MSc Applied Bioinformatics 2024-2025
Data Integration and Interaction Networks

Dr Tomasz Kurowski
t.j.kurowski@cranfield.ac.uk

Introduction to Cytoscape practical – Part I

Working with existing data

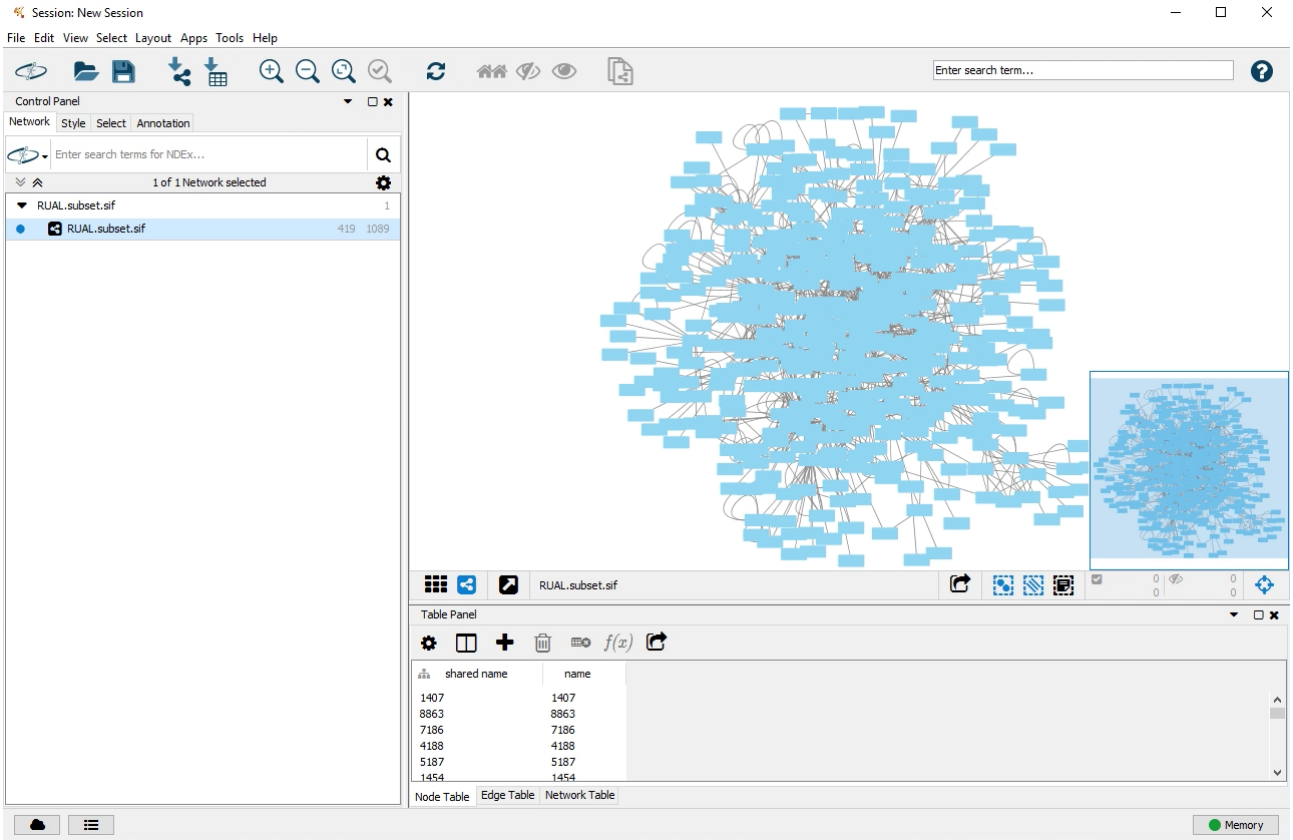
Cytoscape is available for download at **https://cytoscape.org/**. In case you are unable to install it (e.g., you do not have admin rights), an "Archive Distribution", which does not require installation, can also be downloaded from the website.

Launch Cytoscape. For the first exercise, we will look at some interaction data in the *RUAL.subset.sif* dataset (available on Canvas), a portion of a human protein-protein interaction dataset published 2005 Oct 20 in *Rual et al.*, Nature 437(7062):1173-8.

