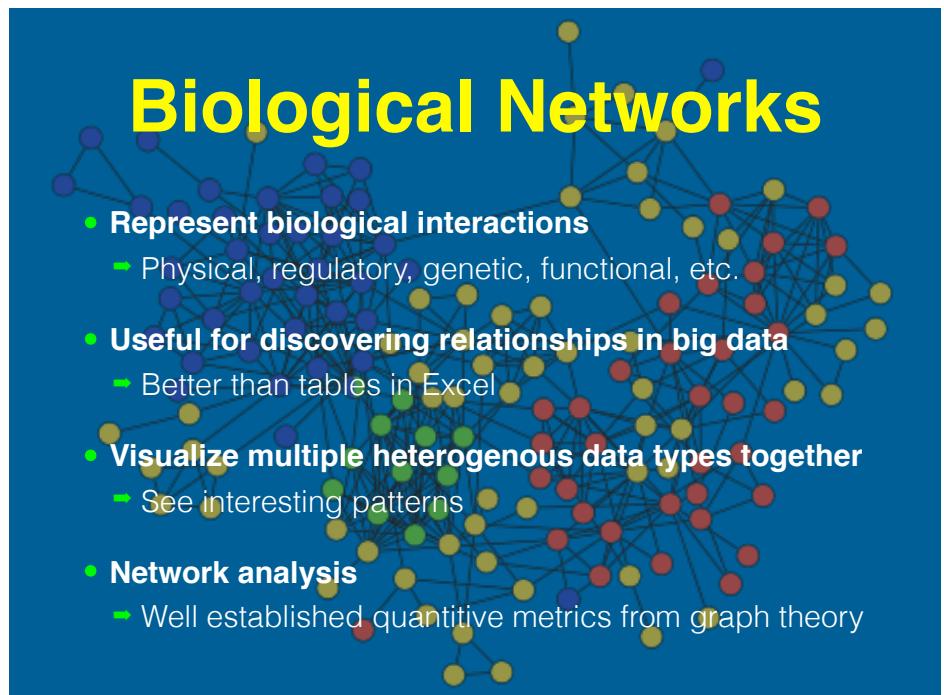


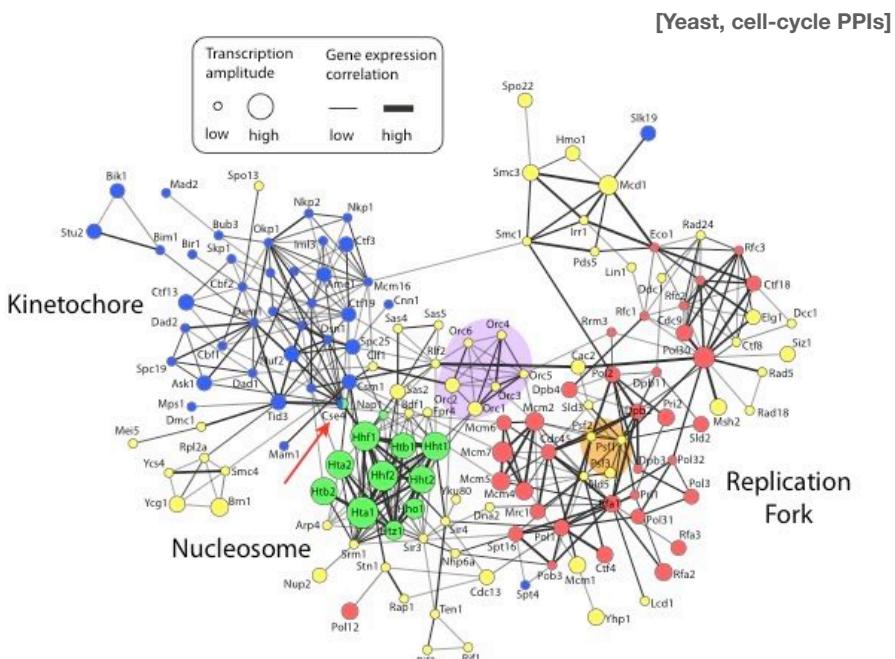
TODAYS MENU:

- ▶ Network introduction
- ▶ Network visualization
- ▶ Network analysis
- ▶ Hands-on:
Cytoscape and R (igraph) software tools
for network visualization and analysis

