**BIMM-143: INTRODUCTION TO BIOINFORMATICS (Lecture 17)**

# Biological Network Analysis (Part 1)

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

Dr. Barry Grant

**Overview:** In this first for two hands-on sessions we will introduce Cytoscape < http://cytoscape.org >; an open source software platform for integrating, visualizing, and analyzing measurement data in the context of networks.

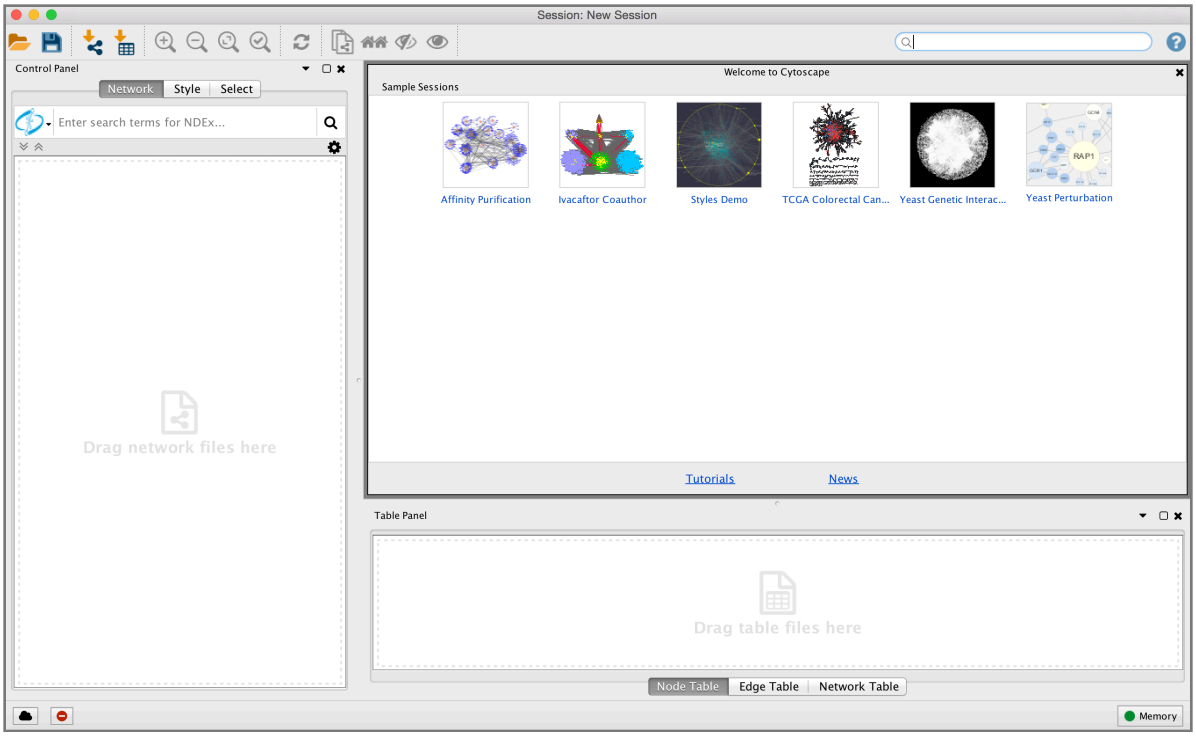
This first session presents a scenario of how expression and network data can be combined to tell a biological story and includes these concepts:

- Visualizing networks using expression data.
- Filtering networks based on expression data.
- Assessing expression data in the context of a biological network.

## Section 1. Loading Network

Start Cytoscape and download the demo network: CytoscapeDemo_01.cys

After launching Cytoscape a new window will appear that looks like this:

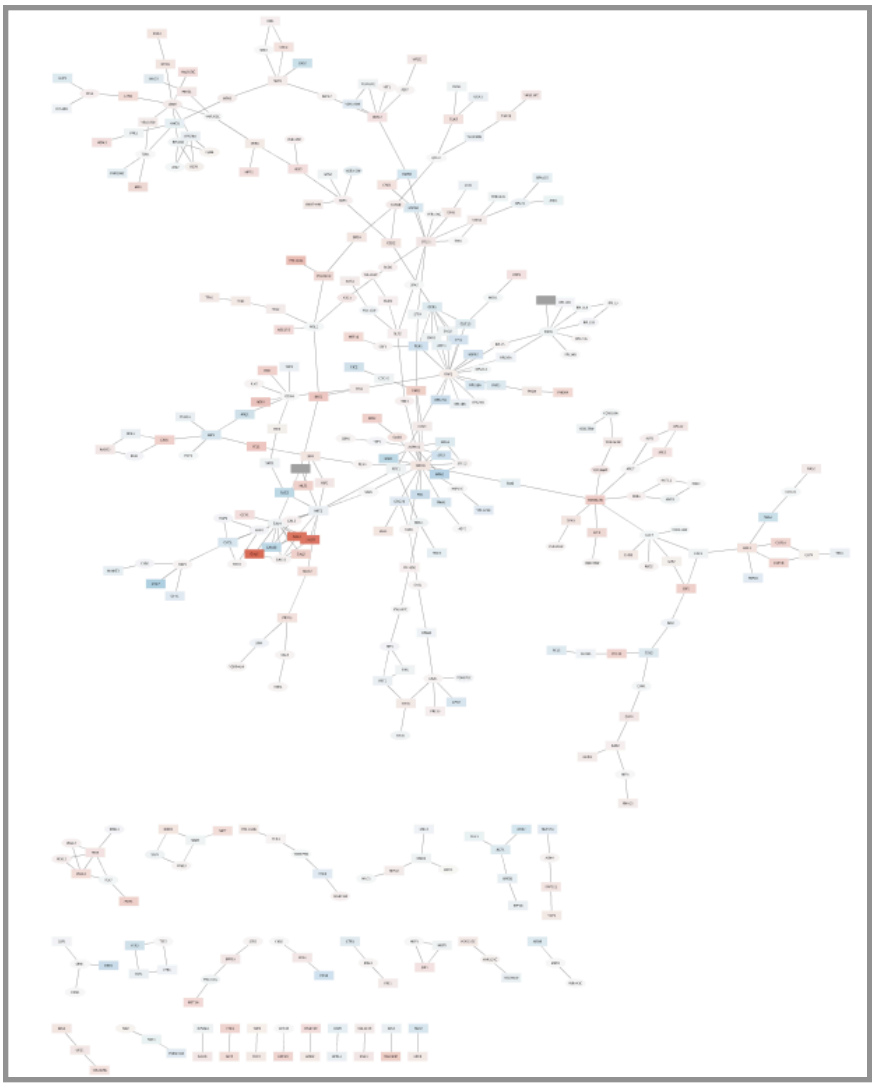


- Open the demo network you downloaded using **File > Open...**

**Note**: When the network first opens, the entire network may not be visible because of the default zoom factor used. To see the whole network, we can use the **View > Fit Content** function.

**Note**: You can also use the *magnifying glass icon* with the two arrows or chevrons. 

You should now see the **full network** below:



## Section 2. Visualizing Expression Data on Networks

Probably the most common use of expression data in Cytoscape is to set the visual properties of the nodes in a network according to expression data.

This creates a powerful visualization, portraying functional relation and experimental response at the same time. Here, we will show an example of doing this.

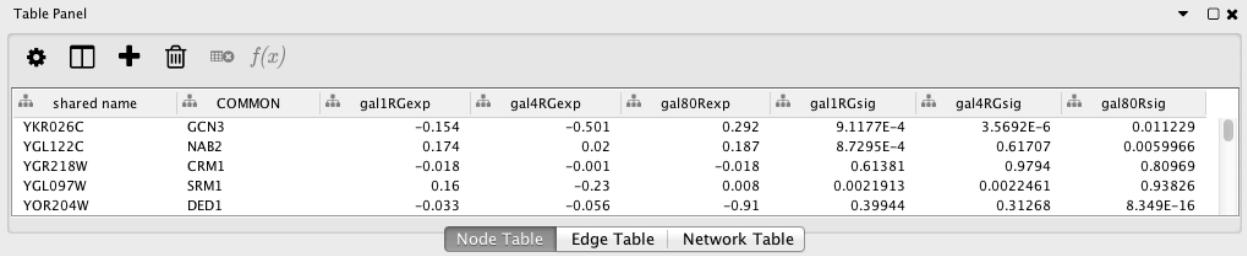
The data used in this example is from yeast, and the genes **Gal1**, **Gal4**, and **Gal80** are all yeast transcription factors. The experiments all involve some perturbation of these transcription factor genes.

