Canadian Bioinformatics Workshops

www.bioinformatics.ca
Module 5: Network Visualization and Analysis

Gary Bader
Pathway and Network Analysis of –omics Data
June 4-6, 2014

Network Analysis Workflow

- A specific example of this workflow:

Networks → Attributes → Analyze → Publish
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

Module 5: Network Visualization and Analysis  bioinformatics.ca

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

Module 5: Network Visualization and Analysis  bioinformatics.ca
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

#### Optional weight

<table>
<thead>
<tr>
<th>Relationships</th>
<th>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 ↔ A6</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 ↔ A6</td>
<td>1</td>
</tr>
<tr>
<td>A6 ↔ A8</td>
<td>1</td>
</tr>
<tr>
<td>A6 ↔ A9</td>
<td>2</td>
</tr>
<tr>
<td>A8 ↔ A9</td>
<td>3</td>
</tr>
</tbody>
</table>

#### Network

![Network Diagram](image)

#### Heatmap

![Heatmap](image)

### Automatic network layout

#### Before layout

![Before Layout](image)

#### After layout

![After Layout](image)
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

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

- Red: Up Regulated Gene Expression
- Green: 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

Network Analysis using Cytoscape
Find biological processes underlying a phenotype
Network Information
Databases
Literature
Expert knowledge
Experimental Data
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, >5000 downloads/month
• Help
  – Documentation, data sets
  – Mailing lists
  – http://tutorials.cytoscape.org
• Annual Conference: TBD, North America 2014
• >200 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 3.1.1

www.cytoscape.org
Network Layout

- Many algorithms available through apps
- Demo: Move, zoom/pan, rotate, scale, align
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
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 5: Network Visualization and Analysis  bioinformatics.ca
Interaction Database Search

Module 5: Network Visualization and Analysis
Module 5: Network Visualization and Analysis

Gene List and Network Analysis Overview

- **Gene List**
- **Gene Attributes**
  - Annotation (Ensembl)
  - Expression
  - Phenotypes
- **Network**
  - Protein-Protein Interactions
  - Functional Interactions
- **Regulatory Network**
  - Regulatory network analysis
  - Gene function prediction
  - Gene set enrichment analysis
- **Network Visualization**
  - Cytoscape
  - BinGO
  - NetMatch
  - STRING GeneMANIA
  - Module detection (network clustering)
  - iRefWeb, GeneMANIA, AgilentLitSearch, STRING
We are on a Coffee Break & Networking Session

Module 5: Network Visualization and Analysis

Gene List and Network Analysis Overview

- Gene List
- Gene Attributes
- Network
  - Protein-Protein Interactions
  - Functional Interactions
  - Regulatory Network
- 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

Annotation (Ensembl) Expression Phenotypes
iRefWeb, GeneMANIA, AgilentLitSearch, STRING
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

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

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

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


Module 5: Network Visualization and Analysis

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


Module 5: Network Visualization and Analysis

## 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

1. High glucose stability upregulates bovine coronary artery smooth muscle cells by increasing HIF-1α and HIF-1β.
2. High glucose enhances phosphorylation of glycogen synthase, 3-phosphoglycerate dehydrogenase, and extracellular signal-regulated kinase (ERK) along with HDAC 4, 5, and 16, respectively.
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

Ferro et al. Bioinformatics 2007
Find Signaling Pathways

• Potential signaling pathways from plasma membrane to nucleus via cytoplasm

Find Expressed Motifs

<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
  – 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 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”