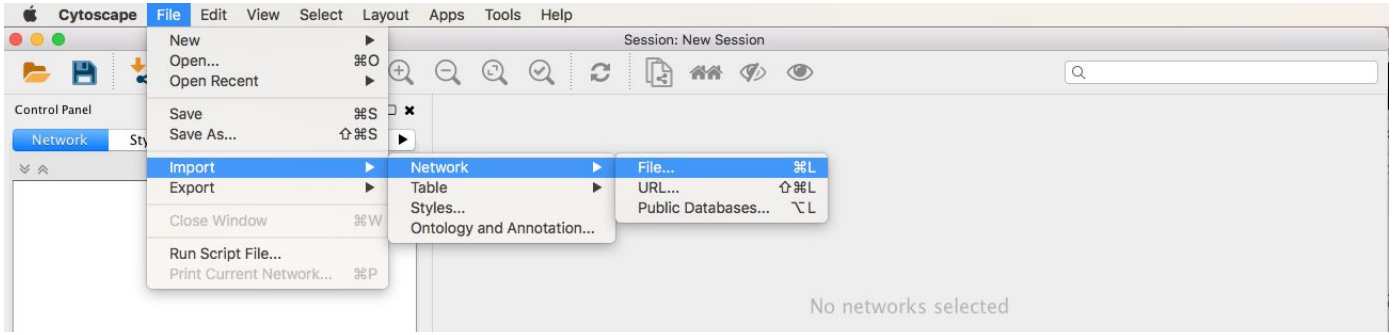


Cytoscape
Loading networks

Two ways:

- Use import network from File:
  - Excel file, comma or tab delimited text
  - Simple interaction file
  - Other network formats
- Use import network from Web Services
  - IntAct (EMBL EBI)
  - Reactome
  - Pathway commons (collection of sources)
  - NCBI Entrez EUtilities

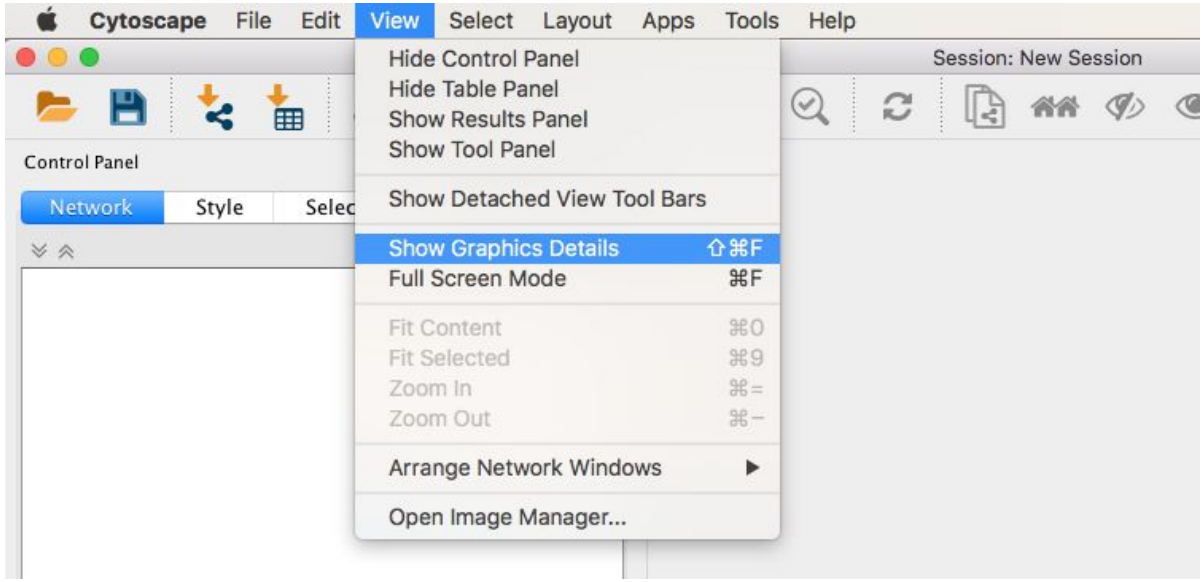
```
In this practical session we are going to use the networks provided in the github repo as sif files
```