Gal1, Gal4, and Gal80 are also represented in the interaction network, where they are labeled according to yeast locus tags: **Gal1 corresponds to YBR020W**, **Gal4 to YPL248C**, and **Gal80 to YML051W**.

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**.

The experimental data is visible in the **Node Table**:



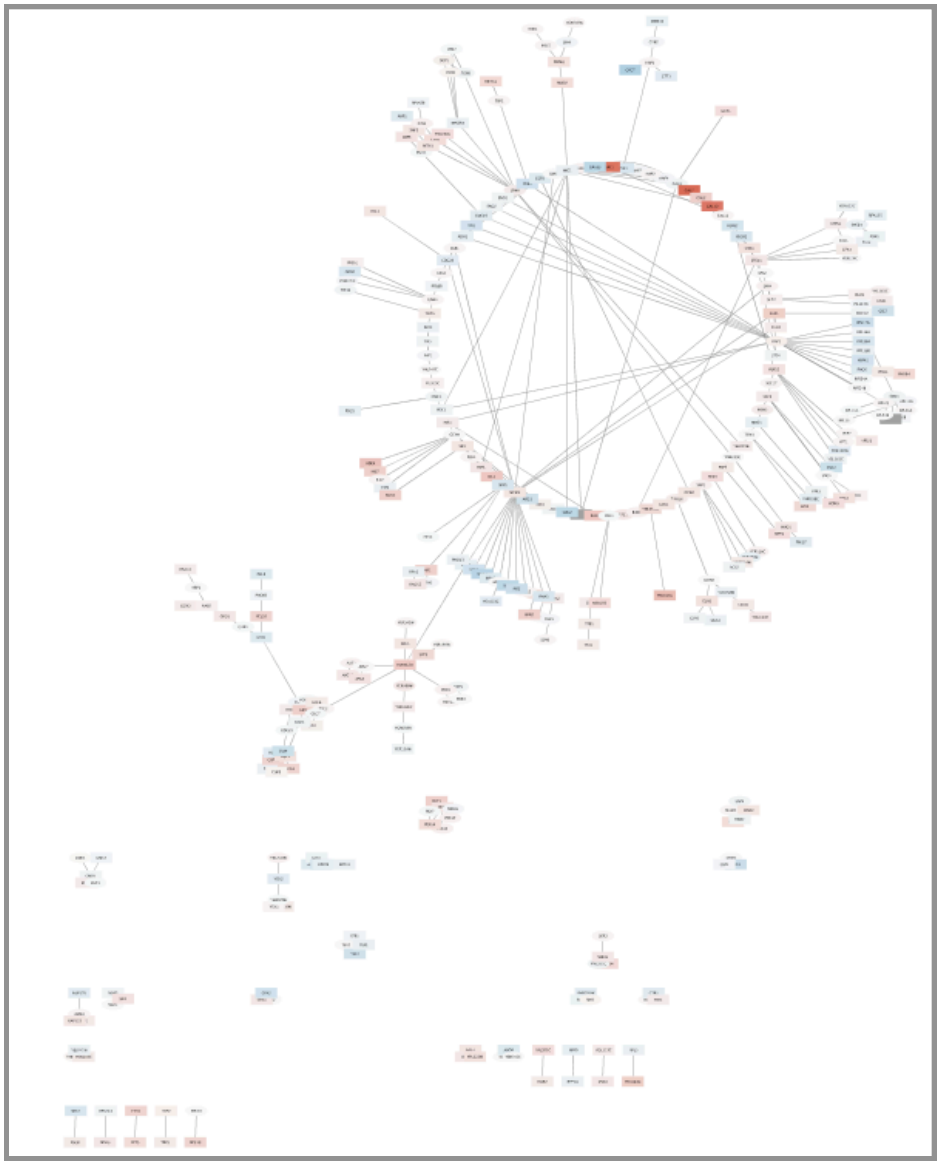

Selecting one or more nodes in the network will update the table to show only the corresponding row(s).