Load the *RUAL.subset.sif* network file into Cytoscape by selecting **File → Import → Network from file...**, then specify the location of the file you just downloaded from Canvas. You may be prompted with a window to create a new Network - just click OK.

This network consists of 1089 interactions observed between 419 human proteins, and is a small subset of a large human interaction dataset. This subset of interactions consists of proteins that interact with the transcription factor protein TP53.

After the file is loaded the network will be listed in the Control Panel on the left. When selected, the network will be displayed in the main window on the right (the canvas) using a specific type of layout. By default, this is likely to be the "Prefuse Force Directed Layout" as shown below.



The layout can be freely modified by dragging the nodes around or selecting a new layout form the **Layout** menu. Try some layouts out now.

Different types of layouts may be more or less appropriate for different types of data, specific datasets, or types of analysis, as they will help elucidate different characteristics of the network to a various extent. The Prefuse Force Directed Layout will work well for the data used in this practical.

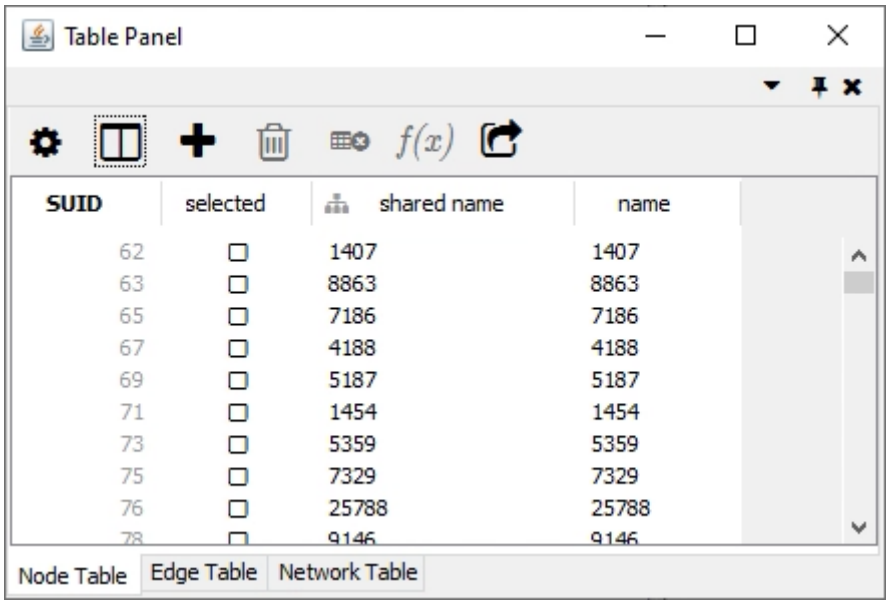
Basic Feature Selection

On the Cytoscape canvas on the right, you can select single nodes by clicking on them with the mouse, or multiple nodes by holding down Shift/Control together with the left mouse button and selecting an area. Edges can be selected in a similar way, either by clicking them individually or with an area selection which will select all edges which intersect the area.

When you want to avoid selecting both nodes and edges at the same time, you can set your preferred selection type by going to **Select → Mouse Drag Selects**.

Cytopanel navigation

Below the canvas there is the **Table Panel**, which displays the characteristics of selected nodes and/or edges (or all of them if none are selected). You can click the Float Window button in the upper right of the panel in order to use it as a separate resizeable window, which can make the tables more readable. You can dock the window back in interface by pressing the pin icon.

