Praxis: Gephi Network Graph **Network Science Fall 2016**
Sean M. Shiverick **(Due** Sep 12 by 11:59pm)

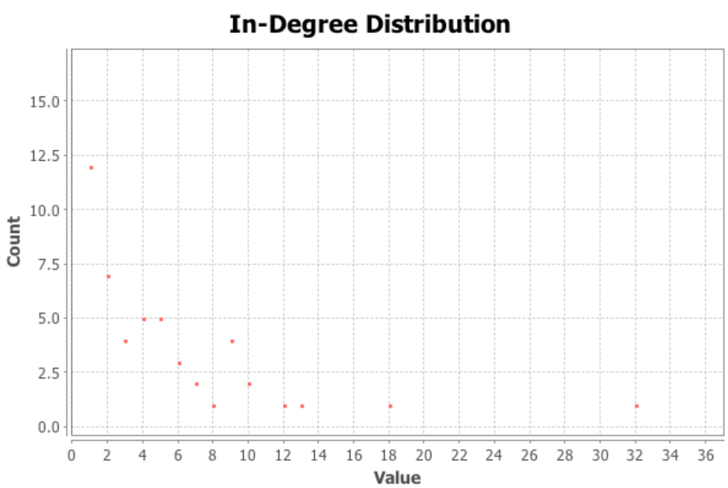
This assignment introduces you to Gephi, a graph visualization software that you can use to calculate statistics of interest about a graph as well as make it visually presentable. Download and install Gephi become familiar with some of its basic functionalities: Now that you have a basic understanding of some of the things Gephi can do, you will put that knowledge to use: First, acquire some real graph data. This site has some graphs which you can download and use with Gephi. They are already in a format that Gephi will recognize. For the first part of the assignment I want you to use the Les Miserables graph. In the second part of the assignment you can use your own graph, one of the other graphs on the site, or any other graph you can find. However, the graph you choose **must** come from real-world data and it should be small enough to manage with Gephi and Jupyter notebook --sorry no million node graphs :(

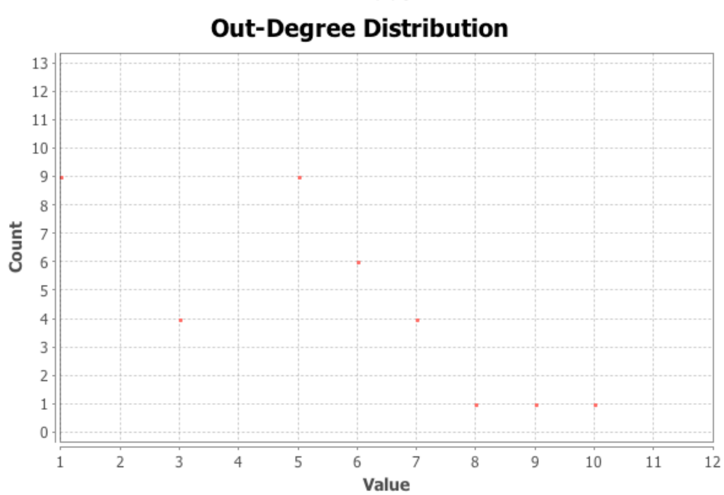
Once you have the Les Miserables graph do the following:

