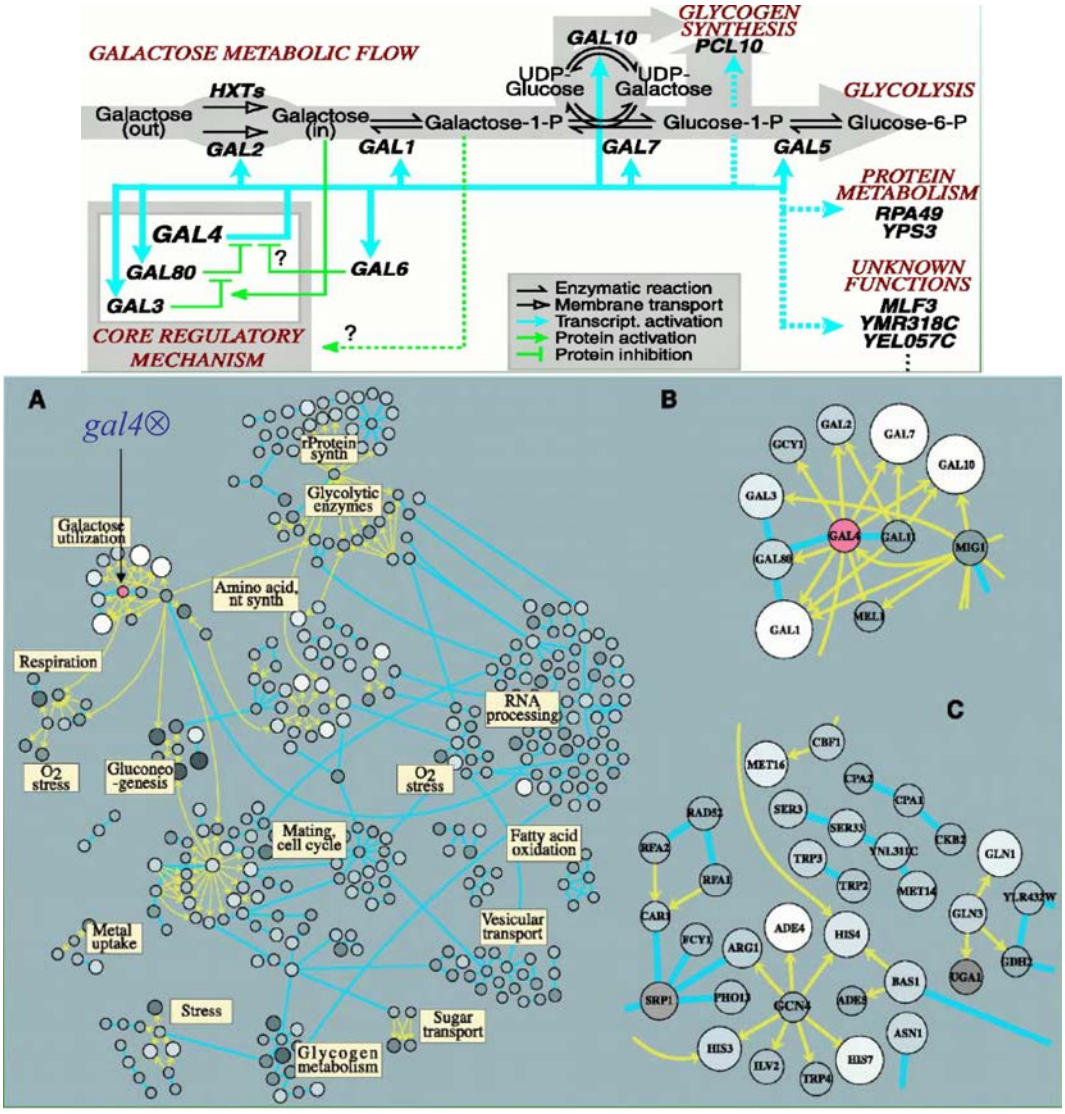
Worksheet 3

Data Integration with a Network

Objective: In this exercise we will integrate gene expression data from gene deletion studies with protein-protein interaction network.

In the study by Ideker *et al*. in Science 2001, the yeast transcription factors Gal1, Gal4, and Gal80 were analyzed for their importance in galactose utilization pathways.



A) Loading Network into Cytoscape

STEP 1: Start Cytoscape 3, in File select ‘Import’ a ‘Network’ from “File’ to load a network structure galFiltered.sif in the simple interaction format (sif).

