BIMM143: INTRODUCTION TO BIOINFORMATICS (Lecture 17)

Biological Network Analysis (Part 1)
http://thegrantlab.org/bimm143/
Dr. Barry Grant

Overview: In this first of two hands-on sessions we will introduce Cytoscape < http://cytoscape.org >; an open source software platform for 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 lab data files: galFiltered.sif and galExpData.csv

After launching Cytoscape a new window will appear that looks like this:
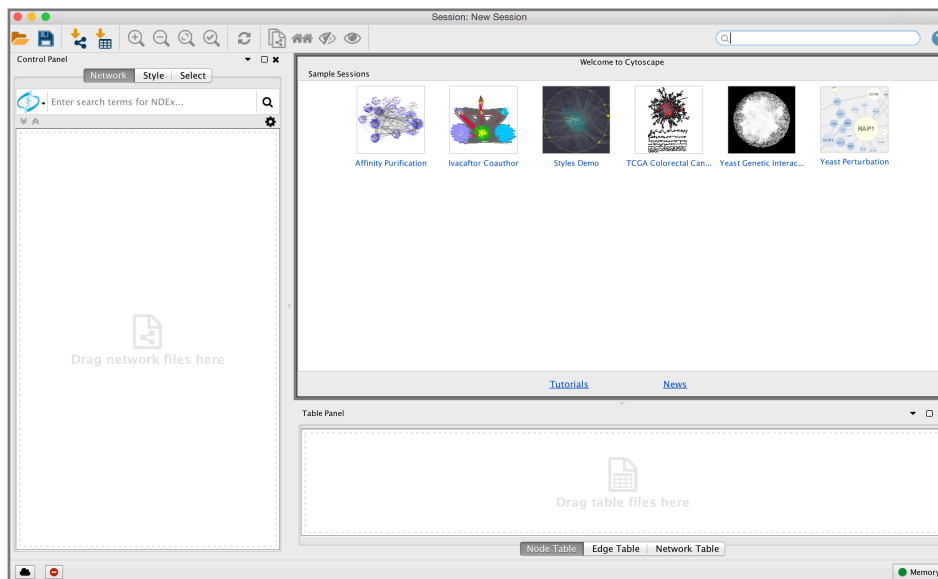
- After starting Cytoscape load the downloaded network file **galFiltered.sif** by clicking **File > Import > Network > File…**.
- By default, the "*Prefuse Force Directed Layout*" is applied to organize the layout of the nodes.
- Cytoscape should now look similar to the following figure:
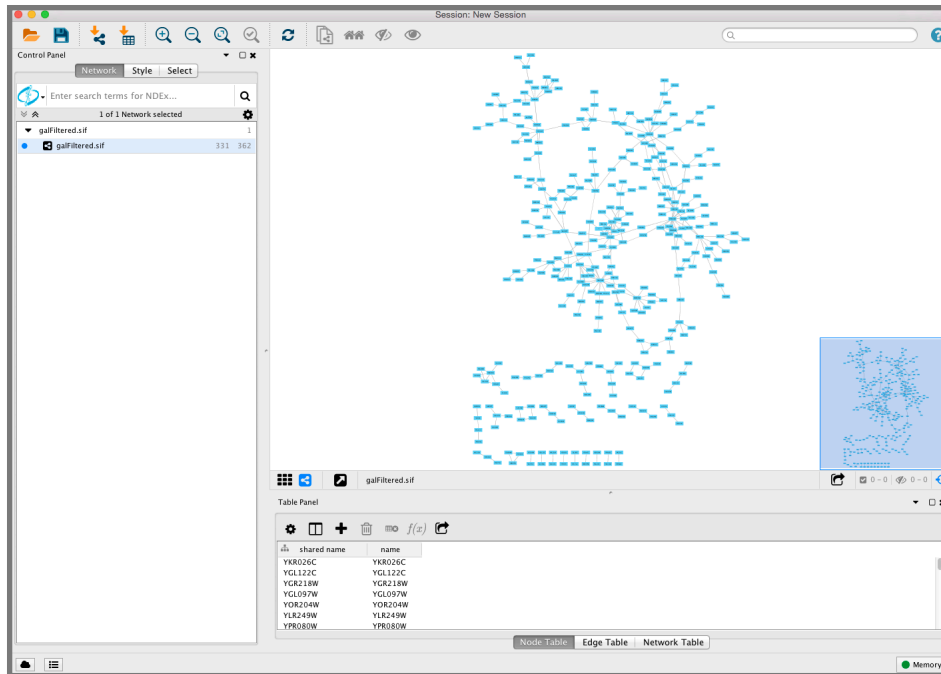
Note: When the network first opens, the entire network may not be visible because of the default zoom factor used. To see the whole network, we can use the **View > Fit Content** function.