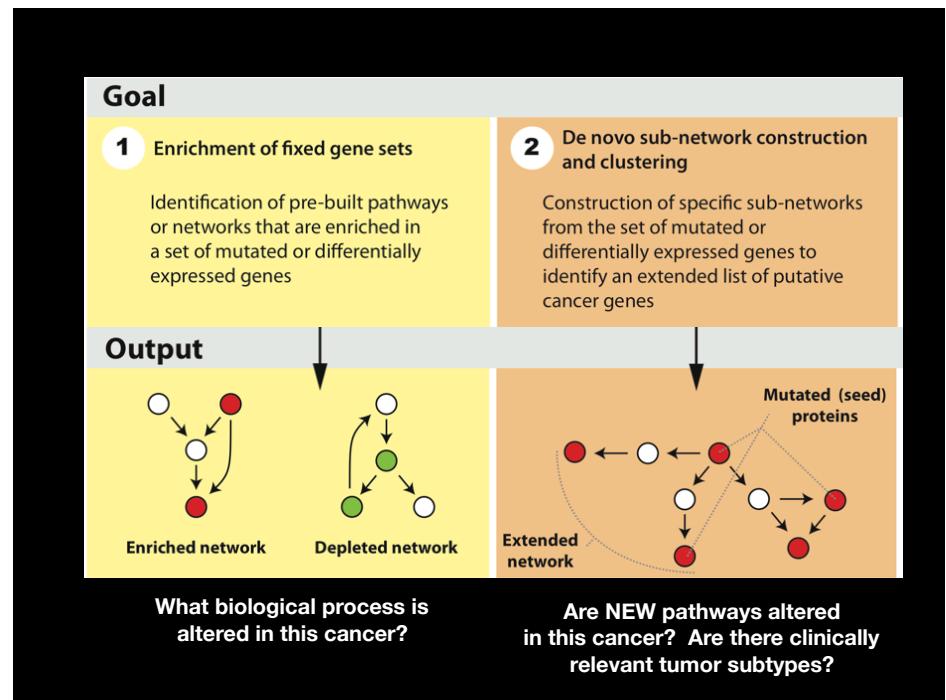
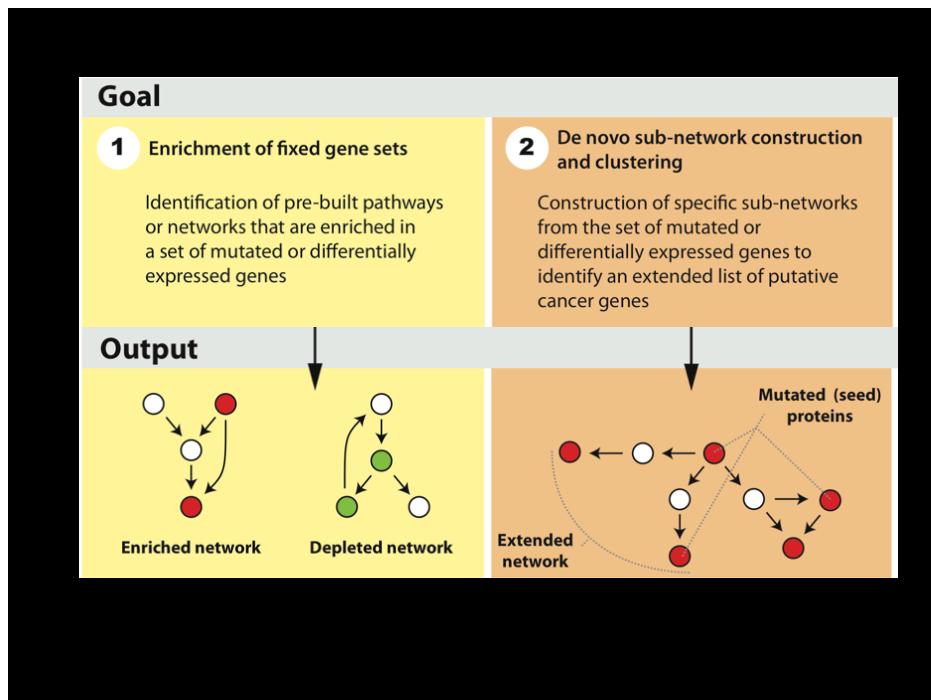
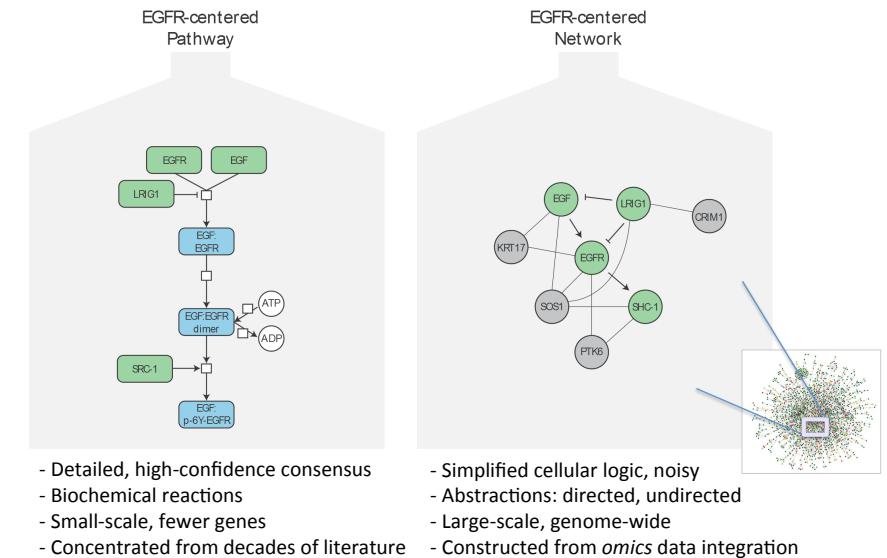
TODAYS MENU:

- ▶ Network introduction
- ▶ Network visualization
- ▶ Network analysis
- ▶ Hands-on:
Cytoscape and R (igraph) software tools
for network visualization and analysis





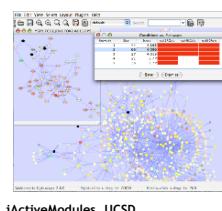
Pathways vs Networks



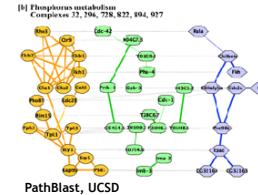
Benefits of Pathway Data vs. transcripts, proteins, SNPs...

