Network visualization and analysis with Cytoscape

Based on slides by Gary Bader (U Toronto)
Network Analysis Workflow

- Load Networks e.g. PPI data
  - Import network data into Cytoscape
- Load Attributes e.g. gene expression data
  - Get data about networks into Cytoscape
- Analyze and Visualize Networks
- Prepare for Publication
- A specific example of this workflow:
Network Visualization and Analysis
Outline

• Network introduction
• Network visualization
• Cytoscape software tool for network visualization and analysis
• Network analysis
Networks

• Represent relationships
  – Physical, regulatory, genetic, functional interactions

• Useful for discovering relationships in large data sets
  – Better than tables in Excel

• Visualize multiple data types together
  – See interesting patterns

• Network analysis
Biological Pathways/Networks?
Six Degrees of Separation

• Everyone in the world is connected by at most six links
• Which path should we take?
• Shortest path by breadth first search
  – If two nodes are connected, will find the shortest path between them
• Are two proteins connected? If so, how?
• Biologically relevant?

http://www.time.com/time/techtime/200406/community.html
Applications of Network Biology

- **Gene Function Prediction** – shows connections to sets of genes/proteins involved in same biological process

- **Detection of protein complexes/other modular structures** – discover modularity & higher order organization (motifs, feedback loops)

- **Network evolution** – biological process(es) conservation across species

- **Prediction of new interactions and functional associations** – Statistically significant domain-domain correlations in protein interaction network to predict protein-protein or genetic interaction
Applications of Network Informatics in Disease

- **Identification of disease subnetworks** – identification of disease network subnetworks that are transcriptionally active in disease.

- **Subnetwork-based diagnosis** – source of biomarkers for disease classification, identify interconnected genes whose aggregate expression levels are predictive of disease state.

- **Subnetwork-based gene association** – map common pathway mechanisms affected by collection of genotypes.

Agilent Literature Search

Mondrian, MSKCC

PinnacleZ, UCSD
What’s Missing?

• Dynamics
  – Pathways/networks represented as static processes
    • Difficult to represent a calcium wave or a feedback loop
  – More detailed mathematical representations exist that handle these e.g. Stoichiometric modeling, Kinetic modeling (VirtualCell, E-cell, …)
    • Need to accumulate or estimate comprehensive kinetic information

• Detail – atomic structures

• Context – cell type, developmental stage
What Have We Learned?

• Networks are useful for seeing relationships in large data sets
• Important to understand what the nodes and edges mean
• Important to define the biological question - know what you want to do with your gene list or network
• Many methods available for gene list and network analysis
  – Good to determine your question and search for a solution
  – Or get to know many methods and see how they can be applied to your data
Network Visualization Outline

- Automatic network layout
- Visual features
- Visually interpreting a network
Network Representations

<table>
<thead>
<tr>
<th>Relationships</th>
<th>Optional weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 ↔ A2</td>
<td>1</td>
</tr>
<tr>
<td>A1 ↔ A3</td>
<td>3</td>
</tr>
<tr>
<td>A2 ↔ A3</td>
<td>1</td>
</tr>
<tr>
<td>A2 ↔ A4</td>
<td>2</td>
</tr>
<tr>
<td>A2 ↔ A5</td>
<td>1</td>
</tr>
<tr>
<td>A3 ↔ A4</td>
<td>1</td>
</tr>
<tr>
<td>A3 ↔ A5</td>
<td>1</td>
</tr>
<tr>
<td>A3 ↔ A7</td>
<td>1</td>
</tr>
<tr>
<td>A5 ↔ A4</td>
<td>1</td>
</tr>
<tr>
<td>A5 ↔ A6</td>
<td>1</td>
</tr>
<tr>
<td>A6 ↔ A8</td>
<td>2</td>
</tr>
<tr>
<td>A6 ↔ A9</td>
<td>3</td>
</tr>
<tr>
<td>A8 ↔ A9</td>
<td>3</td>
</tr>
</tbody>
</table>

Network:
- A1
- A2
- A3
- A4
- A5
- A6
- A7
- A8
- A9

Network:
- $w_{12} = 1$
- $w_{13} = 3$
- $k_7 = 3$
- $c_s = 1/3$
- $c_6 = 1$
- $k_{4}^{in} = 3$
- $k_{4}^{out} = 2$

Heatmap:
- Node A3
- Node A2
- Node A1
- Node A7
- Node A4
- Node A5
- Node A6
- Node A8
- Node A9

Color Scale:
- 0
- 1
- 2
- 3

Legend:
- Node
- Edge
Automatic network layout

Before layout

After layout
Automatic network layout

• Force-directed: nodes repel and edges pull
• Good for up to 500 nodes
  – Bigger networks give hairballs - Reduce number of edges
• Advice: try force directed first, or hierarchical for tree-like networks
• Tips for better looking networks
  – Manually adjust layout
  – Load network into a drawing program (e.g. Illustrator) and adjust labels
Dealing with ‘hairballs’: zoom or filter