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

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**.

2.1 Loading Expression Data

Using your favorite text editor, open the file **galExpData.csv**. The first few lines of the file are as follows:

```
GENE,COMMON,gal1RGexp,gal4RGexp,gal80Rexp,gal1RGsig,gal4RGsig,gal80Rsig
YHR051W,COX6,-0.034,0.111,-0.304,3.75720e-01,1.56240e-02,7.91340e-06
YHR124W,NDT80,-0.090,0.007,-0.348,2.71460e-01,9.64330e-01,3.44760e-01
YKL181W,PRS1,-0.167,-0.233,0.112,6.27120e-03,7.89400e-04,1.44060e-01
```

You should note the following information about the file:
1. The first line consists of labels.
2. All columns are separated by a single comma character (i.e. it is a CSV file!).
3. The first column contains node names, and must match the names of the nodes in your network exactly!
4. The second column contains common locus names. This column is optional, and the data is not currently used by Cytoscape, but including this column makes the format consistent with the output of many expression analysis packages, and makes the file easier to read.
5. The remaining columns contain experimental data, two columns per experiment (one column represents the expression measurement and the second represents the significance value for that measurement), and one line per node. In this case, there are three expression results per node.

- Under the File menu, select **Import → Table → File…**.

- Select the downloaded file **galExpData.csv**.

- The default settings are usually correct, although in some instances you may need to change the **Key** column to indicate which column will be used to match with the network key column.

- Click the **OK** button to import the data.

2.2. Setting node attributes

Now lets set the visual attributes of the nodes (labels and colors) in our network according to the expression data we have just imported.

Label the Nodes
- Open the **Styles** by selecting its tab in the "Control Panel" (the leftmost panel - see figure red annotation box #1).
- We want to use the "Common" name attribute to give the nodes useful names.
- First zoom in on the network so that node labels are visible.
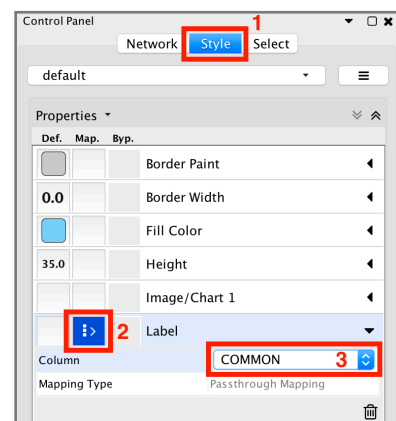
- Click the second column (the middle square labeled **Map.**) of the "Label" row in the **Styles** panel (see above figure annotation box #2). This should produce a drop-down panel with **Column** and **Mapping Type**.
- Change the "Column" to **COMMON** by clicking on the field to the right of the **Column** label. This should bring up a list of columns. Select **COMMON** (see previous figure annotation #3).
- Verify that the node labels on the network have changed to their common names.

Color the nodes
Define the node color of this visual style:
- Click on the middle square (**Map.**) next to the **Fill Color** row in the **Styles** panel (see figure annotation box #1).
- Click the **– select value –** cell in the **Column** section.
- This will produce a drop-down menu of available column names. Select "**gal80Rexp**" (see figure annotation box #2).
- Click the **– select value–** cell in the **Mapping Type** section.
- This will produce a drop-down menu of available mapping types. Select **Continuous Mapping**.
- This action will produce a basic blue to red color gradient.
- [OPTIONAL] Click on the color gradient to change the colors and set different upper and lower limits if you wish.
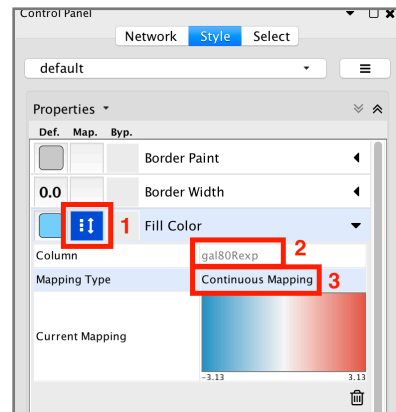

Note: Annoyingly the default node color of pale blue falls within this spectrum. A useful trick is to choose a color outside this spectrum to distinguish nodes with no defined expression value and those with slight repression. To do this click the **Def**. (leftmost) square next to **Fill Color** and choose a dark gray color. Zoom out on the network view to verify that a few nodes have been colored gray.

Set the Node Shape
We imported both expression measurement values and significance values for those measurements. We can use the significance values to change the shape of the nodes so that measurements we have confidence in appear as squares while potentially bad measurements appear as circles.