1 Load the graph into Gephi.

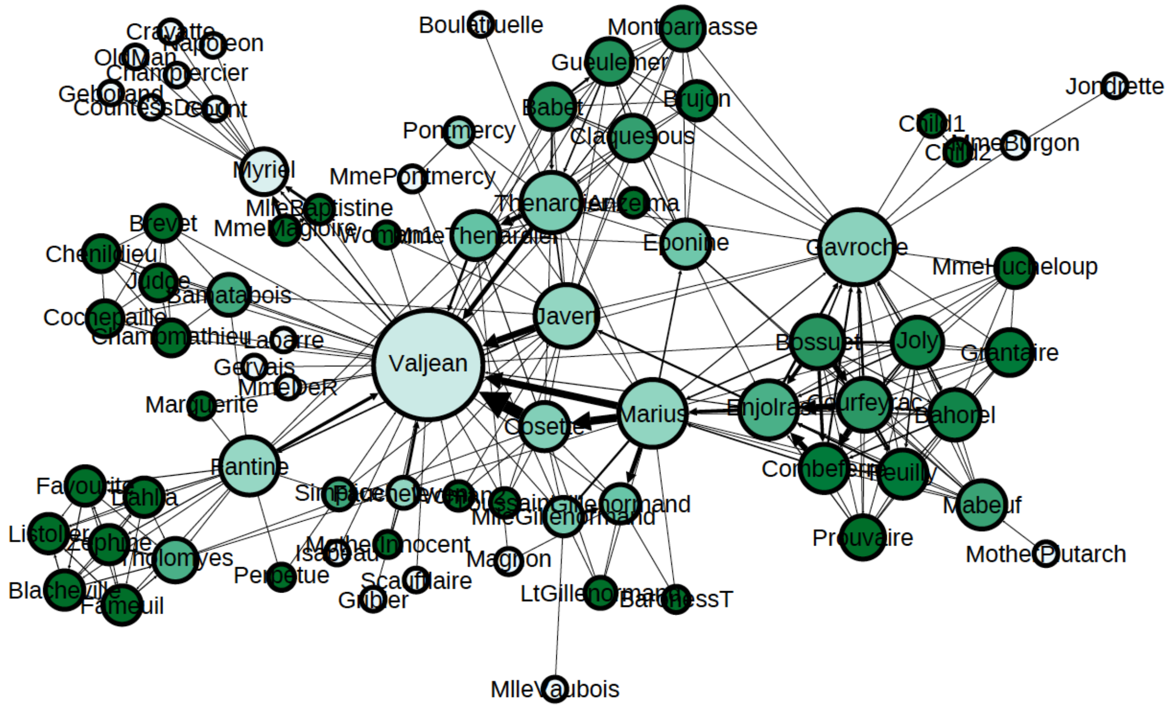
2 Calculate and record each of the following for the graph
- number of nodes: 77
- number of edges: 254
- average degree: 6.597
- average clustering coefficient: (directed) 0.287 (undirected)
- average path length: 2.4

3 Save a plot of the **Real degree distribution** of the nodes (both in/out degree if directed).





4 Then, select a graph layout that you think best visualizes the graph and save it.



Random Graph Based on LesMis Gephi Graph

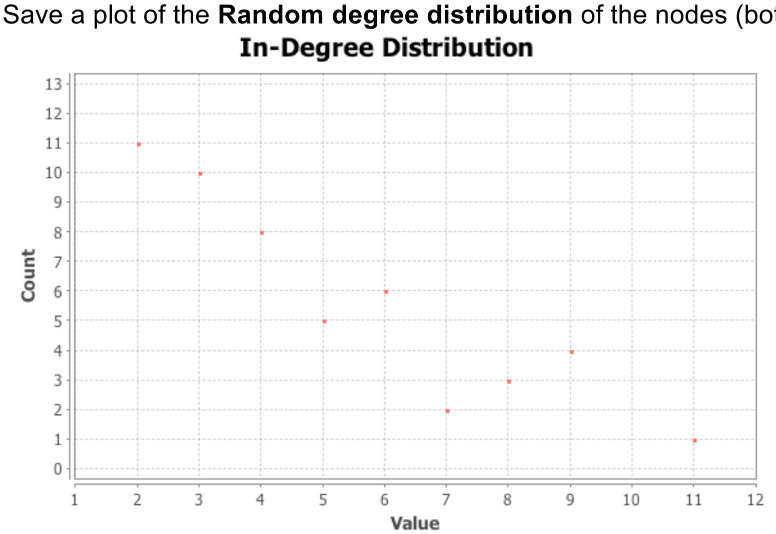
5 Use graph generator tool introduced in the last assignment to create random graph with same number of nodes and about same average degree and number of edges as the real graph.

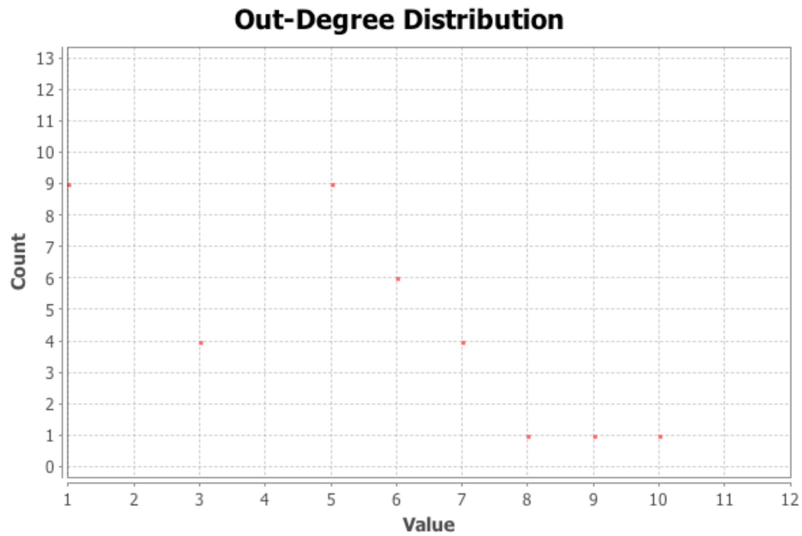
6 Save this random graph to a file format that Gephi can read (see this page for Networkx read/write functions) and perform steps 2-4 with this new graph.

