Exercise #5

In this exercise, we will perform some simple queries to retrieve molecular networks based on a protein, a small-molecule compound, a disease, and a topic in PubMed.

5.1 Protein queries

Go to the menu **File → Import → Network from Public Databases**. In the import dialog, choose **STRING: protein query** as **Data Source** and type your favorite protein into the **Enter protein names or identifiers** field (e.g. SORCS2). You can select the appropriate organism by typing the name (e.g. *Homo sapiens*). The **Maximum number of interactors** determines how many interaction partners of your protein(s) of interest will be added to the network. By default, if you enter only one protein name, the resulting network will contain 10 additional interactors. If you enter more than one protein name, the network will contain only the interactions among these proteins, unless you explicitly ask for additional proteins.

Unless the name(s) you entered give unambiguous matches, a disambiguation dialog will be shown next. It lists all the matches that the stringApp finds for each query term and selects the first one for each. Select the right one(s) you meant and continue by pressing the **Import** button.

- *How many nodes are in the resulting network?*

- *How does this compare to the maximum number of interactors you specified? What types of information does the **Node Table** provide?*

5.2 Compound queries

Go to the menu **File → Import → Network from Public Databases**. In the import dialog, choose **STITCH: protein/compound query** as **Data Source** and type your favorite compound into the **Enter protein or compound names or identifiers** field (e.g. imatinib). You can select the organism and number of additional interactors just like for the protein query above, and the disambiguation dialog also works the same way.

- *How is this network different from the protein-only network with respect to node types and the information provided in the **Node Table**?*

5.3 Disease queries

Go to the menu **File → Import → Network from Public Databases**. In the import dialog, choose **STRING: disease query** as **Data Source** and type a disease of interest into the **Enter disease term** field (e.g. Alzheimer's disease). The stringApp will retrieve a STRING network for the top-N proteins (by default 100) associated with the disease.

The next dialog shows all the matches that the stringApp finds for your disease query and selects the first one. Make sure to select the intended disease before pressing the **Import** button to continue.

- *Which additional attribute column do you get in the **Node Table** for a disease query compared to a protein query? (Hint: check the last column.)*

5.4 PubMed queries

Go to the menu **File → Import → Network from Public Databases**. In the import dialog, choose **STRING: PubMed query** as **Data Source** and type query representing a topic of interest into the **PubMed Query** field (e.g. jet-lag). You can use any query that would work on the PubMed website, but it should obviously a topic with related genes or proteins. The stringApp will query PubMed for the abstracts, find the top-N proteins (by default 100) associated with these abstracts, and retrieve a STRING network for them.

- *Which attribute column do you get in the **Node Table** for a PubMed query compared to a disease query? (Hint: check the last columns.)*

5.5 New search interface

The types of queries described above can alternatively be performed through the new Cytoscape search interface. Click on the drop-down menu with an icon on it, located on the left side below the **Network** tab in the **Control Panel**. Select one of the four possible STRING queries and directly enter your query in the text field. To change settings such as organism, click the **option** button next to the text field. Finally, click the button of the **magnifying glass** to retrieve a STRING network for your query.