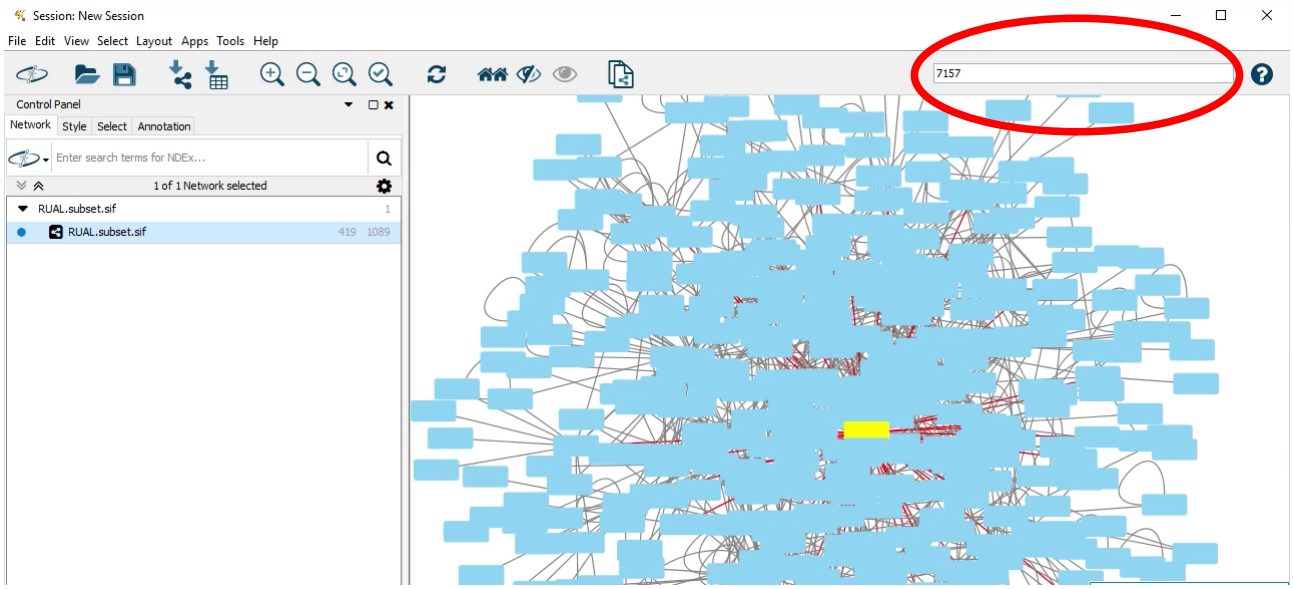


You can choose how many table columns you want to display by clicking on the "Show Column" icon (2nd icon from the top) in the Table Panel.

To view a centred overview of your network in your canvas, click on the Fit Content zoom icon in the top menu bar. The other similar icons can be used to zoom out, zoom in, and zoom to a selection respectively.

More on Node Selection

The nodes in this network are identified by numeric Entrez IDs. The node representing TP53 is numbered 7157. Select this node by putting 7157 in the search field at the top frame, and press Enter.