Calculate and record each of the following for the graph

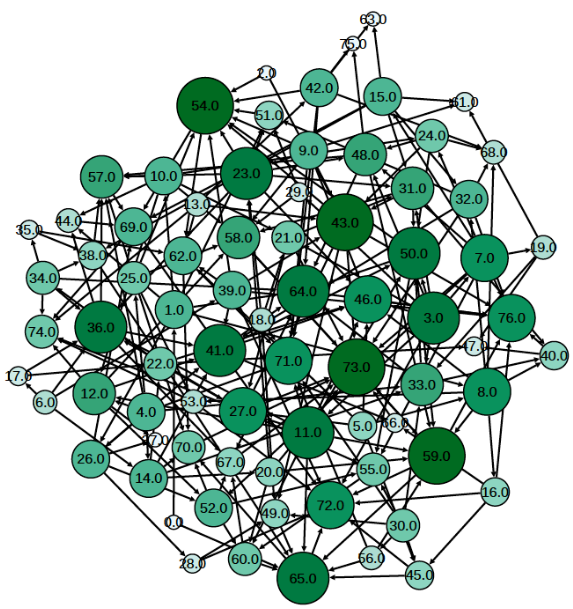
- number of nodes: 77
- number of edges: 247
- average degree: 6.416
- average clustering coefficient: 0.046
- average path length: 2.35

Save a plot of the **Random degree distribution** of the nodes (both in/out degree if directed).





Then, select a graph layout that you think best visualizes the graph and save it.



Answer the following questions:

1. *How do the degree distributions of the real graph and random graph compare? Explain any differences or similarities you see and consider why they might exist.*

For the in-degree distribution, the real graph had more low degree nodes and a greater number of high degree nodes than the random graph. The out-degree distribution followed a Poisson distribution.

2. The real graph was more clustered than the random one, as evidenced by the clustering coefficient. In examples from real life (or in this case, a character map from fictional novel) there are ordinary groupings, that are connected by random elements or weak links, different groups are connected by hubs.

3. Yes, I think that the real graph does exhibit the small-world property because we have clustering with strong link, weak links connecting clusters, and hubs that facilitate connections between groups in the network. The small-world characteristic of the character graph suggorts the idea that real-world networks or systems follow similar organizing principles.

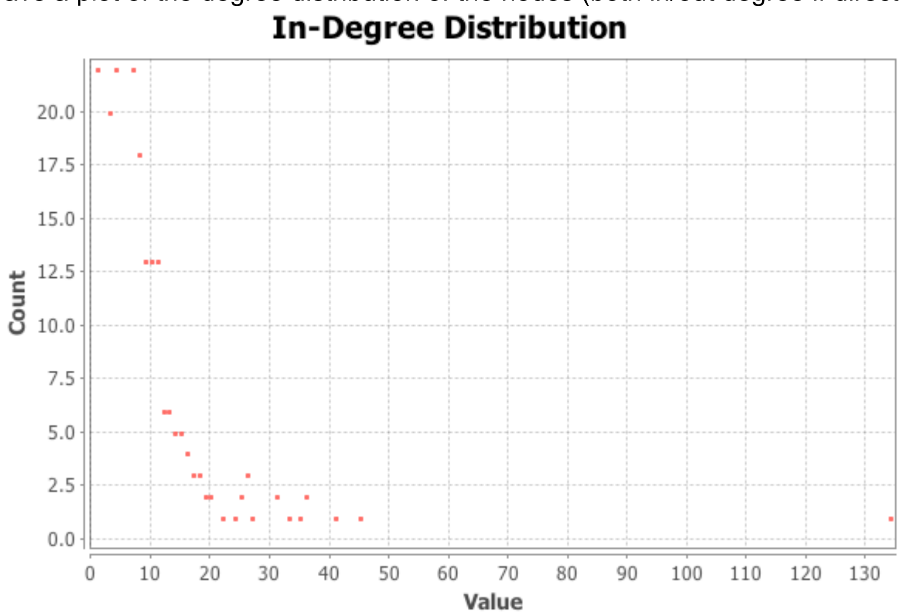
After Les Miserables complete the process again with another real world graph of your choosing.

