BU.330.760 Deep Learning with Unstructured Data
Lab 1. Gephi and Simple Network Analysis

Learning Goal: practice using Gephi to visualize and analyze a social network
Required Skills: knowledge on graph theory

1. Install Gephi from https://gephi.org. Gephi is developed in Java, and it requires Java version 7 and later. Note that on Mac OS X, Java is bundled with the application so it doesn't have to be installed separately. But on Windows and Linux, the system must be equipped with Java. See the system requirements page: https://gephi.org/users/requirements/ for more details.

2. After you install and start Gephi, go to Help -> Check for Updates. Install the Updates, close Gephi and open it again.

3. Download the dataset at https://gephi.org/datasets/lesmiserables.gml.zip which is a coappearance weighted network of characters in the novel Les Miserables. Load the data in Gephi using **File > Open**. Make sure you've selected the "Overview" tab at the top of your screen if not.

4. In **Layout > Choose a layout**, which is located on your left bottom corner, choose "Yifan Hu Proportional" layout, and click the green "Run" button. The green "Run" button will return once the process is complete.

5. To save the visualization, go to **File > Export** or press the "Export" button (lower left corner of the page on the "Preview" window). If you want to see the node labels, go to "Preview" tab, under "Node Labels" section, check "Show Labels" and uncheck "Proportional size" option. Then export again.

6. Describe the macro-level structure using the Yifan Hu Proportional visualization.

    a. Is it a giant, connected component, are there distinct sub-components, or are there isolated components?

7. Go to **Overview > Context** window (located on top right corner) **> Statistics** tab (next to **Filters** tab) **> Network Overview > Modularity**, Press the **Run** button

    a. How many communities have been created?

8. Go to **Data Laboratory** tab (located next to **Overview** tab near the top) and sort the nodes by clicking on top of the column "Modularity Class". Can't see the modularity class? (Or any other column in this assignment?) Click on the light-bulb in the right hand-corner and uncheck columns until "Modularity Class" is visible on your screen.

9. Go to **Overview > Nodes > Partition**, select the parameter "Modularity Class" and press "Apply" button. Select the layout "Yifan Hu Proportional", run it, then go to **Preview** tab at the top. Refresh and save this diagram.

    a. What information does this layout convey?

    b. Are the clusters well separated or is there a great deal of overlap?

    c.   Describe the brokers between any components and cliques.

10. Now, in the **Statistics** tab, run the following:

    a.   What is the average degree (for degree centrality) and network diameter (for betweenness and closeness centrality)?

    b.   What is eigenvector centrality (under **Node Overview** section)?

11. In the previous step you computed a number of graph metrics. Doing so automatically created columns in the **Data Laboratory** tab with values for these metrics. Additionally, in the **Context** window, you can see the distribution of the centrality measures you computed by pressing on the small question mark next to **Run** (you may have to scroll to the right to see it).

    a.   Describe the distribution of degree centrality scores (just called "degree" in Gephi)? What is the min, max, and average degree? Does this appear to resemble a power law distribution?
    b.   Describe the distribution of betweenness centrality scores (click the '?' beside 'Network Diameter').
    c.   Describe the distribution of closeness centrality scores in your network?
    d.   Describe the distribution of eigenvector centrality scores in your network?

12. Using the **Statistics** window, **Graph Density > Run**

    a.   What is the density of the network?

Submission: None for lab 1

Reference:
IEMS 441: Social Network Analysis Lab 1, Noshir Contractor, 2015, Northwestern University.