- Improves sta2s2cal power
 - Fewer tests
- More reproducible
 - E.g. gene expression signatures
- Easier to interpret
 - Familiar concepts e.g. cell cycle
- Identifies mechanism
 - Can explain cause
- Predicts new roles to genes
 - Guilt by association, "birds of a feather"

Applications of Network Biology



jActiveModules, UCSD



PathBLAST, UCSD

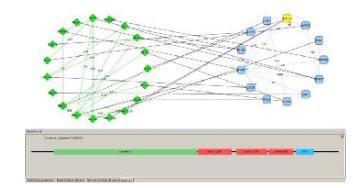
- **Gene Function Prediction** – shows connections to sets of genes/proteins involved in same biological process



MCODE, University of Toronto

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

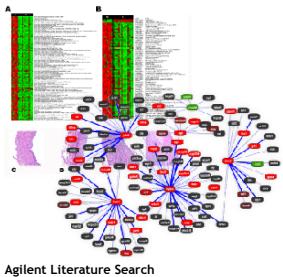
- **Network evolution** – biological process(es) conservation across species



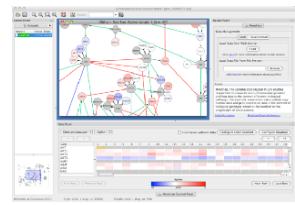
DomainGraph, Max Planck Institute

- **Prediction of new interactions and functional associations** – Statistically significant domain-domain correlations in protein interaction network to predict protein-protein or genetic interaction; allosteric in molecular networks

Applications of Network Informatics in Disease

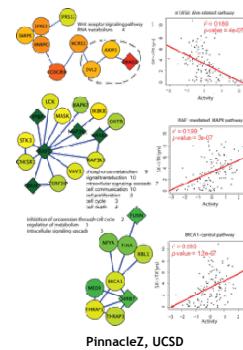


Agilent Literature Search



Mondrian, MSKCC

- **Identification of disease subnetworks** – identification of disease network subnetworks that are transcriptionally active in disease (disease sub-types).
- **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



Slide from: humangenetics-amc.nl

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, ...)
- **Detail** – atomic structures & exclusivity of interactions.
- **Context** – cell type, developmental stage

Slide from: humangenetics-amc.nl

What have we learned so far...

- 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 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

TODAY'S MENU:

► Network introduction

► Network visualization

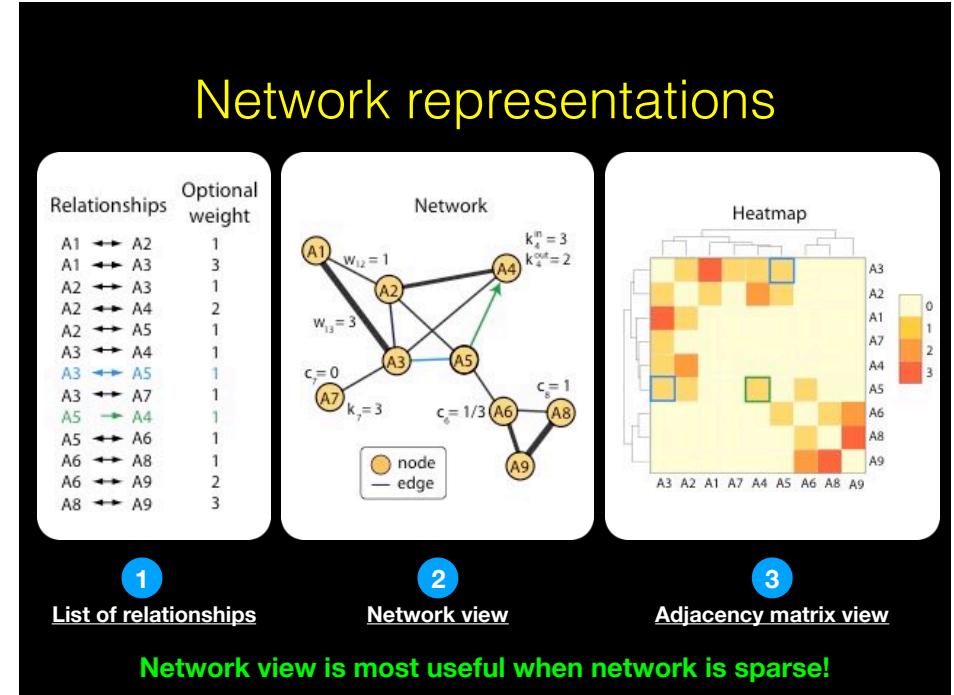
► Network analysis

► Hands-on:

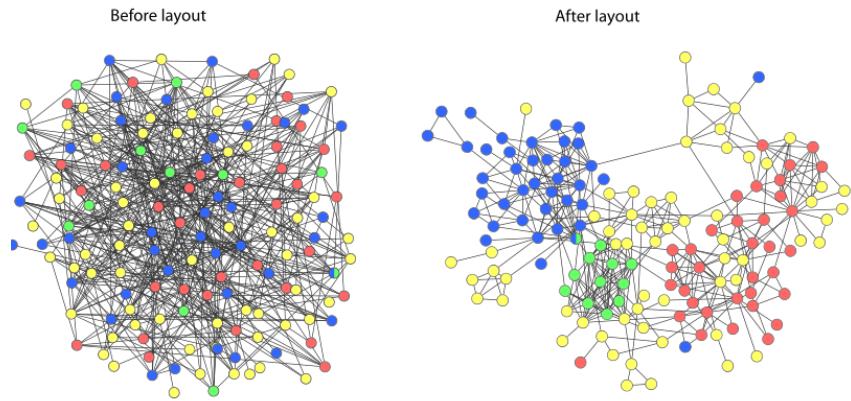
Cytoscape and R (igraph) software tools
for network visualization and analysis

Network Visualization Outline

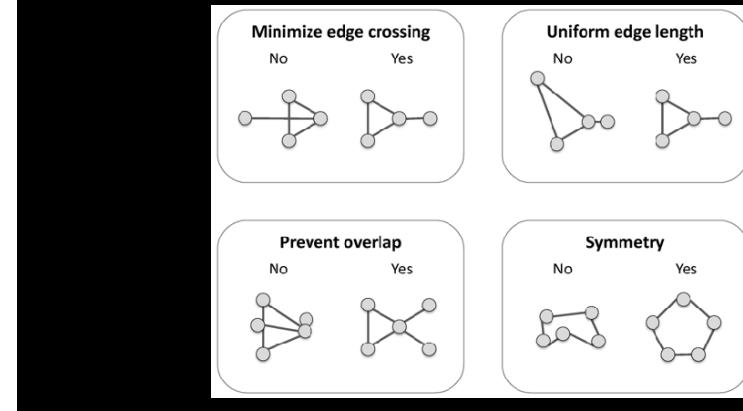
- Automatic network layout
- Visual features
- Visually interpreting a network



Automatic network layout



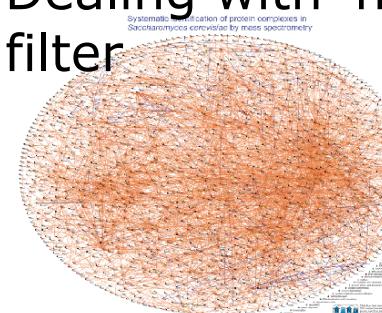
- Modern **graph layouts** are optimized for speed and aesthetics. In particular, they seek to minimize overlaps and edge crossing, and ensure similar edge length across the graph.



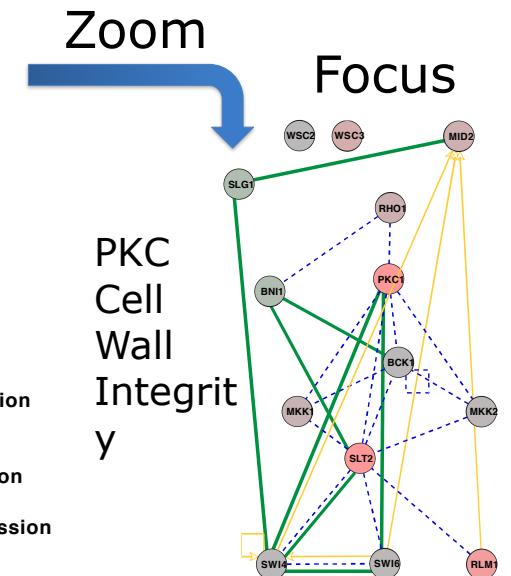
Force-directed layout: nodes repel and edges pull

- Good for up to 500 nodes
 - Bigger networks give hairballs
 - Reduce number of edges
 - Or just use a heatmap for dense networks
- 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

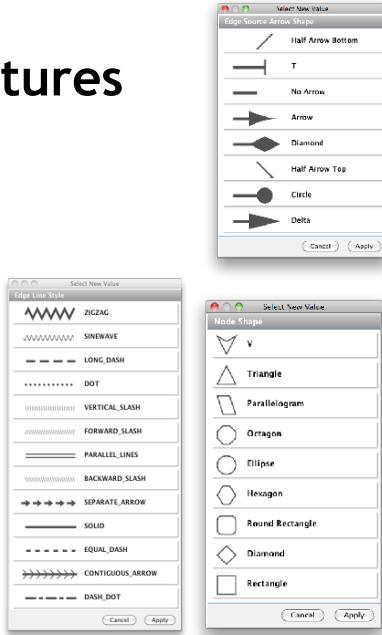


Synthetic Lethal
Transcription Factor Regulation
Protein-Protein Interaction
Up Regulated Gene Expression
Down Regulated Gene Expression