Cytoscape
Loading networks

Tips & Tricks: Network View

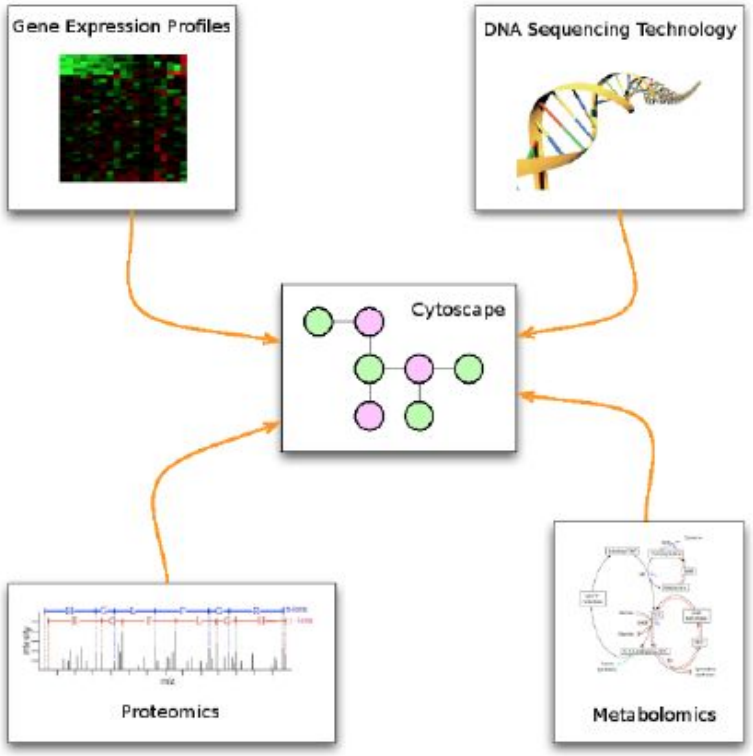
- When you open a large network, you will not get a view by default
- To see what things will look like at full detail: View → Show Graphics Details



Cytoscape

Loading node/edge associated data (attributes)

Nodes and edges can have data/attributes associated with them:



Cytoscape supports multiple data types: Numbers, Text, Logical, Lists.

Upload data (formats):

- excel file
- tab delimited text
- expression matrix

Import data from web services

- NCBI Entrez Gene
- Ensembl Biomart

Create attributes from the node/edge data panel

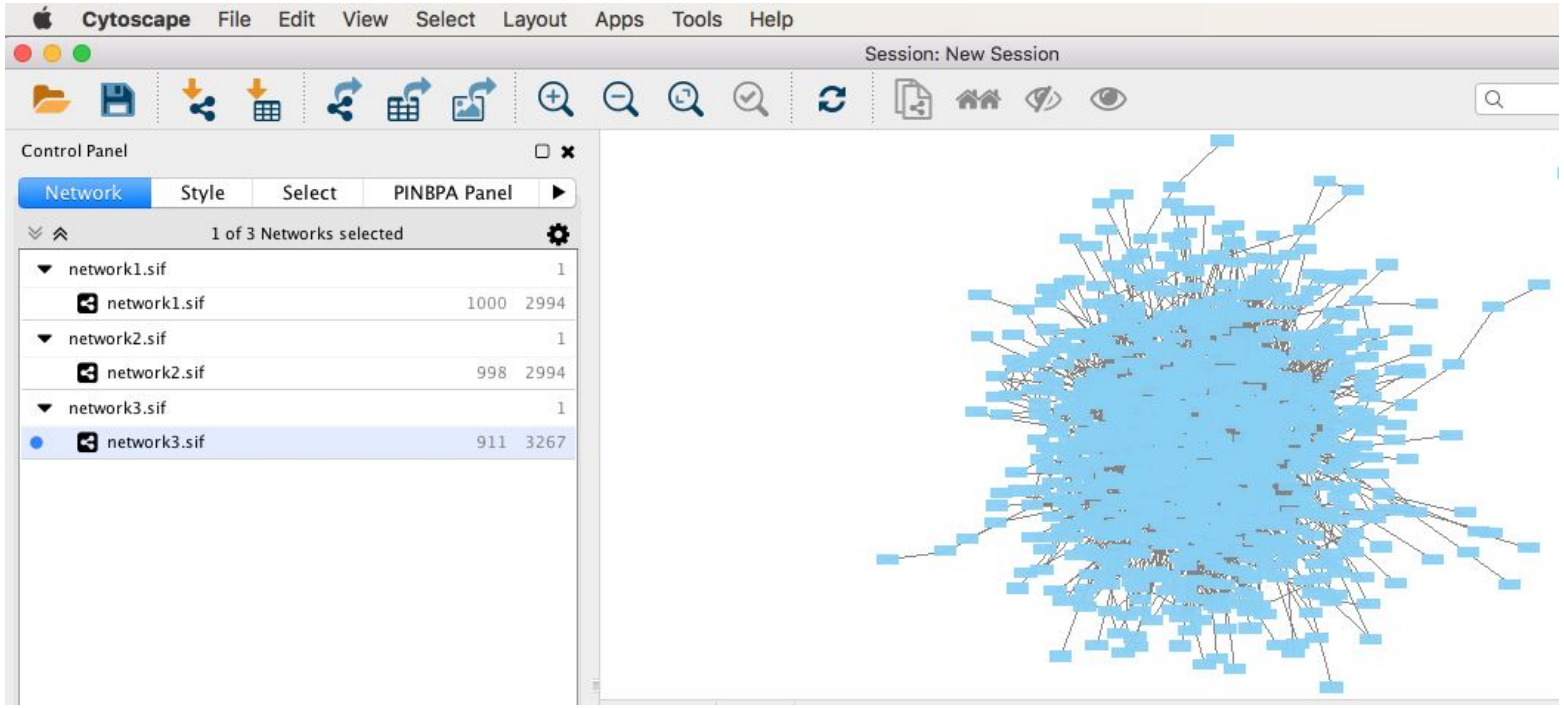
In this practical session we are going to use the node attributes provided in the github repo as **.csv** files