### *2.1. Changing Layouts*

By default, the "*Prefuse Force Directed Layout*" is applied to organize the layout of the nodes.

Let's change the layout to "*Circular Layout"* by selecting **Layout > Circular Layout**.

The network should now look like this:



Cytoscape supports many different layout algorithms. Have a play with different options under Layout and see how it alters the network display.

The available layout options are described in detail in the Cytoscape manual < http://manual.cytoscape.org/en/stable/Navigation_and_Layout.html?highlight=layout#automatic-layout-algorithms >

**Section 3. Filter Interactions**
Our network contains a combination of protein-protein (pp) and protein-DNA (pd) interactions. Next, we will filter out the protein-protein interactions to focus on the protein-DNA interactions.

First, let's select all protein-protein interactions.

- Go to the **Select** tab in the Cytoscape **Control Panel** (the leftmost panel).
- Click on the **+** and create a **Column Filter**.
- Click on **Choose column**... and select **Edge: interaction**
- Now in the text field enter **pp**.

This will select all of the edges with the protein-protein interactions.

If you zoom in on your network you will see some of your edges (the selected ones) in red and some in unselected gray color. You should see that many edges in the network are selected (i.e., colored red).

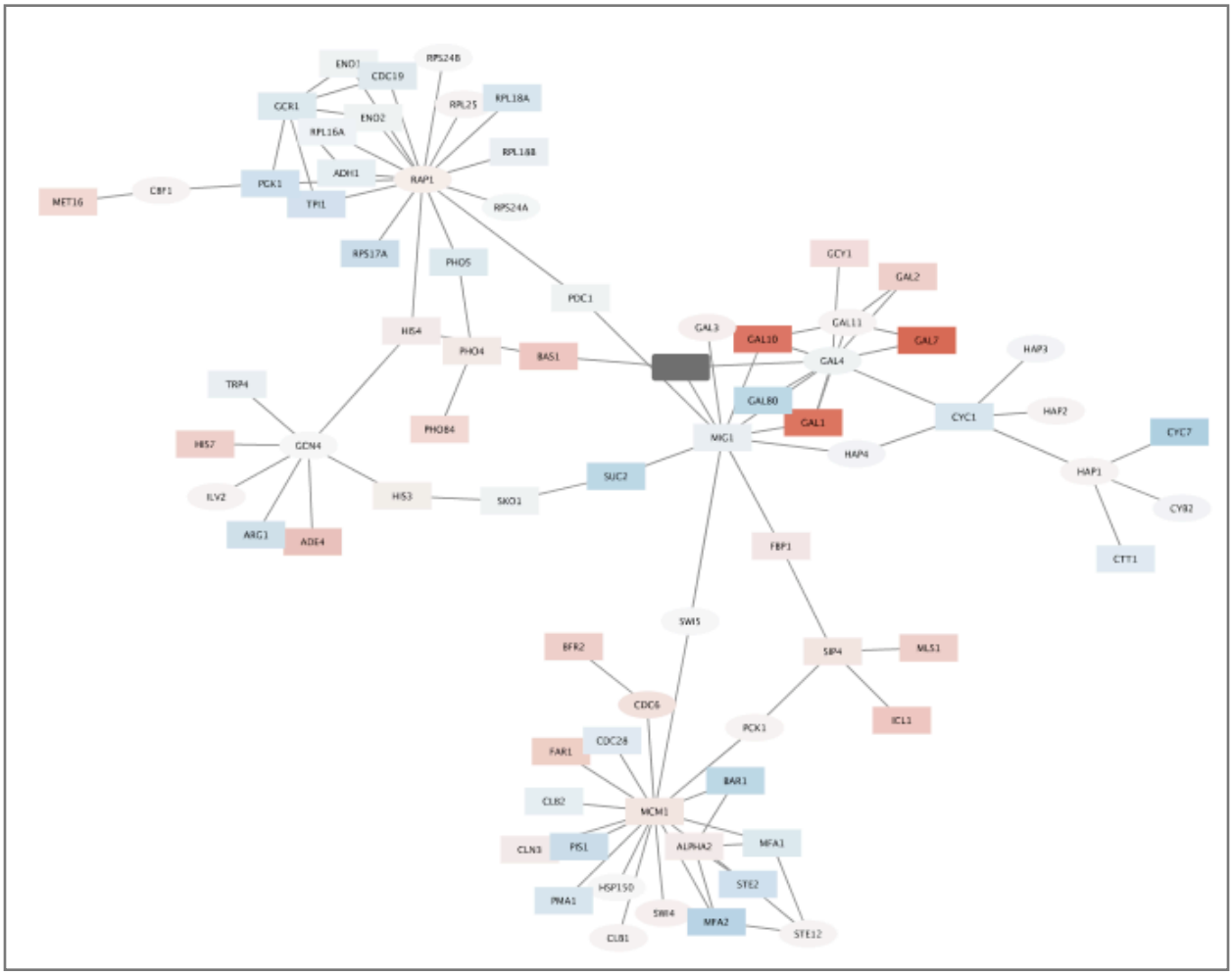
- Since we're only interested in the *protein-DNA edges*, we can delete the protein-protein edges we've just selected.
- To delete selected nodes or edges, you can use the menu **Edit > Delete Selected Nodes and Edges**
- You should now see many unconnected nodes in the network.
- Now, lets clean up the network by applying a force-directed layout under **Layout > Prefuse Force Directed Layout**.
- The largest component of the final filtered and cleaned up network should look like this:

*3.2 Observe the network*
Zoom into this portion of the network and find the three dark red (*i.e. highly induced*) nodes. Notice that these nodes are in the same region of the graph. Also notice:

- Notice that there are two nodes that interact with all three red nodes: **GAL4** (YPL248C) and **GAL11** (YOL051W).

**Note**. You can click on any node and drag it around to more clearly see it's edge associations.

Let's select these two nodes and their immediate neighbors.
- First select **GAL4** (YPL248C) by clicking on it.
- Now, extend the selection by holding down the **shift key** and clicking on **GAL11** (YOL051W).



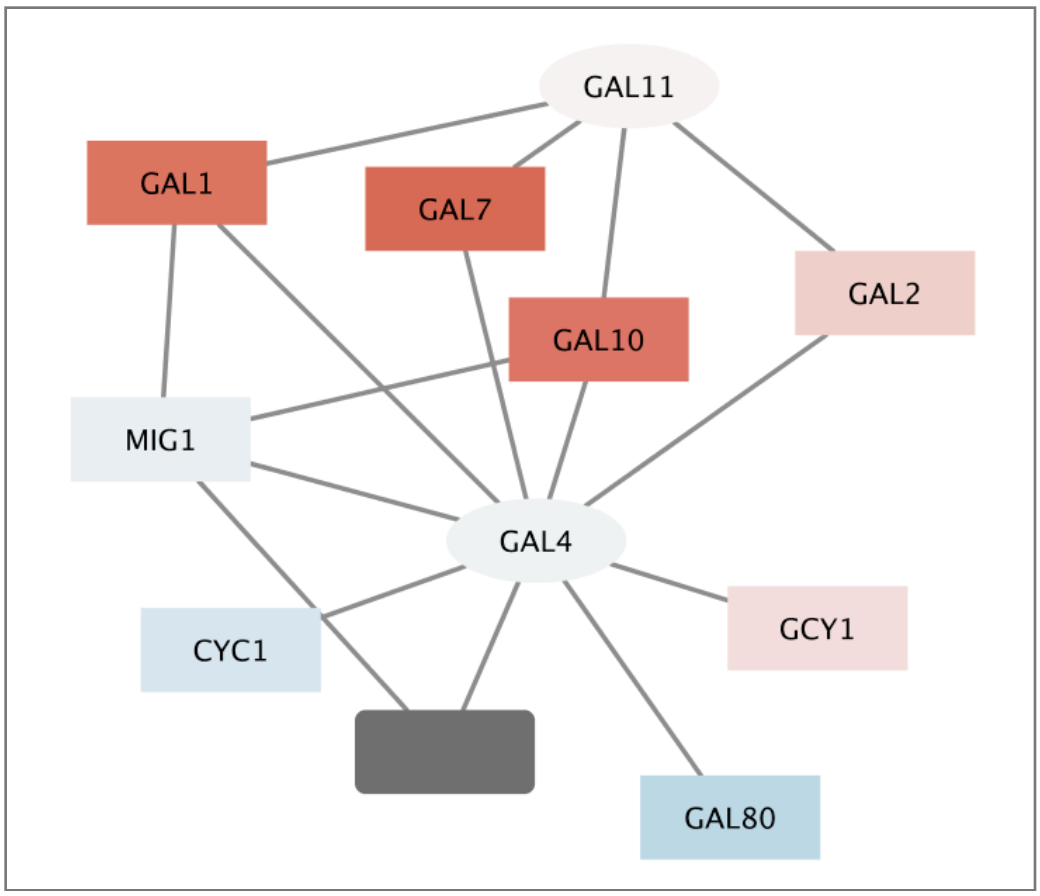

- Finally, select their neighbors by pressing ⌘**-6** on a Mac (or **control-6** on PC), or by the menu **Select > Nodes > First Neighbors of Selected Nodes > Undirected**

It is sometimes useful to create a new network from selected nodes.

- Create a new network by selecting **File > New > Network > Selected nodes, all edges**.

**Note**. Even though we say all edges, only the edges connecting the selected nodes will actually be used.)

As before, we may need to do a **View > Fit Content**. With some layout and zooming, this new network should appear similar to the one shown:




Digging into the biology of this network, it turns out that **GAL4** is repressed by **GAL80**. We will learn more about this below.

*3.3 Exploring Nodes*

Right-click on the node **GAL4** (or Control click if you don't have a multi button mouse).

- Select the menu **External Links > Sequences and Proteins > Entrez Gene**.
- This action will pop-up a browser window and search the Entrez Gene database for the term "**YPL248C**", the id of the node.
- In the results in the browser the first entry should be labeled **GAL4**. Click on this entry.

**Note**: The description of GAL4 (under the section labeled "*General protein information*" on the NCBI Gene page) tells us that GAL4 is "*repressed by GAL80*".

Our data show precisely this:

- Both nodes (**GAL4** and **GAL11**) show fairly small changes in expression, and neither change is statistically significant: they are rendered as light-colored circles.
- These slight changes in expression suggest that the critical change affecting the red nodes (**GAL1**, **GAL7** and **GAL10**) might be somewhere else in the network, and not due to either **GAL4** or **GAL11**.
- Note that our network shows that **GAL4** interacts with **GAL80** (YML051W), which shows a significant level of repression: it is depicted as a red square.
- Note that while **GAL80** shows evidence of significant *repression*, most nodes interacting with **GAL4** show significant levels of *induction*: they are rendered as red rectangles.
- **GAL11** is a general transcription co-factor with many interactions.

***Note***: Putting all of this together, we see that the transcriptional activation activity of **Gal4** is repressed by **Gal80**. So, repression of **Gal80** increases the transcriptional activation activity of **Gal4**. Even though the expression of **Gal4** itself did not change much, the **Gal4** transcripts were much more likely to be active transcription factors when **Gal80** was repressed.

Collectively, this explains why there is so much up-regulation in the vicinity of **Gal4**.

**Section 4. Saving results and dealing with memory issues**
Cytoscape provides a number of ways to save results and visualizations:

- As a session: **File > Save As…**
- As an image: **File > Export as Image…**
- To the web: **File > Export as Web Page…**
- As a graph format file for reading into other software such as R: **File > Export > Network**.

*4.2 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 to avoid crashes.

Destroy/delete views when you don't need them. Save at regular intervals and restart occasionally.

*Note for future versions*: Consider changing this lab to use this input data? http://manual.cytoscape.org/en/stable/Basic_Expression_Analysis_Tutorial.html#basic-expression-analysis-tutorial