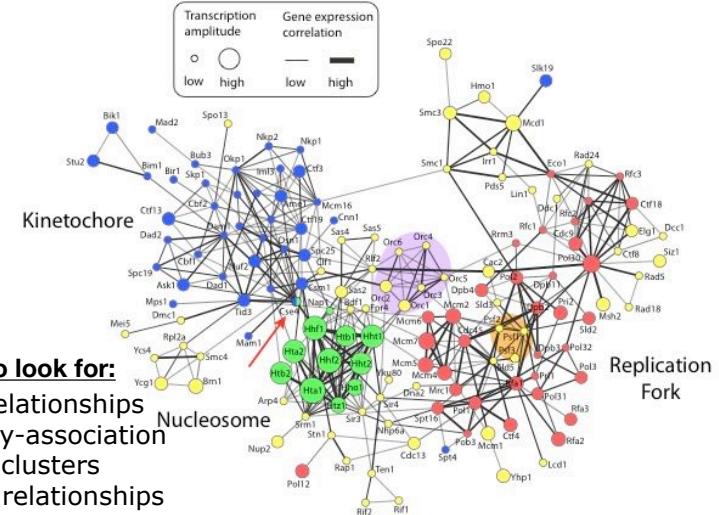


Visual Features

- Node and edge attributes
 - Text (string), integer, float, Boolean, list
 - E.g. represent gene, interaction attributes
- Visual attributes
 - Node, edge visual properties
 - Color, shape, size, borders, opacity...



Visually Interpreting a Network



What to look for:

- Data relationships
- Guilt-by-association
- Dense clusters
- Global relationships

What have we learned so far...

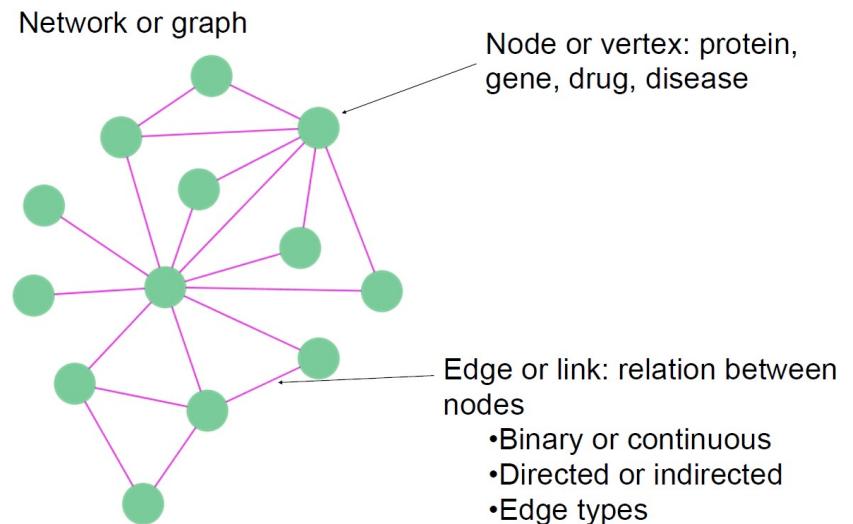
- Automatic layout is required to visualize networks
- Networks help you visualize interesting relationships in your data
- Avoid hairballs by focusing analysis
- Visual attributes enable multiple types of data to be shown at once – useful to see their relationships

TODAY'S MENU:

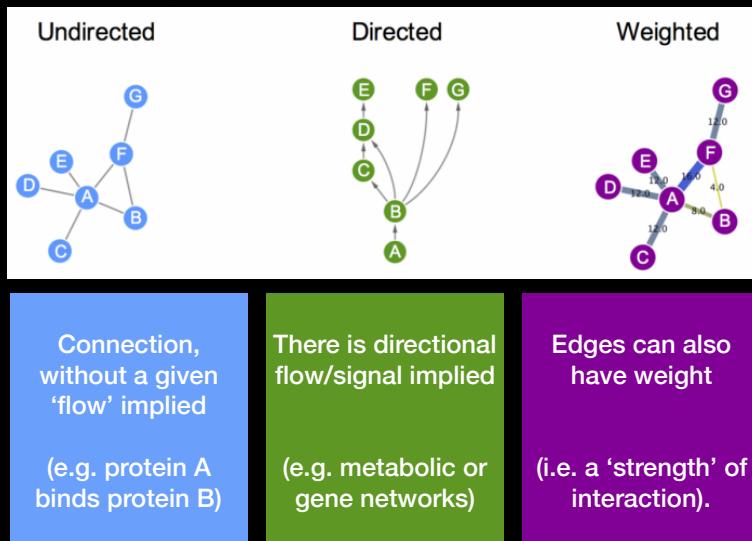
- Network introduction
- Network visualization
- **Network analysis**
- Hands-on:
 - Cytoscape and R (igraph) software tools for network visualization and analysis