Cytoscape

Loading node/edge associated data (attributes)

Tips & Tricks

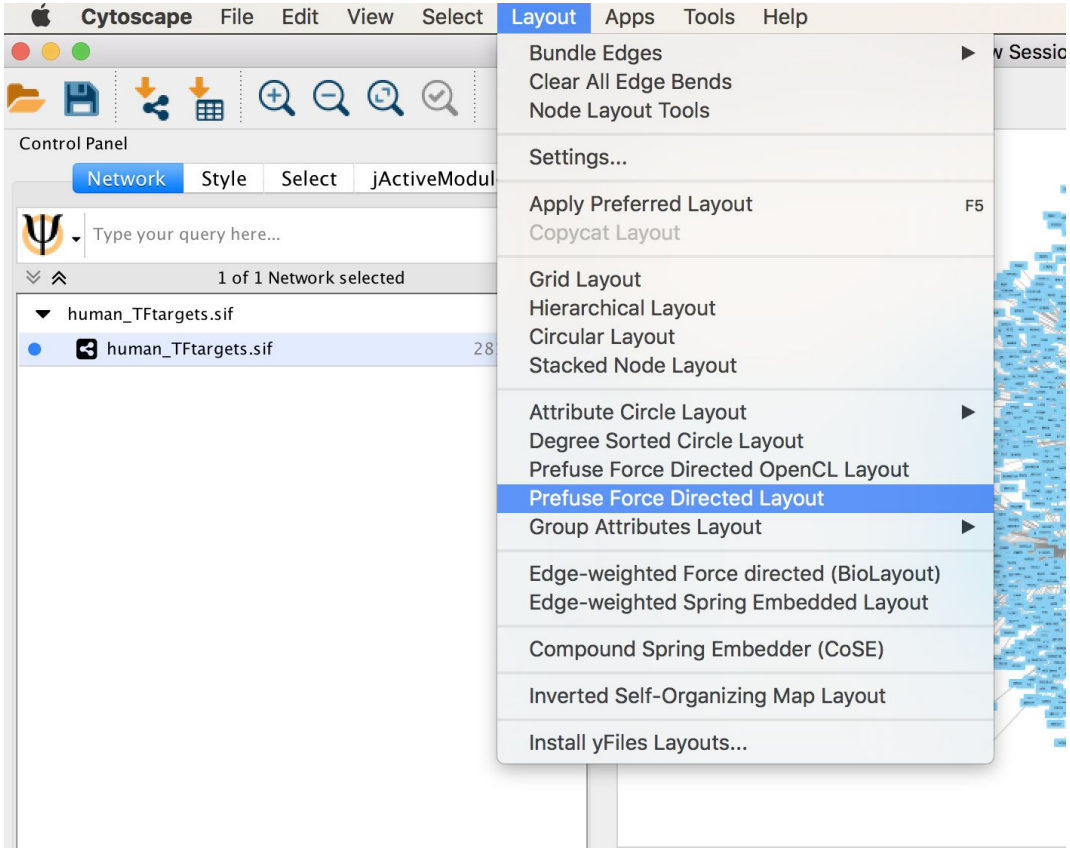
- You can have multiple independent graphs in Cytoscape
- When you upload new node-related data, be sure to select the correct network