Zoom

Focus

PKC Cell Wall Integrity

Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry

- Synthetic Lethal
- Transcription Factor Regulation
- Protein-Protein Interaction

- Up Regulated Gene Expression
- Down Regulated Gene Expression
Visual Features

• Node and edge attributes
  – Text (string), integer, float, Boolean, list
  – E.g. represent gene, interaction attributes
• Visual attributes
  – Node, edge visual properties
  – Colour, shape, size, borders, opacity...
Visually Interpreting a Network

Data relationships
Guilt-by-association
Dense clusters
Global relationships
What Have We Learned?

- Automatic layout is required to visualize networks
- Networks help you visualize interesting relationships in your data
- Avoid hairballs by focusing analysis
- Visual attributes enable multiple types of data to be shown at once – useful to see their relationships
Network Visualization and Analysis using Cytoscape

- Network visualization and analysis using Cytoscape software
- Cytoscape basics
- Cytoscape network analysis examples
Network visualization and analysis

Pathway comparison
Literature mining
Gene Ontology analysis
Active modules
Complex detection
Network motif search

UCSD, ISB, Agilent, MSKCC, Pasteur, UCSF, Unilever, UToronto, UTexas
Network Analysis using Cytoscape

Find biological processes underlying a phenotype
Active Community

http://www.cytoscape.org

• Help
  – Tutorials, case studies
  – Mailing lists for discussion
  – Documentation, data sets

• Annual Conference: Paris, Oct 8-11, 2013

• 10,000s users, 5000 downloads/month

• >160 Plugins/Apps Extend Functionality
  – The app store: http://apps.cytoscape.org/
  – Build your own, requires programming

What Have We Learned?

• Cytoscape is a useful, free software tool for network visualization and analysis
• Provides basic network manipulation features
• Plugins/Apps are available to extend the functionality
yFiles Organic
Network Layout

- 15 algorithms available through plugins
- Demo: Move, zoom/pan, rotate, scale, align
Create Subnetwork
Create Subnetwork
Visual Style

• Customized views of experimental data in a network context

• Network has node and edge attributes
  • E.g. expression data, GO function, interaction type

• Mapped to visual attributes
  • E.g. node/edge size, shape, colour...

• E.g. Visualize gene expression data as node colour gradient on the network
Visual Style

Load “Your Favorite Network”
Visual Style

Load “Your Favorite Expression” Dataset
Map expression values to node colours using a continuous mapper.
Visual Style

Expression data mapped to node colours
Network Filtering
Interaction Database Search
DNA_Repair_(Reactome)

LIG1
Protein
Homo sapiens
Pathway Commons: 6311
Synonyms:
- LIG1
Links:
- UNIPROT: P18858
- UNIPROT: Q32223
- REF_SEQ: NP_000225
- Search iHOP

Visual Legend
Procesive complex: Okazaki fragment complex
Complex
Homo sapiens
Pathway Commons: 5256
Links:
- REACTOME: 68453
- REACTOME: REACT_5537
Gene List and Network Analysis Overview

- Gene List
  - Gene Attributes (Annotation (Ensembl), Expression, Phenotypes)

  Network
    - Protein-Protein Interactions
    - Functional Interactions
    - Regulatory Network
      - iRefWeb, GeneMANIA, AgilentLitSearch, STRING

  Network Visualization
  - Cytoscape
    - Gene set enrichment analysis (BinGO)
    - Regulatory network analysis (NetMatch)
    - Gene function prediction (STRING, GeneMANIA)
    - Module detection (network clustering) (ClusterMaker, ActiveModules, Reactome FI)
VistaClara

- Visualization for gene expression data
- Heat maps, sorting, animation
Cytoscape Lab

• Cytoscape – expression data visualization
  – Load the sample network file: galFiltered.sif
  – Lay it out – try different layouts
  – Load expression data - galExpData.pvals
    • Use File->Import->Attribute from Table
  – Examine node attributes
  – Visualize gene expression data using the Visual Mapper
  – Install the VistaClara plugin from the plugin manager
  – Play the expression data as a movie
BiNGO plugin

• Calculates over-representation of a subset of genes with respect to a background set in a specific GO category
• Input: subnetwork, or list
  – Background set by user
• Output: tree with nodes color reflecting overrepresentation; also as lists

• Caveats: Gene identifiers must match; low GO term coverage, GO bias, Background determining
BiNGO

Hypergeometric p-value
Multiple testing correction (Benjamini-Hochberg FDR)