Introduction to graph theory

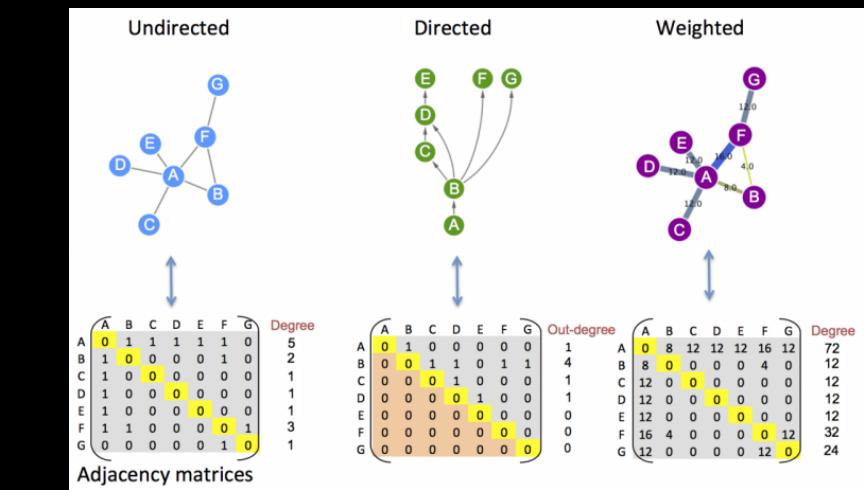
- Biological network analysis historically originated from the tools and concepts of **social network analysis** and the application of **graph theory** to the social sciences.
- Wikipedia defines graph theory as:
 - "[...] the study of graphs used to model pairwise relations between objects. A graph in this context is made up of **vertices** connected by **edges**".
- In practical terms, it is the set of concepts and methods that can be used to visualize and analyze networks



Types of network edges

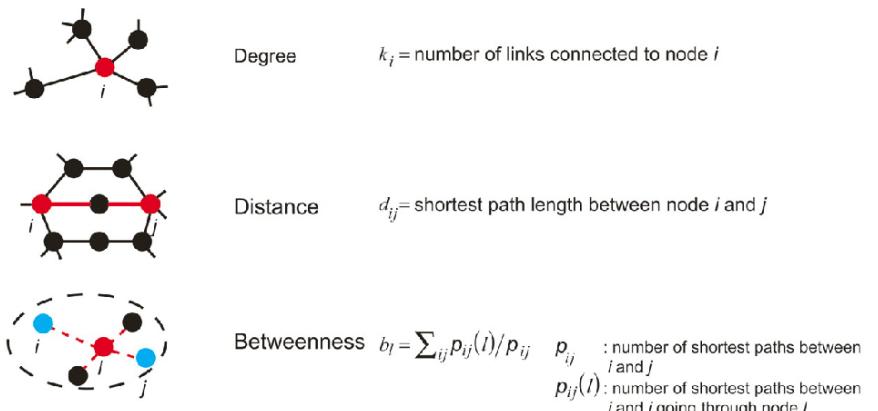


- Every network can be expressed mathematically in the form of an adjacency matrix

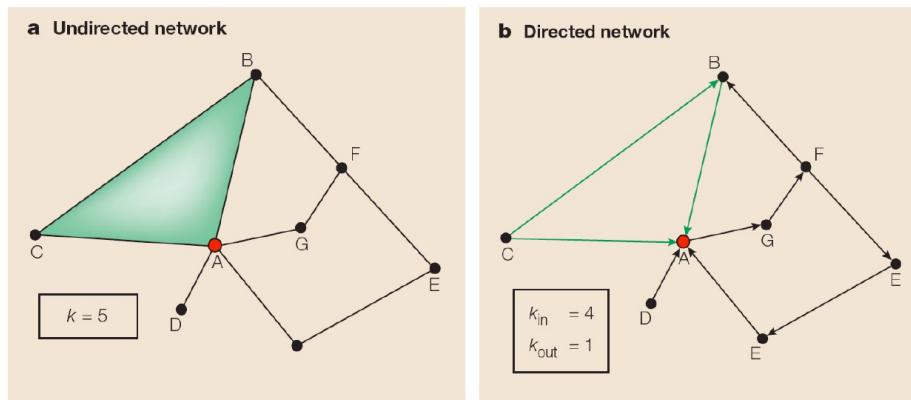


Network topology

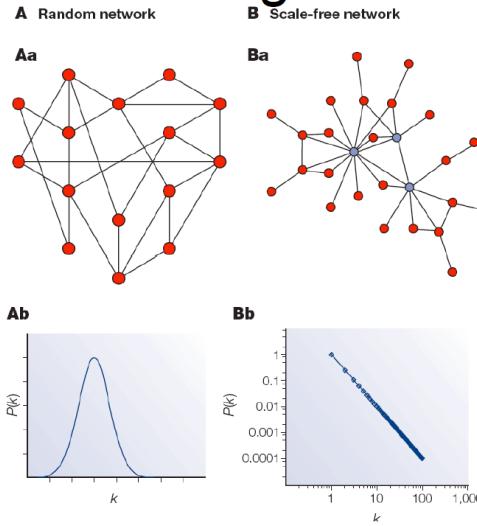
- Topology is the way in which the nodes and edges are arranged within a network.
- The most used topological properties and concepts include:
 - Degree** (i.e. how many node neighbors)
 - Communities** (i.e. clusters of well connected nodes)
 - Shortest Paths** (i.e. shortest distance between 2 nodes)
 - Centralities** (i.e. how ‘central’ is a given node?)
 - Betweenness** (a measure of centrality based on shortest paths)



Network Measures: Degree



Degree Distribution



$P(k)$ is probability of each degree k , i.e. fraction of nodes having that degree.