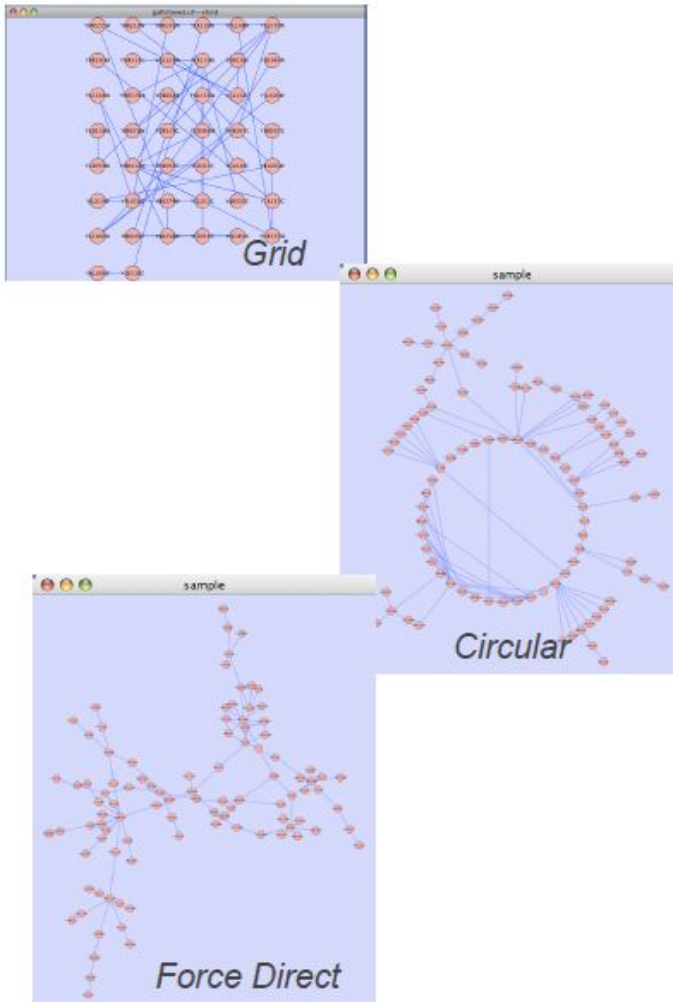


Cytoscape

Visualize Networks

Layouts several algorithms available

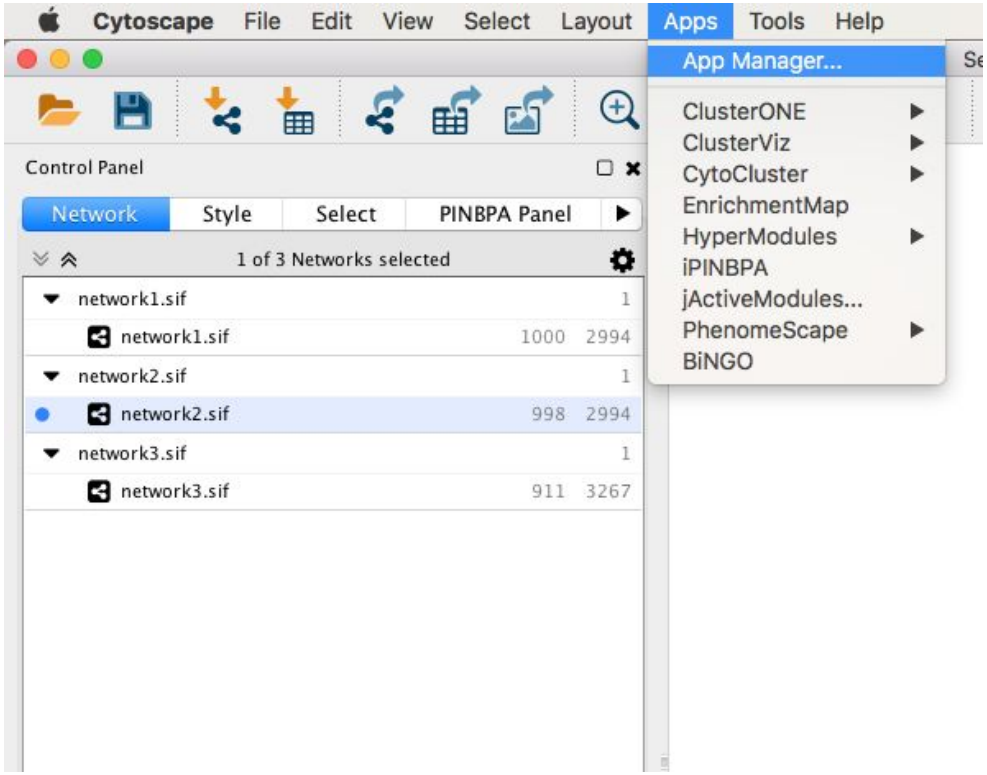




Cytoscape
Extended Functionality

Cytoscape extends its functionality with Apps:

- Developed by third parties
- Listed at http://apps.cytoscape.org/
- Major categories:
  - Online databases import
  - Graph Analysis
  - Genomics Integrated Analysis
  - Clustering
  - Visualization
  - Network Generation
- Usually available through
  - Apps Manager in the toolbar
  - Download from the Apps websites




```
For this practical session we are going to install the modules: jActiveModules, BiNGO, MCODE.

Install them via App Manager.
```

