Elizabeth Fabio
BDA-594

Web Exercise 6: Social Network Analysis (Gephi)

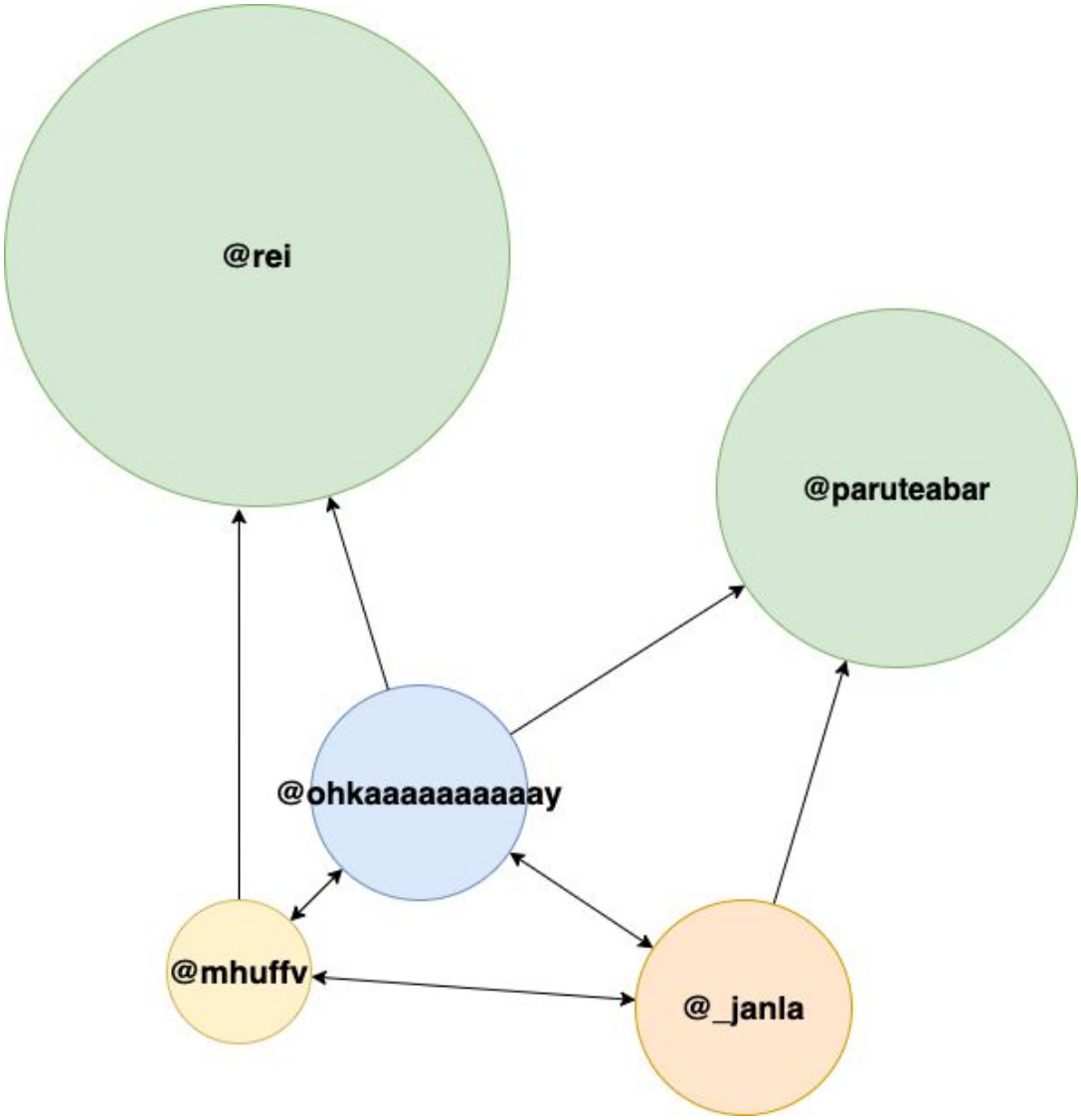
1. Explain what is a “node”? What is a “Edge” in this exercise example? Please provide a social network example for at least FIVE nodes and SIX edges. Draw the network graph in your report. (You can draw on a paper and then take a photo using your phone, or scan it into a digital picture. You can also use any graphic software to draw the network graph.)

   A node is an entity or point in a network. The connections between nodes are reflected by their edge. In the case of the Tweets Social Network example, the edge reflects Twitter users tagging other users (out-degree) or Twitter users being tagged by others users (in-degree).

   A simplified example of a social network can be seen in the next page (IG-SNA.png). The nodes reflect Instagram followers labeled with their username. The center node is my username (@ohkaaaaaaaaaay). Some nodes reflect users that I follow who follow me back (@mhuffv, @_janla). Other nodes reflect users I follow that do not follow me back (@rei, @paruteabar). The arrows represent who is following who. The direction of the arrow on the edge can reflect if one user is following another user (out-degree) or if one user is being followed by another user (in-degree). The size of the node also reflects how many followers that Instagram account has. Since @rei is a popular outdoor company, they have the most followers (2.3 million). As for @mhuffv, he has the smallest node size with only 394 followers. Since the node with my username is the central point (blue), the colors reflect the follower relationship to me. @rei and @paruteabar are green to reflect I only follow them and @mhuffv and @_janla are yellow to reflect we follow each other.

