

Lab8: Network Properties & Visualization

Due by Friday May 31st, 2019 @ 12pm

The goal of this lab is to use Gephi to study some properties of various networks and methods of how to visualize them. We will mainly focus on biological network today. You will have an opportunity to import, manipulate and visualize some real biological networks. Through this lab, you should have a better understanding of: (i) basic data frame for the field of network science; (ii) network statistics interpretation and implementation; (iii) basic network visualization methods.

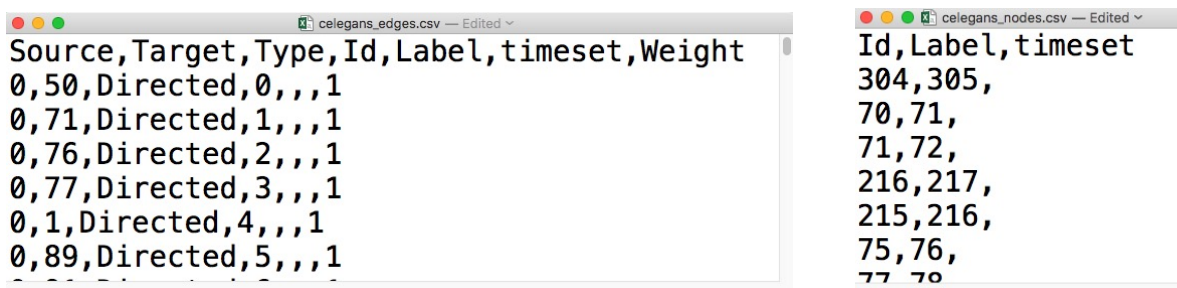
Here is a list of instructions (section 1-4) to start:

1 Importing data:

Often times, part of the difficulties in doing network analysis is actually getting data in the right or importable format. Here let's examine the typical format for network data and two methods to import into Gephi.

First, if you're lucky enough to get a pre-formatted network data file such as `.gdf` file or `.gexf` file, then you only need to open the file in Gephi and Gephi will automatically recognize the data frame and create all of the edges and nodes (and all of the pre-calculated attributes).

However, most times your network data will be in `.csv` file format, which is just a text file that has comma separated entries. One often need two `.csv` files to specify the network: `edges.csv` and `node.csv`.



As you can see from the csv file, for each edge, you need to specify what is source node and target node this edge (need two nodes to form an edge). Furthermore, you can also specify whether the edge is directed or undirected (Type), the ID & label for the edge, and the weight of the edge. Also if your network is dynamical (edges change in time), you can also include the timeset for the appearance of each edge in time.

For the node.csv file, one would include the ID and label for each node (also timeset if the node is dynamical).

In order to import the edge.csv and node.csv files into Gephi. You need to:

- Select new project once you enter Gephi.
- Click on “Data Laboratory” on the top left.
- Click on “Nodes/Edges” on the top left under the Data Table tab, then click “Import Spreadsheet” and select your node.csv/edge.csv file.
- Follow the instructions to finish importing.

2 Basic network manipulation through Gephi

Once you imported the network, then we can start some simple manipulation.

You can view and access the network graph in the overview section (top left). Once you are under the graph tab, you can see your network visually. At this point you can use mouse/trackpad and cursor to move the network around. Zoom in & out using trackpad or mouse (using two fingers on trackpad up and down).

Right now your graph should look quite random. So the first thing to do is to choose a reasonable layout and see if there's any underlying structure of this network. Click the layout section on the top left and try different layouts (recommend you try "ForceAtlas 2", "Fruchterman Reingold", and "Yifan Hu").

After choosing an appropriate layout, one should change the size and color of nodes & edges based on their properties. For example, you can color the nodes based on their degree, or you can change the size of nodes based on their centrality.

Furthermore, to find label/coordinate for a node, click the big "T" at bottom for "Show node labels". For Diseaseome will need to decrease font size and then zoom in. (Click button for links to right to see what network looks like when all edges are removed.)

3 Network statistics

There are lots of different network statistics you can try out. The network statistics module is on the right (next to the filters). You can see that there are different network statistics you can run. Essential statistics that you should try out first are average degree, network diameter, modularity, and etc. . .

4 Node & edge attributes

You can get the node & edge attributes from the Data laboratory section.

Before you start, you should go to, <https://github.com/gephi/gephi/wiki/Datasets>, and download the biological datasets (there should be three: `celegans.gexf`, `diseaseome.gexf`, and `yeast.gexf`). You should also download the dataset I sent out through email (can also find under week 8 in CCLE [name: lab8_data.zip]).