Applications of network analysis in biology

1. Characterization of molecular networks at the systems level
   - Identify emergent patterns (structure, organization, relevant sub-networks) that can not be observed if the individual molecules were examined individually

2. Use of biological networks as a comprehensive scaffold of molecular interactions
   - To functionally characterize experimental data (high-throughput)
   - To make phenotypical predictions

3. Prediction of new interactions and networks

Characterization of molecular networks at the systems level

Looking at different network properties can provide valuable information about:

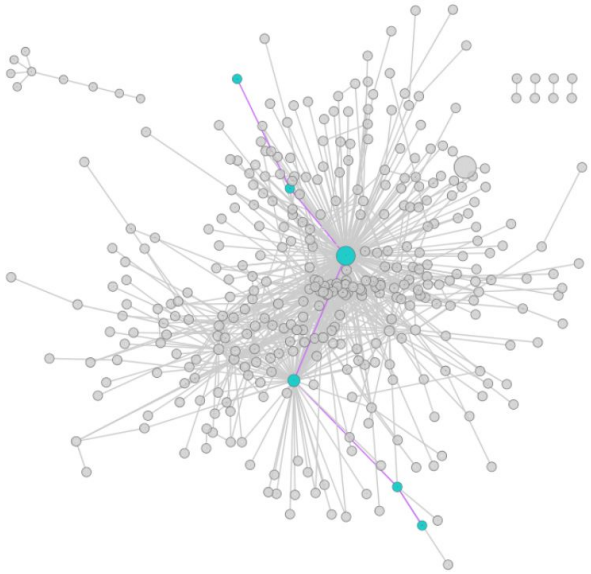
- the internal organization of a biological network (topological properties)

Characterization of molecular networks at the systems level

Looking at different network properties can provide valuable information about:

- the internal organization of a biological network (topological properties)

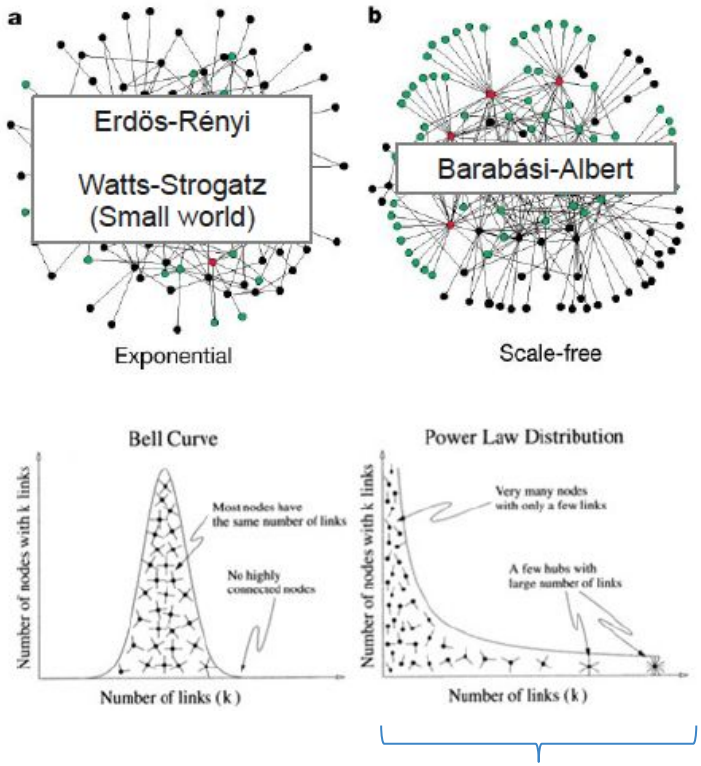
Topology Parameters	
Global	**Local**
- **Components**: Set of nodes connected between them but disconnected to the other nodes. - **Shortest Path**: The path with the smallest number of edges between two nodes. - The largest shortest path is the network **diameter**. - **Degree Distribution**: Counts the number of nodes with degree k and gives the probability that a selected node has exactly k edges. The degree distribution allows us to distinguish between different classes of networks.	- **Degree**: the number of direct interactors. Hub is a node with an exceptional high degree. - **Betweenness**: the frequency with which a node lies on the shortest path between other proteins. - **Closeness**: the average distance to all other proteins. - **Clustering Coefficient**: the degree to which nodes in a graph tend to cluster together.