- Click the **Map**. cell next to the **Shape** row in the **Style** panel.
- Click the **– select value –** cell next to **Column**.
- This will produce a drop-down menu of available column names. Select "**gal80Rsig**".
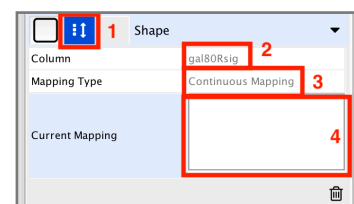
- Click the **– select value –** cell next to **Mapping Type**.
- This will produce a drop-down menu of available mapping types. Select **Continuous Mapping**.
- This will create an empty icon in the **Current Mapping** row of the **Shape** section. Double-click on this icon.

- This action will pop-up a new window entitled "*continuous mapping editor for node shape*" (see figure right).
- Click the **Add** button.
- This action will split the range of values with a slider down the middle with a node shape icon to either side of the slider.
- Double-click on the left node icon (a circle).
- This will pop-up a node shape selection dialog.
- Choose the **Rectangle** shape and click the **Apply** button.
- The continuous shape selection dialog should now show both a square and a circle node shape icon.
- Click on the black triangle and move the slider to the left, move it to the lower value of **0.05**, our threshold for significance.
- Close the continuous shape selection dialog and verify that some nodes have a square shape and some nodes have a circular shape.
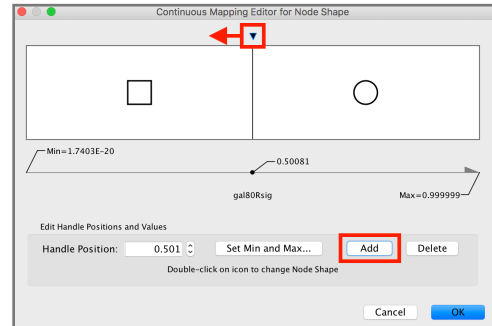
Note that the experimental data we imported is visible in the **Node Table**:

| shared name	COMMON	gal1RGexp	gal4RGexp	gal80Rexp	gal1RGsig	gal4RGsig	gal80Rsig
YKR026C	GCN3	−0.154	−0.501	0.292	9.1177E−4	3.5692E−6	0.011229
YGL122C	NAB2	0.174	0.02	0.187	8.7295E−4	0.61707	0.0059966
YGR218W	CRM1	−0.018	−0.001	−0.018	0.61381	0.9794	0.80969
YGL097W	SRM1	0.16	−0.23	0.008	0.0021913	0.0022461	0.93826
YOR204W	DED1	−0.033	−0.056	−0.91	0.39944	0.31268	8.349E−16

Node Table | Edge Table | Network Table

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

2.3. 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 two different network layout results are shown below. Note that Cytoscape supports many different layout algorithms. Have a play with different options under **Layout** and see how it alters the network display.
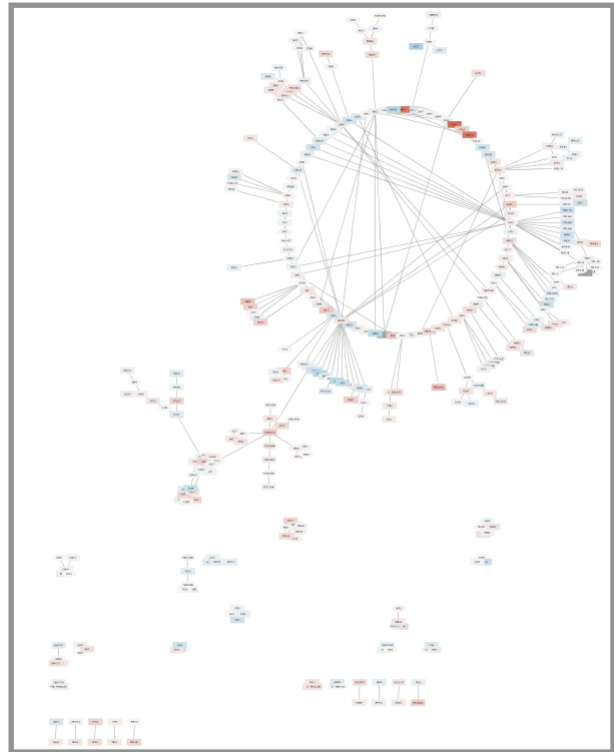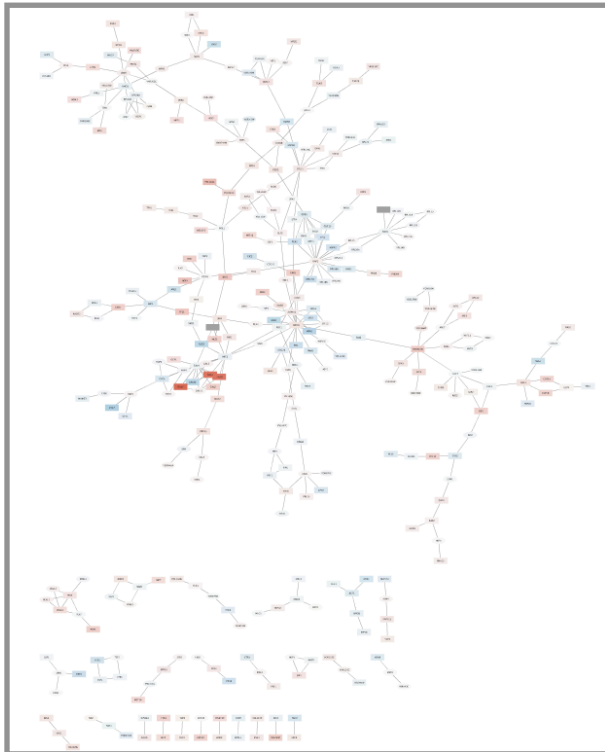
Note: The available layout options are described in detail in the Cytoscape manual
< http://manual.cytoscape.org/en/stable/Navigation_and_Layout.html?
highlight=layout#automatic-layout-algorithms >