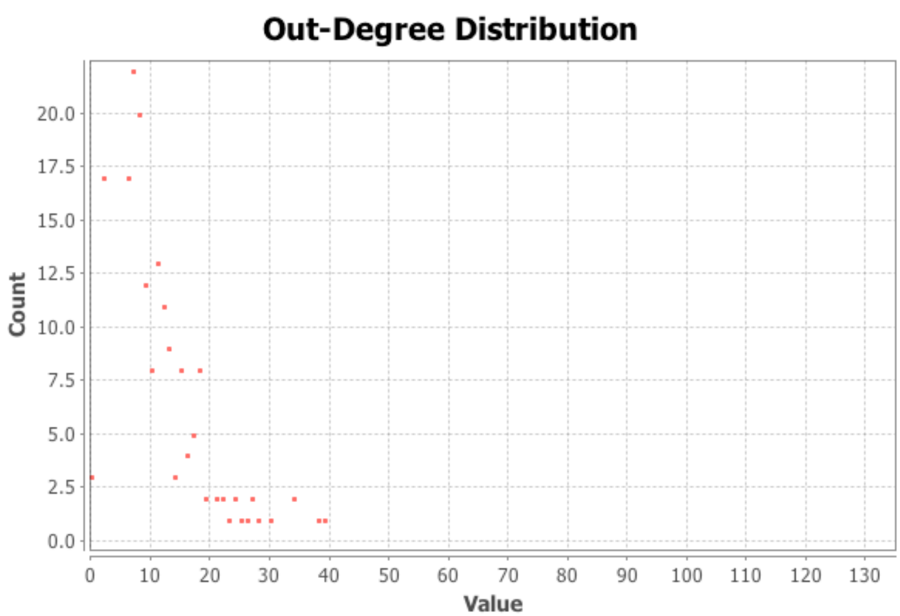
Graph based on graph of C. Elegans neural network:

1 Load celegans.gsm graph into Gephi.

2 Calculate and record each of the following for the graph
- number of nodes: 297
- number of edges: 2345
- average degree: 15.791
- average clustering coefficient: 0.169
- average path length: 3.992

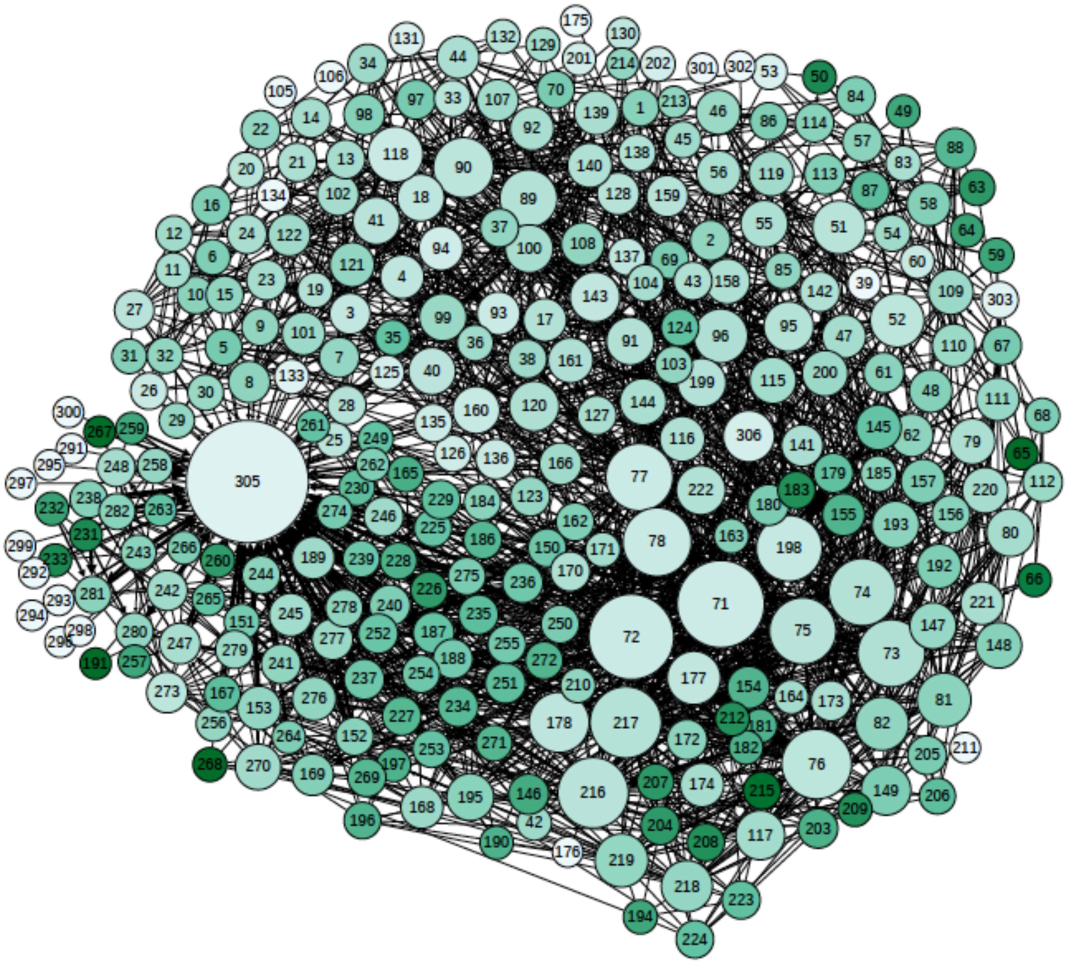
3 Save a plot of the degree distribution of the nodes (both in/out degree if directed).