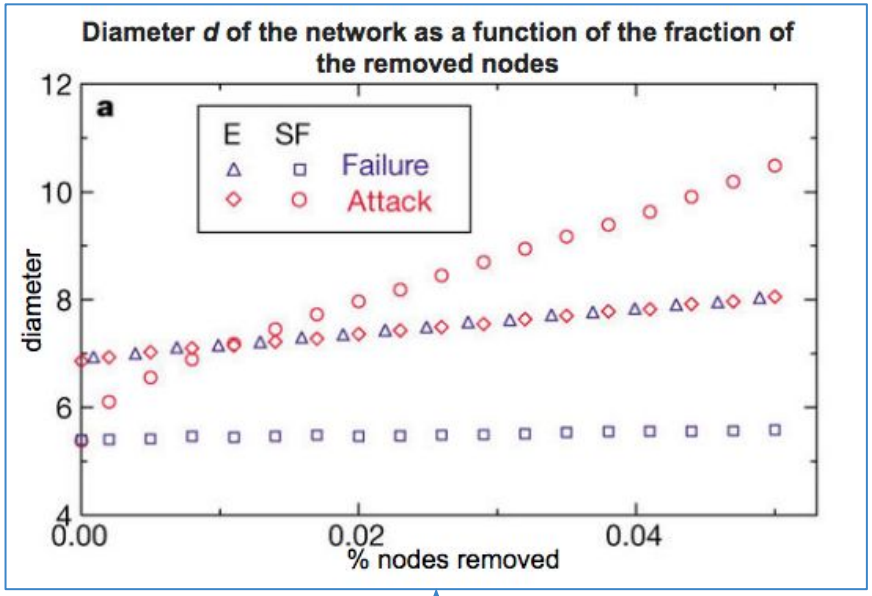
Characterization of molecular networks at the systems level

Looking at different network properties can provide valuable information about:

- the internal organization of a biological network (topological properties)
- the evolutionary constraints that have shaped an organism's molecular network



Distribution observed in most biological networks



High degree of **robustness** against accidental node failures

Characterization of molecular networks at the systems level

Looking at different network properties can provide valuable information about:

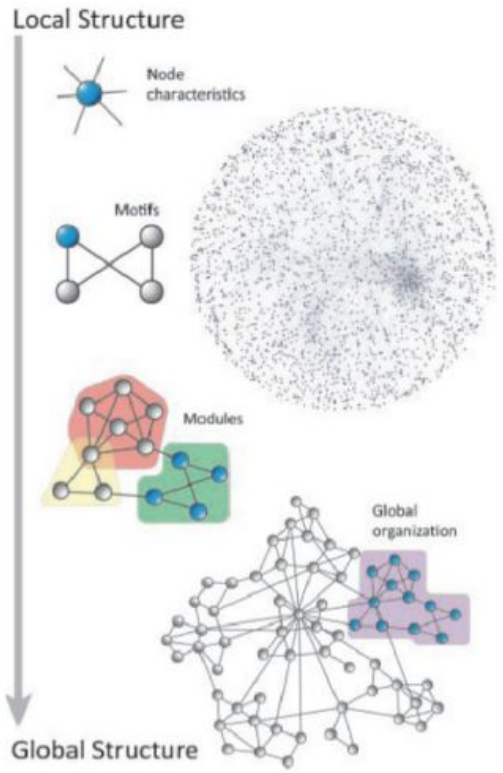
- the internal organization of a biological network (topological properties)
- the evolutionary constraints that have shaped an organism's molecular network
- the distribution of molecules among biological processes (modular properties)

Cellular functions are likely to be carried out in a highly modular manner.

The levels of organization of complex networks are:

- **Hierarchy** describes how the various structural elements are combined
- Groups of highly interconnected nodes are called **modules** (or communities)
- Each module can reduced to a set of triangles (**motifs**). A high degree of triangles is reflected by the clustering coefficient (the signature of a network potential modularity)

A hierarchical architecture implies that sparsely connected nodes are part of highly clustered areas, with communication between the different highly clustered neighbourhoods being maintained by a few highly connected nodes



Characterization of molecular networks at the systems level

Looking at different network properties can provide valuable information about:

- the internal organization of a biological network (topological properties)
- the evolutionary constraints that have shaped an organism's molecular network
- the distribution of molecules among biological processes (modular properties)
- detect coverage and bias issues