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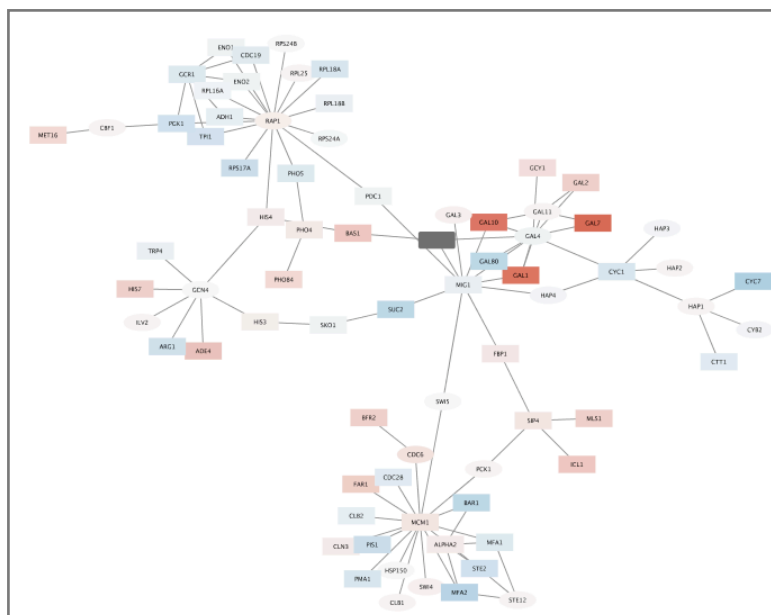
3.2 Observe the network
Zoom into this portion of the network and find the three dark red (*i.e. highly induced*) nodes. Notice that these nodes are in the same region of the graph. Also:

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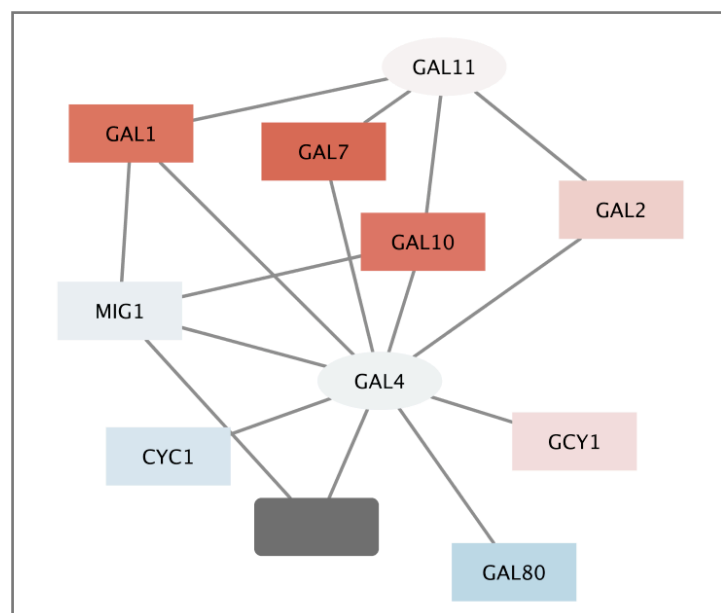
3.3 Exploring Nodes
Right-click on the node **GAL4** (or Control click if you don't have a multi button mouse).

- 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 blue rectangle.

- 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. <u>Saving results and dealing with memory issues</u>
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 a saved network with visual settings from sections 1-3 is available here: :
CytoscapeDemo_01.cys