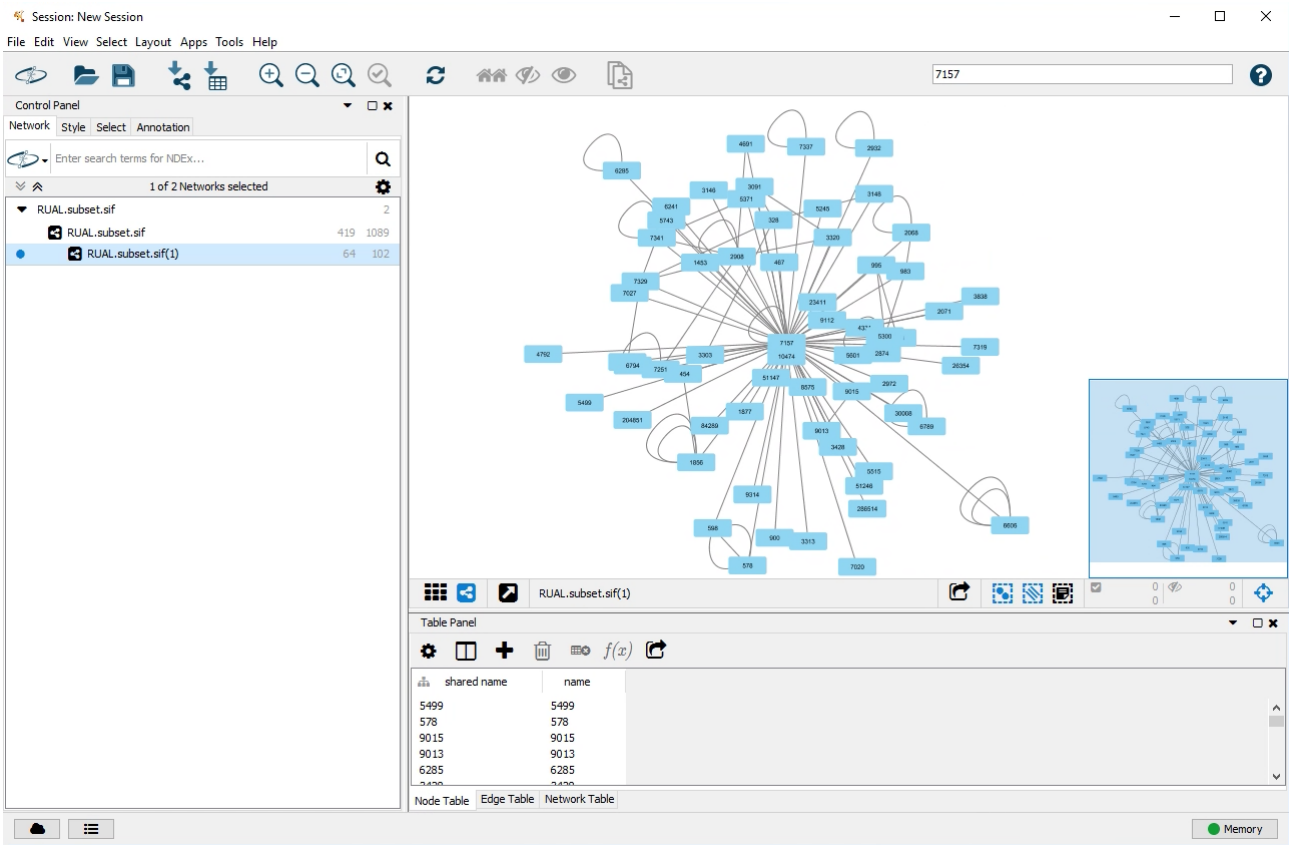
The TP53 node at the centre of your network should turn yellow, i.e. become selected.

Select the nodes that interact directly with TP53 by going to **Select → Nodes → First neighbors of selected nodes → Undirected**. You should see a sub-network with several yellow nodes in the centre, as shown below. At the bottom of the canvas (highlighted in the image below) you can see that you have selected 64 nodes and 64 edges. Copy the selected nodes and their edges into a separate network by selecting **File → New Network → From selected nodes, all edges**.

Clean up your canvas:

1. Make sure your subnetwork window is maximised.
2. Use the **Zoom out to display all of current Network** icon to zoom into this network.

Your display should now appear as shown below:



Loading Node Attributes

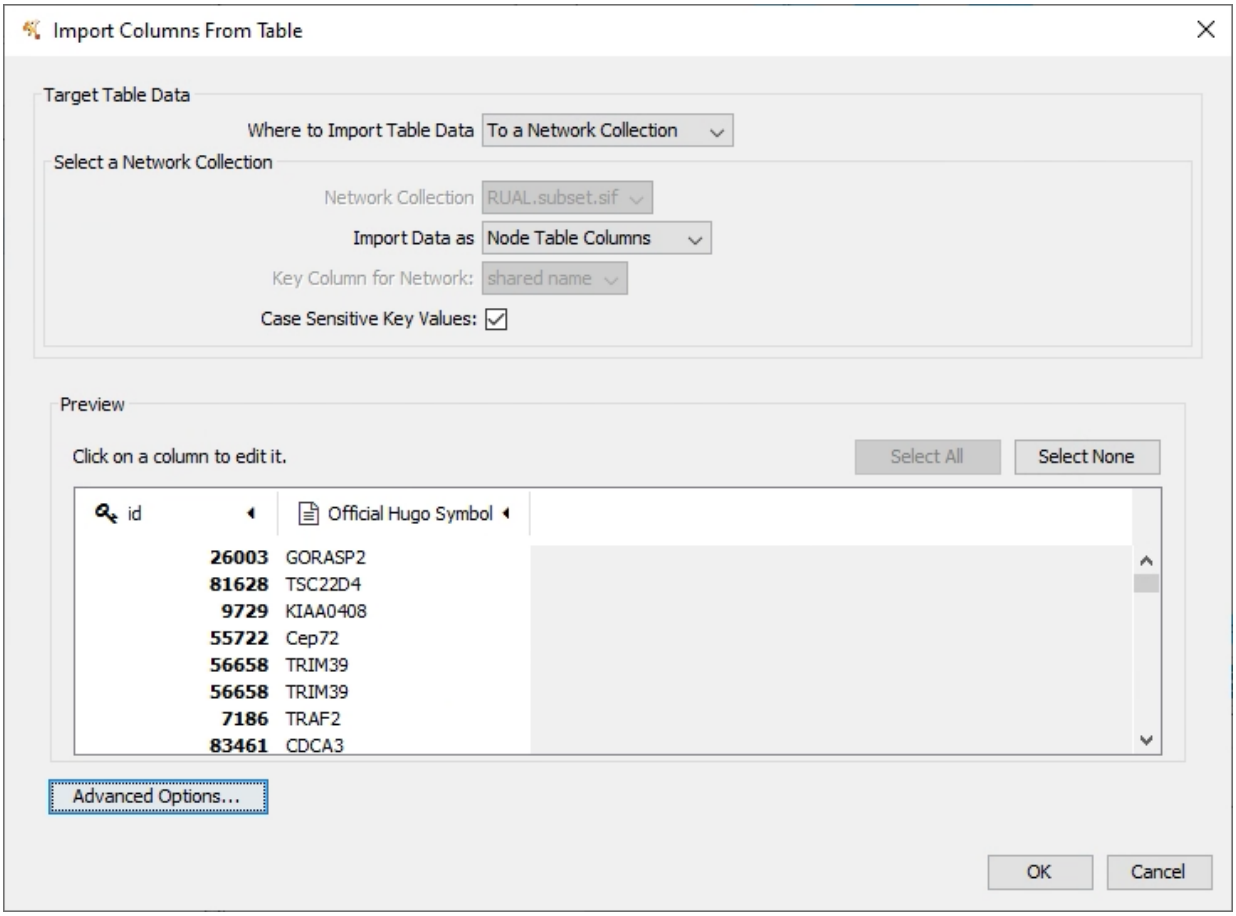
The nodes of this network are identified by numeric IDs. Looking closely at these nodes, you will see their numeric labels. A companion attribute file maps these numeric IDs to standard gene names. Load this attribute file as follows:

1. Go to **File → Import → Table from File**...
2. Select the file you downloaded from Canvas: **RUAL.csv**, and click **Open**.

3. A popup window labelled **Import Column from Table** will appear. Configure the file import options in **Advanced Options**, selecting the appropriate delimiter character so that the file contents are split into columns as shown below.