Exercise 1

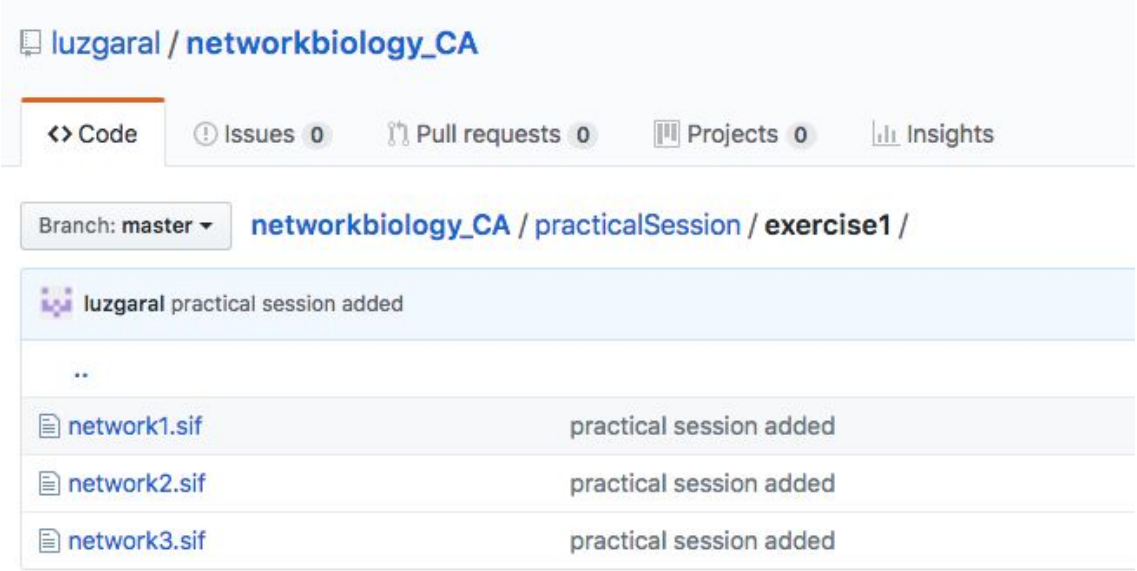
Characterizing networks topology

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.

Exercise 1

Characterizing networks topology



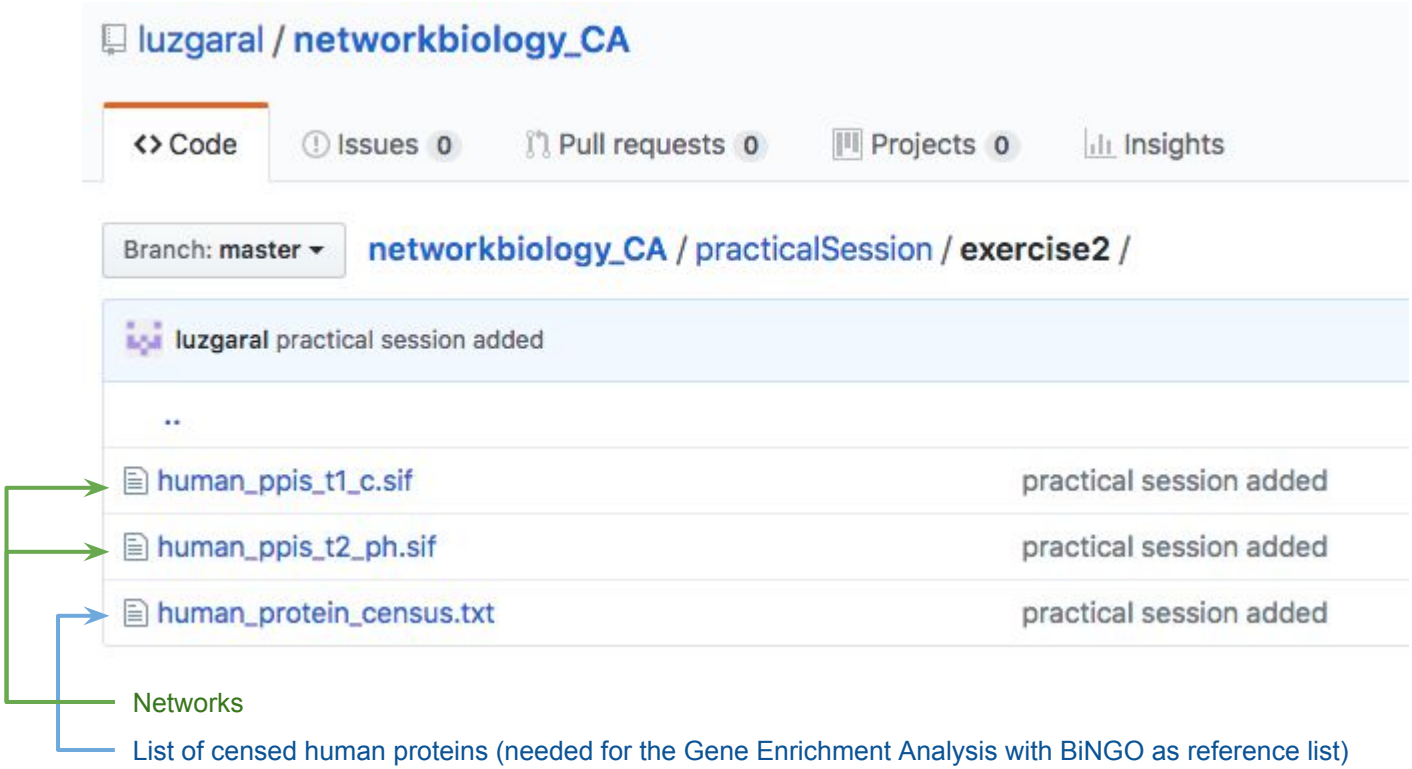
Exercise 2

Characterizing function and modularity

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).