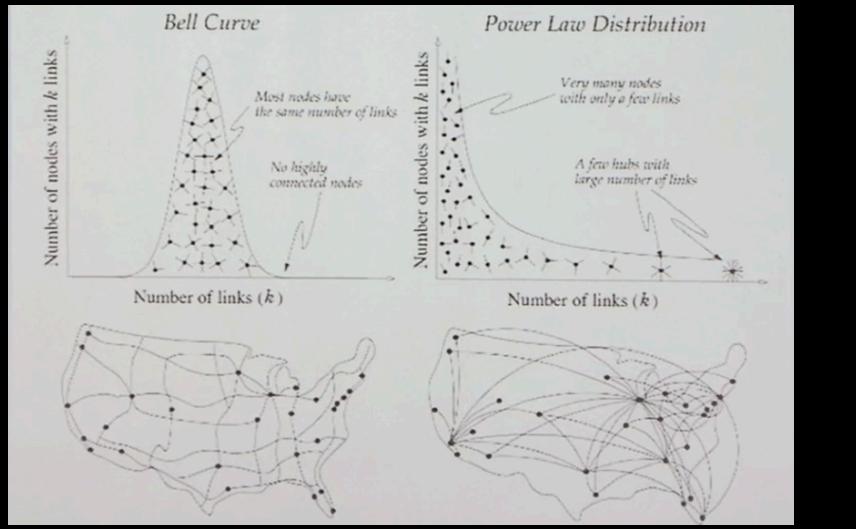
For random networks, $P(k)$ is normally distributed.

For real networks the distribution is often a power-law:

$$P(k) \sim k^{-\gamma}$$

Such networks are said to be **scale-free**

Random graphs vs scale free



Scale-Free Networks are Robust

- Complex systems (cell, internet, social networks), are resilient to component failure
- Network topology plays an important role in this robustness
 - Even if ~80% of nodes fail, the remaining ~20% still maintain network connectivity
- *Attack vulnerability* if hubs are selectively targeted
- In yeast, only ~20% of proteins are lethal when deleted, and are 5 times more likely to have degree $k > 15$ than $k < 5$.

Implications

- Many biological networks (protein-protein interaction networks regulatory networks, etc...) are thought to have hubs, or nodes with high degree.
- For protein-protein interaction networks (PPIs) these hubs have been shown to be older [1] and more essential than random proteins [2]
 - [1] Fraser et al. *Science* (2002) 296:750
 - [2] Jeoung et al. *Nature* (2001) 411:41

TODAYS MENU:

- **Network introduction**
- **Network visualization**
- **Network analysis**

‣ **Hands-on:**

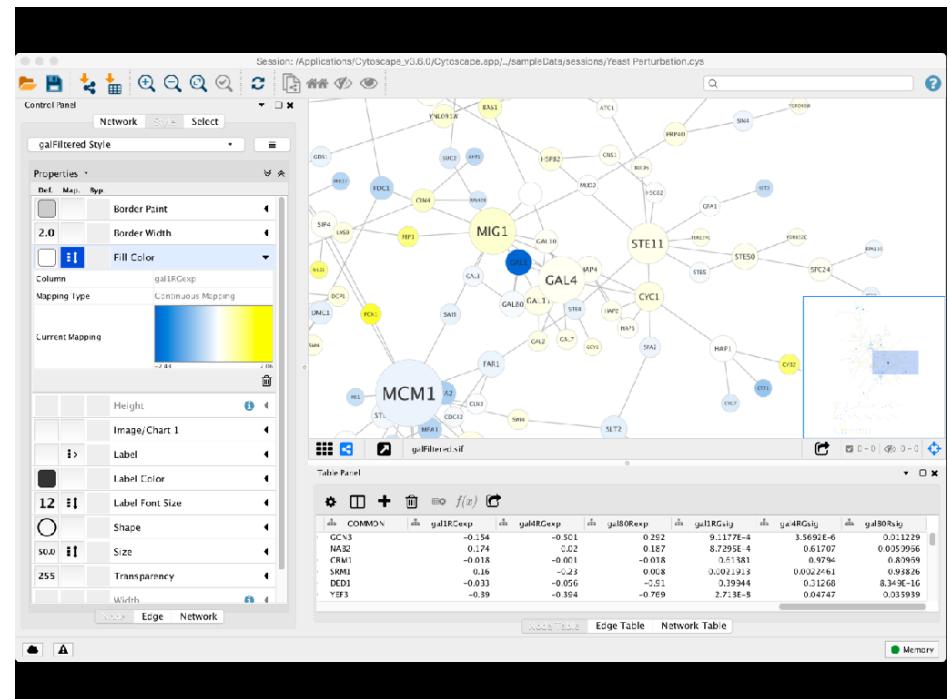
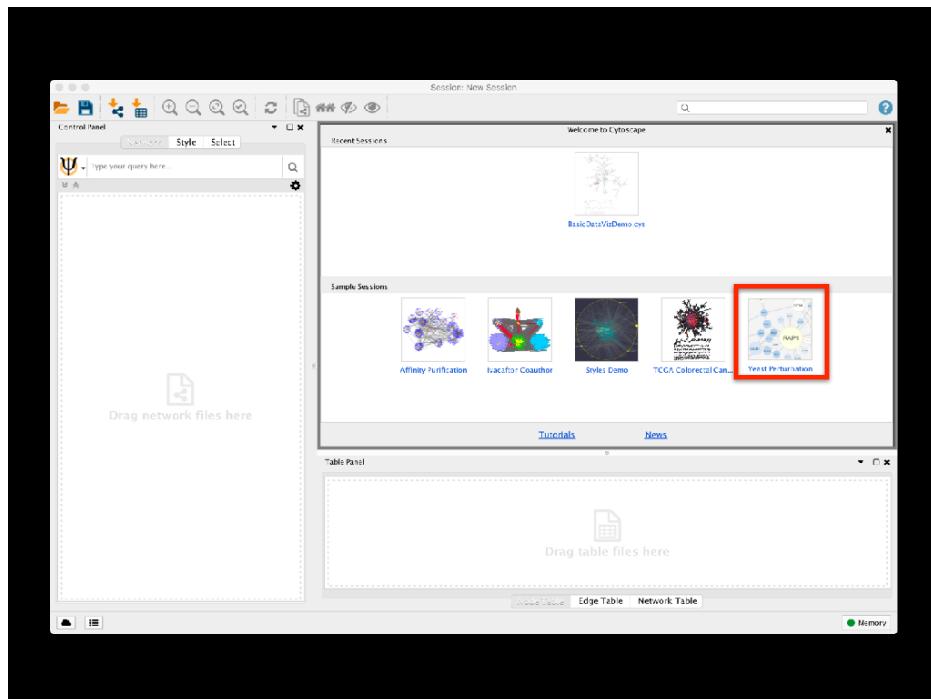
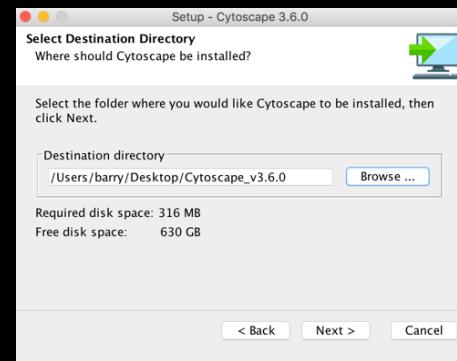
- Cytoscape and R (igraph) software tools for network visualization and analysis

Practical issues

- Major tools for the **creation, manipulation** and **visualization** of biological networks include:
 - ➡ Cytoscape
 - ➡ R packages (`igraph`, `graph`, `tidygraph`)
- Tools for network analysis and modeling include:
 - ➡ Cytoscape apps/plugins
 - ➡ R packages (`igraph` and others)
 - ➡ ByoDyn, COPASI

<http://cytoscape.org/download.php>

Note: If you are on a classroom Mac then be sure to install to your **Desktop** directory!



Cytoscape Memory Issues

- Cytoscape uses lots of memory and 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
- Since version 2.7, Cytoscape does a much better job at "guessing" good default memory sizes than previous versions but it still not great!
 - ➡ Java doesn't give us a good way to get the memory right at start time

Cytoscape 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

Hands-on: Part 1

https://bioboot.github.io/bimm143_W18/lectures/#17

- The data used in **part 1** is from yeast, and the genes Gal1, Gal4, and Gal80 are all yeast transcription factors. The experiments all involve some perturbation of these transcription factor genes.
- In this network view, the following node attributes have been mapped to visual style properties in cytoscape:
 - ➡ The "gal80exp" expression values are used for Node Fill Color.
 - ➡ The Default Node Color, for nodes with no data mapping, is dark grey.
 - ➡ Nodes with expression values that are significant are rendered as rectangles, others are ovals.
 - ➡ The common name for each gene is used as the Node Label.

Hands-on: Part 2

https://bioboot.github.io/bimm143_W18/lectures/#17

- The data used in **part 2** is from an ocean metagenomic sequencing project - where all the genetic material in a sample of ocean water is sequenced.
- We will use the R package **igraph** and the bioconductor package **RCy3** together with Cytoscape.
- Many of these microbial species in these types of studies have not yet been characterized in the lab.
 - ➡ Thus, to know more about the organisms and their interactions, we can observe which ones occur at the same sites.
 - ➡ One way to do that is by using **co-occurrence networks** where you examine which organisms occur together at which sites.

Summary

- Introduced the major themes of biological network analysis and the concepts underlying network **visualization** and **analysis**
- **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
- The R **igraph** package has extensive network analysis functionality beyond that in Cytoscape
- The R bioconductor package **RCy3** package allows us to bring networks and associated data from R to Cytoscape so we can have the best of both worlds.

Network Analysis Overview