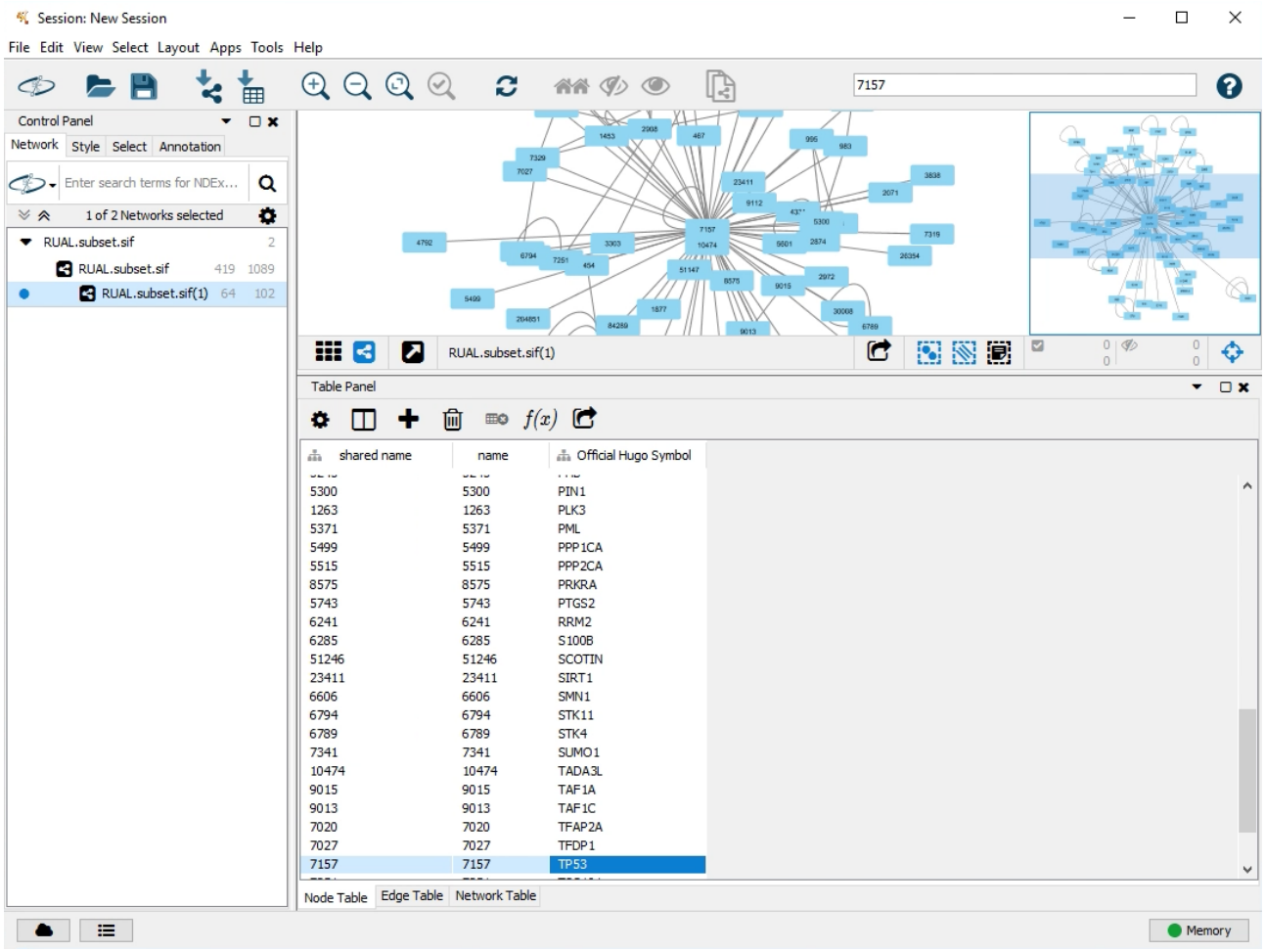
4. Make sure the first column which contains the identifier is marked as a Key column, while the second (which you should rename to „Official Hugo Symbol") is marked an an Attribute column.

5. Click Ok to import the data.



Now you can view the added attributes in Cytoscape. They should be visible as a new column in the Table View for nodes, as shown below with TP53.

Dr Tomasz Kurowski
t.j.kurowski@cranfield.ac.uk



Defining visual styles

We will now modify the visual style of our network to display the HUGO code names in the canvas view.

Open the Style tab in the Control Panel. You will see a drop down menu allowing you to select a visual style for your network visualisation (**default** initially), a list of modifiable properties, and a tab selection at the bottom allowing you to set properties for Nodes, Edges, and the Network separately.

Notice that you can switch quickly between visual styles using the **Current Style** pull-down menu. Switch between some of the pre-defined visual properties and watch how the canvas changes.

