Network analysis using Cytoscape

Dr. Hadar Benyamini
The Ein Kerem Bioinformatics unit
The Hebrew University

Outline
• Introduction to (biological) networks
  • Graph analysis
  • Properties of biological networks
  • PPI detection methods
• Cytoscape basics
  • Sources for PPI
  • Import & export
  • Selection and styling
  • Data Integration
  • Enrichments
• Hands-on examples

Widely studied networks
• Biological networks
• The internet
• Social networks

Molecular networks: most frequent application of Cytoscape

(Biological) networks
• Network: a system with sub-units that are linked into a whole
• Biological network: any network that applies to biological systems
• A mathematical representation of connections found in ecological/ evolutionary/ physiological studies

Graph: a structure of related objects
• $G = (V, E)$
• A set of $V$ vertices/ nodes, with a set of $E$ edges /arcs/ lines
• An edge is associated with two vertices
• Edges: directed, undirected, weighted
Network Representations

**List**

<table>
<thead>
<tr>
<th>Protein1</th>
<th>Type</th>
<th>Protein2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProteinA</td>
<td>pp</td>
<td>ProteinB</td>
</tr>
<tr>
<td>ProteinA</td>
<td>pp</td>
<td>ProteinC</td>
</tr>
<tr>
<td>ProteinB</td>
<td>pp</td>
<td>ProteinD</td>
</tr>
<tr>
<td>ProteinC</td>
<td>pd</td>
<td>ProteinD</td>
</tr>
</tbody>
</table>

**Matrix (rows, columns)**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Graph theory: some definitions

- **Degree**
- **Shortest path**
- **Centrality**

Molecular networks

- **Metabolic networks**
- **Signalling networks**
- **Gene regulatory networks**

Protein-protein interaction networks

- Protein-protein interactions: physical, selective, biological context dependent
- Interactome: All PPIs of an organism/organ/biological condition

Rapid Accumulation of PPI Data

- **Human physical PPI**
- **Mouse physical PPI**

Physical, non-redundant protein-protein interactions

Taken from BioGrid stats, Feb. 2010 to Feb. 2018

Information quantity - BioGrid

- [Human PPI stats](#)
- [Mouse PPI stats](#)
PPI networks properties

Small world effect
Scale free networks
High transitivity

Communities/ clusters/ modules

Respiratory electron transport chain
Lipid homeostasis

PPI detection methods

Low throughput, small scale, high resolution

High throughput, large scale, low resolution

197,019
Affinity Capture-MS

57,106
Affinity Capture-Western

51,225
Two-Hybrid

39,688
Co-fractionation

29,827
Reconstituted Complex

10,483
Biochemical Activity

Human PPI – BioGrid database

Building and analyzing PPINs

PPI retrieval

Topological analysis
Clustering
Data integration
Annotation enrichment

Insights and visualizations

The Era of Network Biology

• Parts list
• Genomics

• Blueprint
• Network Medicine

• Diagnostic tools
• Metabolomics, proteomics

• Replace Parts
• Gene Therapy

A. Barabasi, TEDMED 2012
Added value: **Network Environment**

- Visualization
- Relationship between genes and gene sets
- Graph analysis tools
- Data integration
- Open-source, apps

**Basic Cytoscape Workflow**

- Data Integration
- Expression
- Network
- Annotation
- Data Analysis
- Enrichments
- Graph analysis
- Visualization

**Outline**

- Retrieving networks
- Selection and filtering
- Layouts, clustering
- Data integration
- Styles
- Enrichments

Can be done outside Cytoscape

**Initial window**

- Toolbar: with command buttons, name is shown when mouse hovers
- Control Panel: Network, style, select
- Network Panel: opens with links to demos
- Table Panel: node, edge, and network data

**Building and analyzing PPINs**

- PPI retrieval
- Experimental (in house)
- Databases (Primary & meta-)
- Databases (Predictive)
- Topological analysis
- Clustering
- Data integration
- Annotation enrichment
- Insights and visualizations
Selection/ filtering

- By column, degree, or topology

Select List From interactome (Mouse Network)

Selection of Up-regulated

Layouts

Network Analyzer+ Node Size by Degree
General tips

• For heavy applications can be slow and memory consuming (yet always improving)
• Save sessions and views
• Sometimes there is inconsistency between Linux/Windows