Canadian Bioinformatics Workshops

www.bioinformatics.ca
Module 3
Network Visualization and analysis with Cytoscape

Gary Bader
Pathway and Network Analysis of –omic Data
June 13, 2016

Network Analysis Workflow

- A specific example of this workflow:

Module 3: Network Visualization with Cytoscape
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

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
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 and Analysis using Cytoscape

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

http://cytoscape.org

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

UCSD, ISB, Agilent, MSKCC, Pasteur, UCSF
Manipulate Networks

Filter/Query

Automatic Layout

Interaction Database Search

The Cytoscape App Store

http://apps.cytoscape.org

Pathway analysis
Gene expression analysis
Complex detection
Literature mining
Network motif search
Pathway comparison
Active Community

http://www.cytoscape.org

• 10,000s users, >8,000 downloads/month

• Help
  – Documentation, data sets
  – Mailing lists
  – http://tutorials.cytoscape.org

• Annual Conference

• >280 Apps Extend Functionality
  – 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
• Apps are available to extend the functionality

Cytoscape Demo

Version 3.4.0
www.cytoscape.org
Network overview
CytoPanels
Network manager
Attribute browser
Canvas
Desktop
Network Layout

- Many algorithms available through apps
- Demo: Move, zoom/pan, rotate, scale, align
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 Expression” Dataset

Map expression values to node colours using a continuous mapper
Visual Style

Expression data mapped to node colours

Network Filtering

Module 3: Network Visualization with Cytoscape
Module 3: Network Visualization with Cytoscape

bioinformatics.ca
Module 3: Network Visualization with Cytoscape

- **GeneMANIA**
- **Enrichment Map**
- **Reactome FI**
- **jActiveModules**
- **ClusterMaker**
- **PubMed**
- **g:Profiler**
- **Pathvisio**
- **GeneMANIA Network Assessor**

**FYI**

**Mechanistic drill down**

- Drill down to mechanism (e.g., overlay genomics data on pathway or network, pathway diagrams, Cytoscape)

**Identify interesting pathways**

- Pathway enrichment analysis for rare events (sensitive test on original samples)
- Rank-based pathway enrichment analysis (GREAT g:Profiler)
- Last-based pathway enrichment analysis

**Identify interesting networks**

- Find network of gene relationships (Reactome FI, GeneMANIA)
- Visualize and identify network modules and their functions (Cytoscape, Enrichment Map)

**Measure mRNA expression** (e.g., RNA-Seq)
- Compute differential expression, methylation (e.g., exp. vs. control, time series)
- Score methylation at gene promoters
- Identify known or predicted targets

**Measure protein expression** (e.g., Mass spec)
- Measure DNA methylation
- Measure protein binding to DNA or RNA (e.g., ChIP, CLIP)
- Identify target genes

**Measure protein interaction network** (e.g., Y2H, Mass spec)
- Compare to known networks (GeneMANIA Network Assessor)
- Develop mechanistic model explaining data

**Measure DNA methylation**
- Find network of gene relationships
- Score methylation at gene promoters

**Measure SNP or CNVs in cases and controls**
- Identify significant or recurrent gene-associated variants in cases

**Measure differentially expressed, methylated (e.g., exp. vs. control, time series)**
- Identify and filter variants

**Sequence cancer and/or normal genomes**
- Identify and filter variants

**Identify target genes**
- List-based pathway enrichment analysis
- Identify target genes

**Integrate additional information** e.g., disease signatures, miRNA/TF targets, drug targets
- Develop mechanistic model explaining data
BiNGO app

• 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
Find Active Subnetworks

- Active modules (jActiveModules app)
  - 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


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 ClusterMaker2 Cytoscape app

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 app
Module 3: Network Visualization with Cytoscape

Cytoscape Network produced by Literature Search.

Abstract from the scientific literature

Sentences for an edge

1. \( \text{High glucose puffs} \) might protect in human coronary artery smooth muscle cells by increasing TSC2 expression.
2. \( \text{High glucose puffs} \) were protective in human coronary artery smooth muscle cells by increasing TSC2 expression.

Network Motifs – NetMatchStar app

- 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://apps.cytoscape.org/apps/netmatchstar

Analysis Lab

Find Network Motifs – NetMatchStar app

Module 3: Network Visualization with Cytoscape

bioinformatics.ca
Finding specific biological relevant TF-PPI sub-networks

Ferro et al. Bioinformatics 2007

Find Signaling Pathways
• Potential signaling pathways from plasma membrane to nucleus via cytoplasm

Signaling pathway example
NetMatch query
MAP Kinase Cascade

Nucleus - Growth Control Mitogenesis

NetMatch query
Shortest path between subgraph matches

NetMatch Results
Find Expressed Motifs

Find specific subgraphs where certain nodes are significantly differentially expressed

<table>
<thead>
<tr>
<th>Protein</th>
<th>Differential Expression Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>YLR075W</td>
<td>1.7255E-4</td>
</tr>
<tr>
<td>YGR085C</td>
<td>2.639E-4</td>
</tr>
<tr>
<td>YPR102C</td>
<td>3.7183E-4</td>
</tr>
</tbody>
</table>

Cytoscape 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 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 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
  – Cytoscape tries to “guess” good default memory settings based on how much memory you have, but it can fail
Cytoscape Tips & Tricks

• CytoscapeConfiguration directory
  – In your user/home directory
  – Your defaults and any apps downloaded from the app store will go here
  – Sometimes, if things get really messed up, deleting (or renaming) this directory can give you a “clean slate” (it will remove all your apps)

• Install and test apps one at a time. If an app interferes with Cytoscape, you will be able to identify it.