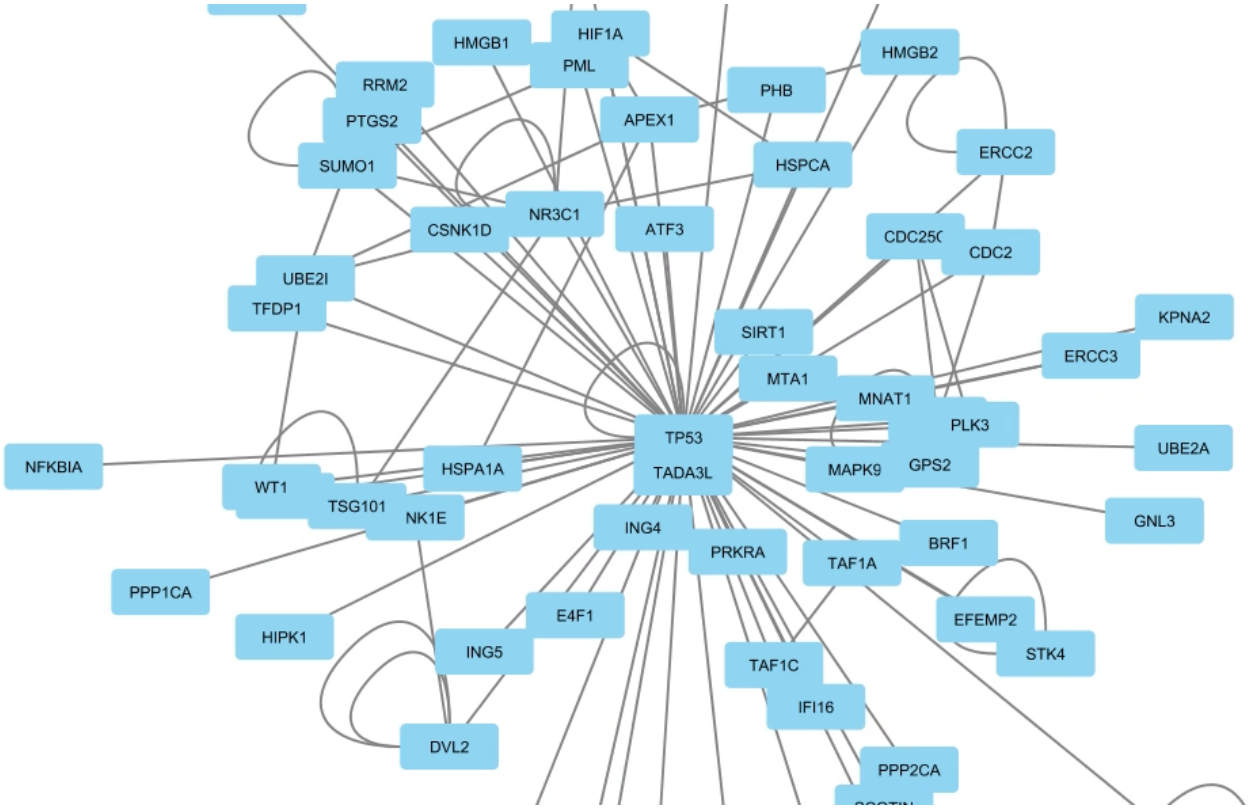
Now we will define our own visual style.

1 Click the button next to the Current Style drop-down menu and click "**Copy style...**". Name this new style **Class 1**.

2 Open the **Label** tab in your Class 1 style.

3 Use the drop-down list in the tab to select Official HUGO Symbol as the column used for labelling nodes.

The nodes should now be labelled with their HUGO symbols as shown below.



Edges

This dataset contains many types of edges: some representing experimentally-determined interactions (Y2H and coAP, from yeast two-hybrid and co-immunoprecipitation respectively), and some obtained from the literature (non_core, core, and hyper_core, corresponding to low, moderate, and high confidence literature search results). We will change the look of each edge by its interaction type.

1. Go back to the Style tab in the Control Panel to further define your Class 1 visual style.
2. Select the **Edge** tab at the bottom to modify the visual style of edges.
3. Find and open the **Stroke Color (Unselected)** tab.
4. Select the **interaction** column as the column used.
5. Select **Discrete Mapping** as the **Mapping Type**. A list should now appear listing the types of interactions in this network, including Y2H, coAP, core, hyper_core, non_core.
6. Click the space next to Y2H and then click the small icon with three dots that appears. Select a colour for the Y2H edges.
7. Repeat for the other edge types. Select your colours so that the Y2H and coAP colours are similar (e.g. green and blue), and the core, hyper_core, and non_core colours are similar (e.g. orange, red, and pink). This will allow you to see if each edge was determined experimentally or through literature, and will further allow you to see the edge type.
8. The edges in the network should now be coloured by type. Which is the most common type of edge? Which is the least common?

To see details on specific edges, perform the following:

1. In the **Table Panel**, click on the **Edge** tab at the bottom.
2. In the main Cytoscape Window, under the **Select** menu, choose **Mouse Drag Selects → Edges Only**.
3. Select groups of edges in the network canvas.
4. Observe how the list of edges changes in the **Edge** table as you select additional edges.