4 Then, select a graph layout that you think best visualizes the graph and save it.

Gephi Graph of C. Elegans Neural Network



Random Graph Based on C. Elegans Neural Network Gephi Graph

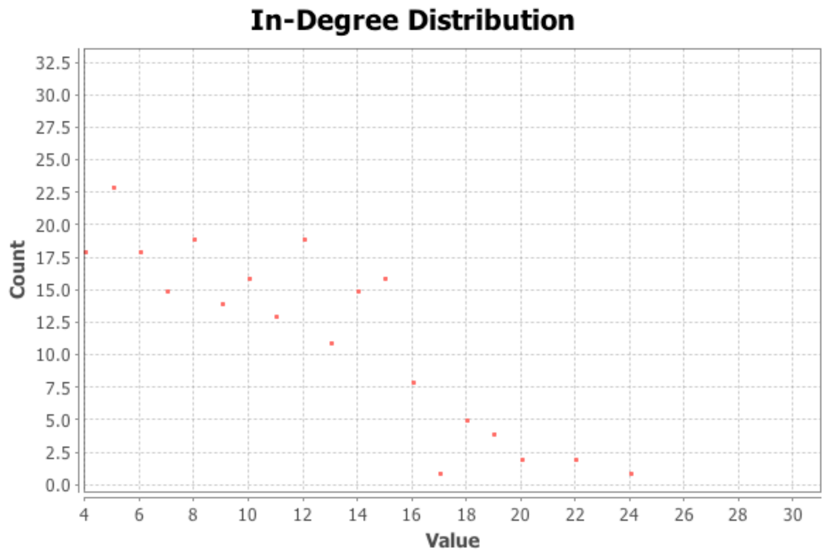
5 Use random graph generator tool to create a random graph with the same number of nodes and about the same average degree and number of edges as the real graph.

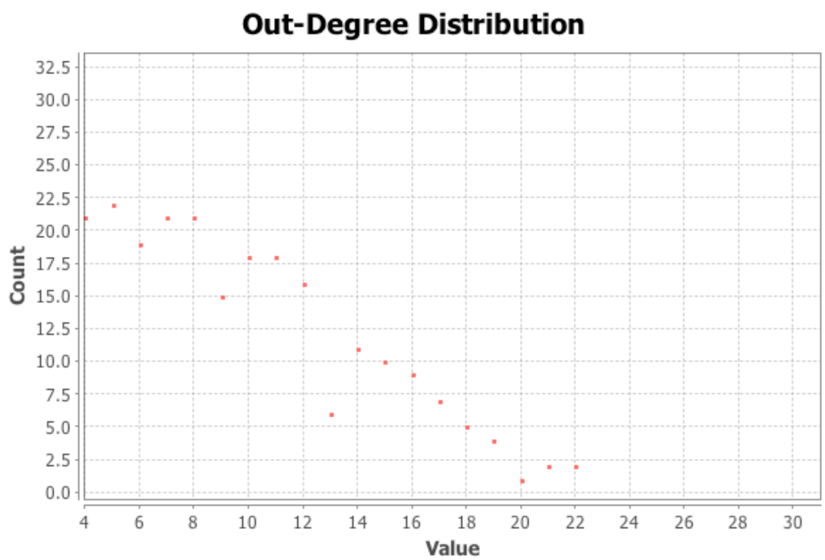
6 Save trandom graph to a file format that Gephi can read (see this page for Networkx read/write functions) and perform steps 2-4 with this new graph.