Exercise 2

Characterizing function and modularity



Applications of network analysis in biology

1. Characterization of molecular networks at the systems level
   - Identify emergent patterns (structure, organization, relevant sub-networks) that can not be observed if the individual molecules were examined individually

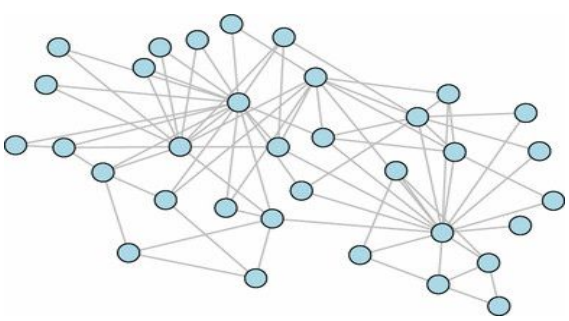
2. Use of biological networks as a comprehensive scaffold of molecular interactions
   - To functionally characterize experimental data (high-throughput)
   - To make phenotypical predictions

3. Prediction of new interactions and networks

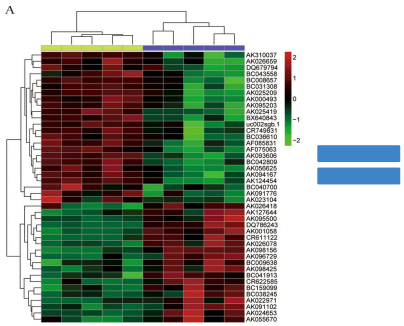
Functional characterization of large-scale experimental data

Common examples:

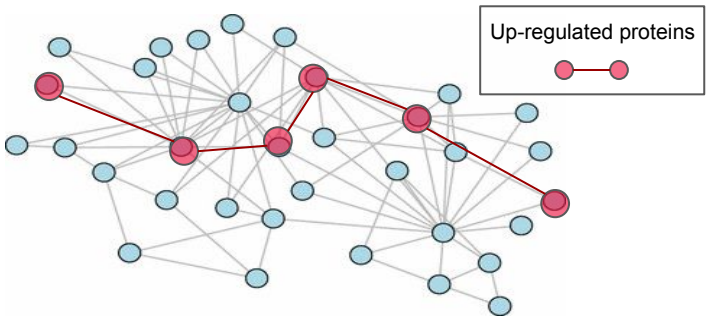
1. Identification of signalling pathways with differential gene expression upon drug treatment



Protein-protein interactions involved in cell-cycle

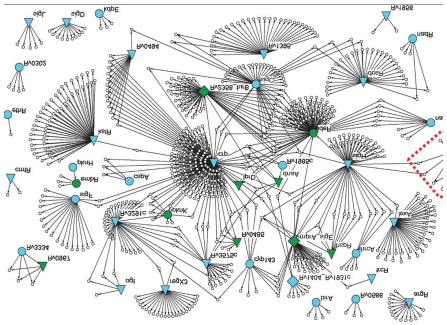


Differential gene expression after drug treatment

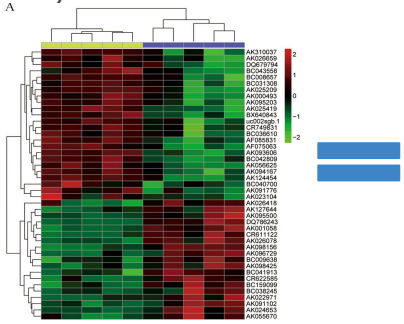


Molecular insights into the mechanism of action of the drug

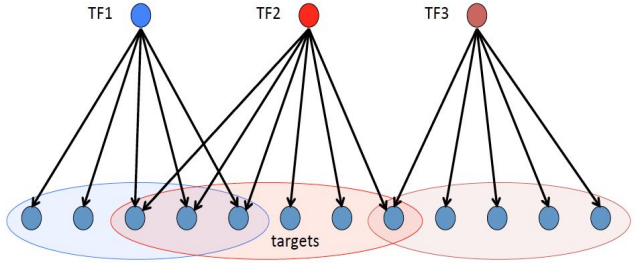
2. Identification of the Gene Regulatory Network involved in cell differentiation



TF-gene regulatory network



Differential gene expression between 2 differentiation stages



Identification of the TFs responsible of the differentiation response

Exercise 3

Network-based characterization of Differential Gene Expression data

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).

Exercise 3

Network-based characterization of Differential Gene Expression data