This network will contain a combination of protein-protein (pp) and protein-DNA (pd) interactions.

B) Loading Gene Expression Data into Cytoscape

STEP 2: Under the File menu, select 'Import' to input a 'Table' data in a 'File' galExpData.pvals. This file contains gene expression measurements for three pertubation experiments. In each experiment, the level of one key protein was perturbed artificially.

After a brief load, a status window will appear. In the 'Table Panel', you can examine the experimental data columns (three) for each conditions and the type of significance values included

STEP 3: Now we will use the Node Attribute Browser to custom-browse through the expression data, as follows.

i. In the Cytoscape canvas, select any node by clicking on the node.
ii. In the Table Pannel, hit the 'Change Table Mode' and change it to 'Show selected'. Then hit the 'Show Column' button, and select the attributes of name, gal1RGexp, gal4RGexp, and gal80Rexp.
iii. Under the Table Pannel, you should see your nodes listed with their expression values.

C) Coloring Nodes

It is common to use expression data in Cytoscape to set the visual attributes of the nodes in a network. This visualization can be used to portray functional relation and experimental response at the same time. First, we will play with how to indicate increased or decreased expression of a gene. The steps for doing this are as follows:

STEP 4: Go to VizMapper to set the visual properties.

STEP 5: By clicking on the 'Option' and select the 'Copy Visual Style'. Name it as "Gal80" to duplicate the default style. Click on the Define button to edit your style.

STEP 6: The default tab defines the Node Color of this visual style. Double clicking the 'Node Fill Color' and select 'RGB'.

i. Under Visual Mapping Browser, double click the node fill color, select the attribute "Gal80RGexp". In the 'Mapping Type', select 'Continuous Mapping'.
ii. This specifies that each node will be colored on a color continuum according to Gal80 expression, as follows: Large negative values (indicating high repression) are colored black. Large positive values (indicating high induction) are colored white.
iii. Finally, click on 'Show All'. You should see most nodes colored black, grap, and white.

D) Using *p*-values

Here, we will explore an example of using expression values and p-values together in setting visual properties.

STEP 7: Select some nodes at random, and look at their expression values and *p*-values under the Node Attribute Browser. Notice how the expression data value ranges from about -3 to +3 in these cases, the *p*-value ranges from 0 to 1, as they should.

STEP 8: Now, we will explore setting node shapes according to *p*-values.

i. Double click the Node Width tab, select gal80RGsig and in the 'Mapping Type', select 'Continuous Mapping'. Double click the Node Height tab, select gal80RGsig and in the 'Mapping Type', select 'Continuous Mapping'.
ii. Double click on the 'Current Mapping' will bring you to the Continuous Mapping Editor. You can change the minimum and maximum node sizes to match with the range of *p*-value.
iii. Click on 'Show All'. On your Cytoscape canvas, your node sizes should change. This will have the effect of depicting nodes with significant *p*-values as large circles. Changes in larger nodes are more likely to be significant than changes in smaller ones.

E) Biological Analysis Scenario

This section presents one scenario on how expression data can be combined with network data to tell a biological story.

STEP 9: Select the neighborhood of GAL4.

i. Enter 'YPL248C' (this is the GAL4 gene) into the selection box and hit return.
ii. Select -> Nodes-> First neighbors of selected node (this gets the immediate neighbors of GAL4).
iii. Hit 'New Network from Selection' button.
iv. In the new sub-network, apply a graph layout algorithm using the yFiles Hierarchic layout.

Notice that all three black (highly induced) nodes are in the same region of the graph. With a little exploration in the node attribute browser, you should see the following:

i. The two nodes that interact with all three black nodes are GAL11, a general transcription cofactor with many interactions) and GAL4.
ii. Both nodes show fairly small changes in expression, and neither change is statistically significant. These slight changes in expression suggest that the critical change affecting the black nodes might be somewhere else in the network.
iii. GAL4 interacts with GAL80, which shows a significant level of repression.
iv. Note that while GAL80 shows evidence of significant repression, most nodes interacting with GAL4 show significant levels of induction.

STEP 10: Go to the NCBI website (http://www.ncbi.nlm.nih.gov/), and search the Gene database for "YPL248C" (another name for the GAL4 gene). The items returned should include Gal4. Click on the link for Gal4 to get more information.

Q: Is Gal80 activating or inhibiting the activity of Gal4? Does this make sense given your subgraph? Explain your reasoning in a sentence or two.