Calculate and record each of the following for the graph

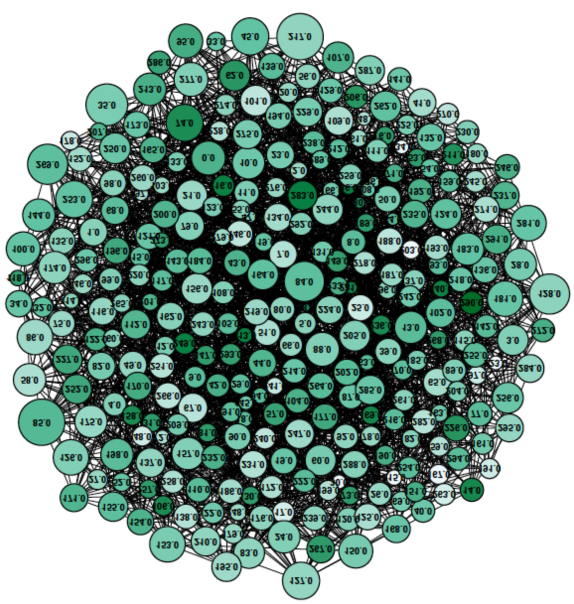
- number of nodes: 297
- number of edges: 2338
- average degree: 15.74
- average clustering coefficient: (directed) 0.028 (undirected)
- average path length: 2.69

Save a plot of the degree distribution of the nodes (both in/out degree if directed).





Then, select a graph layout that you think best visualizes the graph and save it.



Answer the following questions below for the C. elegans Neural Network and Random Graph:
1. How do the degree distributions of the real graph and random graph compare? Explain any differences or similarities you see and consider why they might exist.

The scale for the real neural network for C. elegans and the random network are different; however, for both the in-degree and out-degree distributions of the real c. elegans neural network both show high numbers of nodes with low degree and a steep decay with a few (or no) nodes with high degree. By contrast, the in-degree and out-degree distributions for the random graph both tapered off, with few nodes having more than 20 degrees. The real network distribution showed several nodes with between 30 to 40 degrees and one or two with very high degrees, which are the hubs.

2. The real graph is much more clustered than the random one as shown by the difference in the clustering coefficients (0.169 versus 0.028, repectively), and this tell us that the real graph is much more organized in terms of local clusters or communities tied by weak links similar to the small world pattern than is the random graph.

3. It does appear that the real graph of the C. elegans neural network exhibits the small-world property. The pattern of small-world network organization with small clusters of nodes interconnected with other clusters by random links is an efficient pattern of organization, which reduces the distance within a network and increases the connectivity between nodes. It is very fascinating to see this in a real world example of a neural network system of a biological organism.

(When complete, upload a PDF of work and Jupyter notebook used to generate random graphs)

Using Gephi (Tutorial)

- Download and load the graph into Gephi (lesmis.gml)
- OVERVIEW: Drag tool (configure radius), mouse selection (ctrl to delete, remove nodes)
- Add labels, adjust font side, Edge size and opacity of edges
- **Statistics tab** [run] average degree, clustering coefficient, average path length
- **DATA Laboratory** contains all info for graph: Change names, add edges, nodes, etc.
- **Appearance Tab** (upper left) use to adjust appearance color, size of Nodes, Edges
- Adjust Node Attribute to *Clustering Coefficient* by color, Edges according to weight
- **Layout:** ForceAtlas2 is nice Node layout for graphs with community structure
- *Adjust Scaling* (size of graph), LinLog mode (Scale-Free Graph with large degree dist)
- PREVIEW: Add labels, adjust label size (8), unclick proportional label size, edge arrow width
- Click **Refresh** after each change
- **Export** Gephi Graph as pdf file and save to desktop
-