Exercise 3 – Network Propagation Online

Objectives:

- to learn about the "WebPropagate" server, and to use it to expand beyond a set of seed proteins, for a give functional process
- to apply network propagation to a poorly delineated cellular subsystem, as well as to a human disease process.

The WebPropagate service is a browser-based wrapper around standard network propagation algorithms, published a few years ago by the lab of Roded Sharan in Tel Aviv. Similar software can also be installed and run locally, but the webservice has the advantage of convenient use and it is fully pre-loaded with the relevant networks. WebPropagate is based on the latest network information from the BioGrid database.

1) characterize Telomere-binding proteins

- let's assume we are interested in the set of proteins that bind and protect telomeres (chromosome ends) in human cells.
- to collect a set of "seed proteins", point your browser to Wikipedia and search for the page on "shelterin" (use the English-language Wikipedia; shelterin is the name of a recently described telomere-binding complex).
- on the Wikipedia page, scroll a bit down to reach the section "Subunits". There, you will find six protein names listed, which together form the protein-complex. We will use these six proteins as seeds, but we will omit one of them: "POT1". We expect the network propagation algorithm to recover that protein; this is effectively a positive control.
- now, open a second tab in your browser and point it to the WebPropagate server: https://anat.cs.tau.ac.il/WebPropagate/
- now, let's enter five of the six proteins (omitting POT1) into the server. Under "Species", select "Homo sapiens", leave all other parameters as they are, and enter the five seed proteins into the box labeled "Seeds" (separated by spaces, on one line).
- then, click "Submit" and wait. The results will appear either to the right of the page, or directly below the input (depending on your browser); so you might need to scroll down to see them.
- a network should have appeared, which contains in green the five seed proteins we have entered, and in red a set of roughly 20 proteins produced by the network propagation. Notice how "POT1" is indeed among them, relatively well connected!
- now, click on one of the green seed proteins, and you should be shown how many proteins are annotated to be interacting with that particular protein according to BioGrid. Notice how this set of interactors is much larger (depending on which protein you clicked) than what has been selected by network propagation … this is the "added value"; the algorithm filters the relevant, tightly connected, "local" interactors from the large list of potential interactors. You can use shift-click on the same protein to remove again the layer of interactors just added.
-

2). characterize and expand a human disease set.

- let's try the same process with a set of human proteins known to be associated with a particular disease.
- point your browser to the DISEASES database: https://diseases.jensenlab.org/
- in the search box, enter "Parkinson", and select the first entry from the list of matching diseases (labelled "Parkinson's disease", DOID:14330).
- now, you will be presented with a list of the ten most likely disease-associated human proteins (from Text-mining thousands of papers).
- from this list, enter the proteins into the WebPropagate server, as before, but leave out "SNCA" (synuclein) as a positive control.
- when the results come back, SNCA should indeed be listed again among the significant local neighbors, if everything worked out well. Notice that it is the central-most neighbor in the layout.
- do any of the other proteins listed make any sense?

.

3). try a seed set of your own.

- now, having familiarized yourself with WebPropagate, go ahead and select a "seed" protein set of your own choosing. Perhaps you can try a different disease, or a set of proteins from a lab experiment?
- did you get a useful output? If not, any idea why one set of seeds might work, and another might not?