Maere, S., Heymans, K. and Kuiper, M
Bioinformatics 21, 3448-3449, 2005
Cerebral

http://www.pathogenomics.ca/cerebral
Find Active Subnetworks

• Active modules
  – Input: network + p-values for gene expression values e.g. from GCRMA
  – Output: significantly differentially expressed subgraphs

• Method
  – Calculate z-score/node, $Z_A$ score/subgraph, correct vs. random expression data sampling
  – Score over multiple experimental conditions
  – Simulated annealing used to find high scoring networks

Ideker T, Ozier O, Schwikowski B, Siegel AF
Active Module Results

Network: yeast protein-protein and protein-DNA network
Expression data: 3 gene knock out conditions (enzyme, TF activator, TF repressor)

Note: non-deterministic, multiple runs required for confidence of result robustness

Network Clustering

- Clusters in a protein-protein interaction network have been shown to represent protein complexes and parts of pathways
- Clusters in a protein similarity network represent protein families
- Network clustering is available through the ClusterMaker Cytoscape plugin

<table>
<thead>
<tr>
<th>Rank</th>
<th>Score</th>
<th>Size</th>
<th>Names</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>5.417</td>
<td>12,60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Create a new child network.  Save  Done
Text Mining

• Computationally extract gene relationships from text, usually PubMed abstracts
• Useful if network is not in a database
  – Literature search tool
• BUT not perfect
  – Problems recognizing gene names
  – Natural language processing is difficult
• Agilent Literature Search Cytoscape plugin
• iHOP (www.ihop-net.org/UniPub/iHOP/)
Cytoscape Network produced by Literature Search.

Abstract from the scientific literature

Sentences for an edge

2. High glucose induced phosphorylation of phosphatidylinositol 3-kinase (PI 3-K) and extracellular signal-regulated kinase (ERK)1/2 along with bcl-xL and bfl-1/A1 upregulation.
Analysis Lab

Find Network Motifs - Netmatch plugin

• Network motif is a sub-network that occurs significantly more often than by chance alone
• Input: query and target networks, optional node/edge labels
• Output: topological query matches as subgraphs of target network
• Supports: subgraph matching, node/edge labels, label wildcards, approximate paths
• http://alpha.dmi.unict.it/~ctnyu/netmatch.html
Finding specific biological relevant TF-PPI sub-networks
Find Signaling Pathways

- Potential signaling pathways from plasma membrane to nucleus via cytoplasm

Signaling pathway example

NetMatch query

NetMatch Results

- Shortest path between subgraph matches
Find Expressed Motifs

Find specific subgraphs where certain nodes are significantly differentially expressed

Protein | Differential Expression Significance
---|---
YLR075W | 1.7255E-4
YGR085C | 2.639E-4
YPR102C | 3.7183E-4
“Root graph”

- “There is one graph to rule them all....”
- The networks in Cytoscape are all “views” on a single graph.
- Changing the attribute for a node in one network will also change that attribute for a node with the same ID in all other loaded networks.
- There is no way to “copy” a node and keep the same ID.
- Make a copy of the session.
Cytoscape 2.8 Tips & Tricks

• Network views
  – When you open a large network, you will not get a view by default
  – To improve interactive performance, Cytoscape has the concept of “Levels of Detail”
    • Some visual attributes will only be apparent when you zoom in
    • The level of detail for various attributes can be changed in the preferences
    • To see what things will look like at full detail:
      – View→Show Graphics Details
Cytoscape 2.8 Tips & Tricks

• Sessions
  – Sessions save pretty much everything:
    • Networks
    • Properties
    • Visual styles
    • Screen sizes
  – Saving a session on a large screen may require some resizing when opened on your laptop
Cytoscape 2.8 Tips & Tricks

• Logging
  – By default, Cytoscape writes its logs to the Error Dialog: Help ➔ Error Dialog
  – Can change a preference to write it to the console
    • Edit ➔ Preferences ➔ Properties...
    • Set logger.console to true
    • Don’t forget to save your preferences
    • Restart Cytoscape
  – (can also turn on debugging: cytoscape.debug, but I don’t recommend it)
Cytoscape 2.8 Tips & Tricks

• Memory
  – Cytoscape uses lots of it
  – Doesn’t like to let go of it
  – An occasional restart when working with large networks is a good thing
  – Destroy views when you don’t need them
  – Java doesn’t give us a good way to get the memory right at start time
    • Since version 2.7, Cytoscape does a much better job at “guessing” good default memory sizes than previous versions
Cytoscape 2.8 Tips & Tricks

• .cytoscape directory
  – Your defaults and any plugins downloaded from the plugin manager will go here
  – Sometimes, if things get really messed up, deleting (or renaming) this directory can give you a “clean slate”

• Plugin manager
  – “Outdated” doesn’t necessarily mean “won’t work”
  – Plugin authors don’t always update their plugins immediately after new releases
Cytoscape 3.0