Elizabeth Fabio
BDA-594

IG-SNA.png



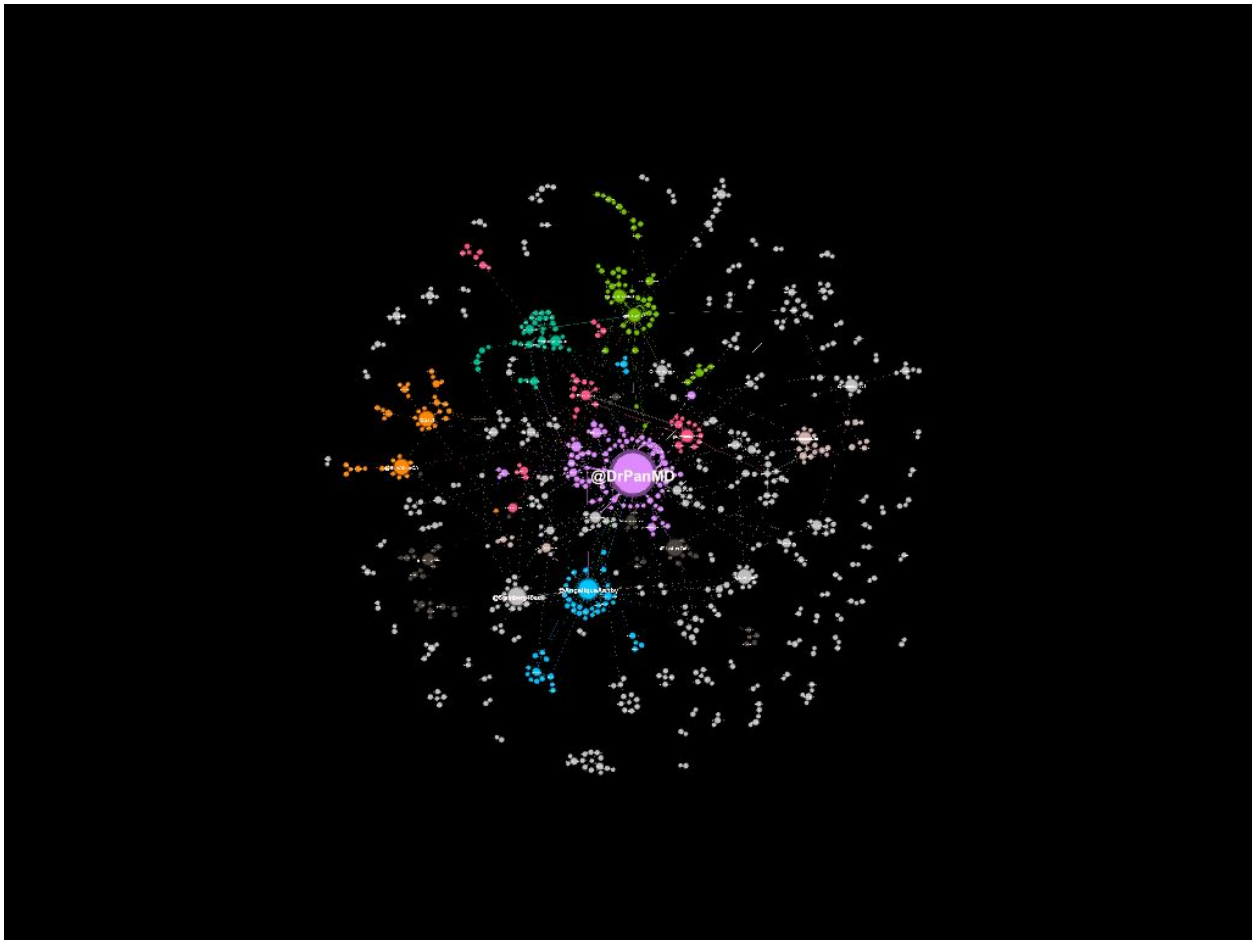
2. Display the In-degree network screenshot (one layout) and the Out-degree network screenshot (another layout) in Gephi using different layouts. Explain which layout scheme do you select for the In-degree network and the Out-degree network in this exercise. (Note: DO NOT use the same layout for both networks).

   The layout I choose for the in-degree network is the **Force Atlas 2**. The following conditions were made:

   - **Node size**: Ranking > In-Degree (Min: 20, Max: 200)
   - **Node color scheme**: Partition > Modularity Class
   - **Layout**: Force Atlas 2
     - **LinLog Mode**
     - **Prevent Overlap**

   Since we are focusing on an in-degree network, I thought the node size should reflect this. Also the color palette was adjusted to observe clustering. With both of these conditions, we should see clustered units to have one bigger central node with smaller nodes surrounding it. This is to show the center node to have a higher in-degree and the smaller nodes pointing at the central node to have a smaller in-degree. The LinLog Mode was checked to make sure nodes that are connected were clustered together. Also, the Prevent Overlap behavior was checked so nodes would not overlap each other. Overall, the Force Atlas 2 layout not only showed node attraction but also node repulsion.

SNA-ForceAtlast2-InDegree.png



The layout I chose for the out-degree network is the **OpenOrd**. This was to practice highlighting node clustering. It is originally based on the Frutcherman-Reingold layout and incorporates adjusting the iteration (Default: 750, Minimum: 100). The smaller the iteration the more dense the node network appears. The number of iterations are controlled via a simulated annealing type schedule (liquid, expansion, cool-down, crunch, and simmer). I kept most of the default values except the following listed:

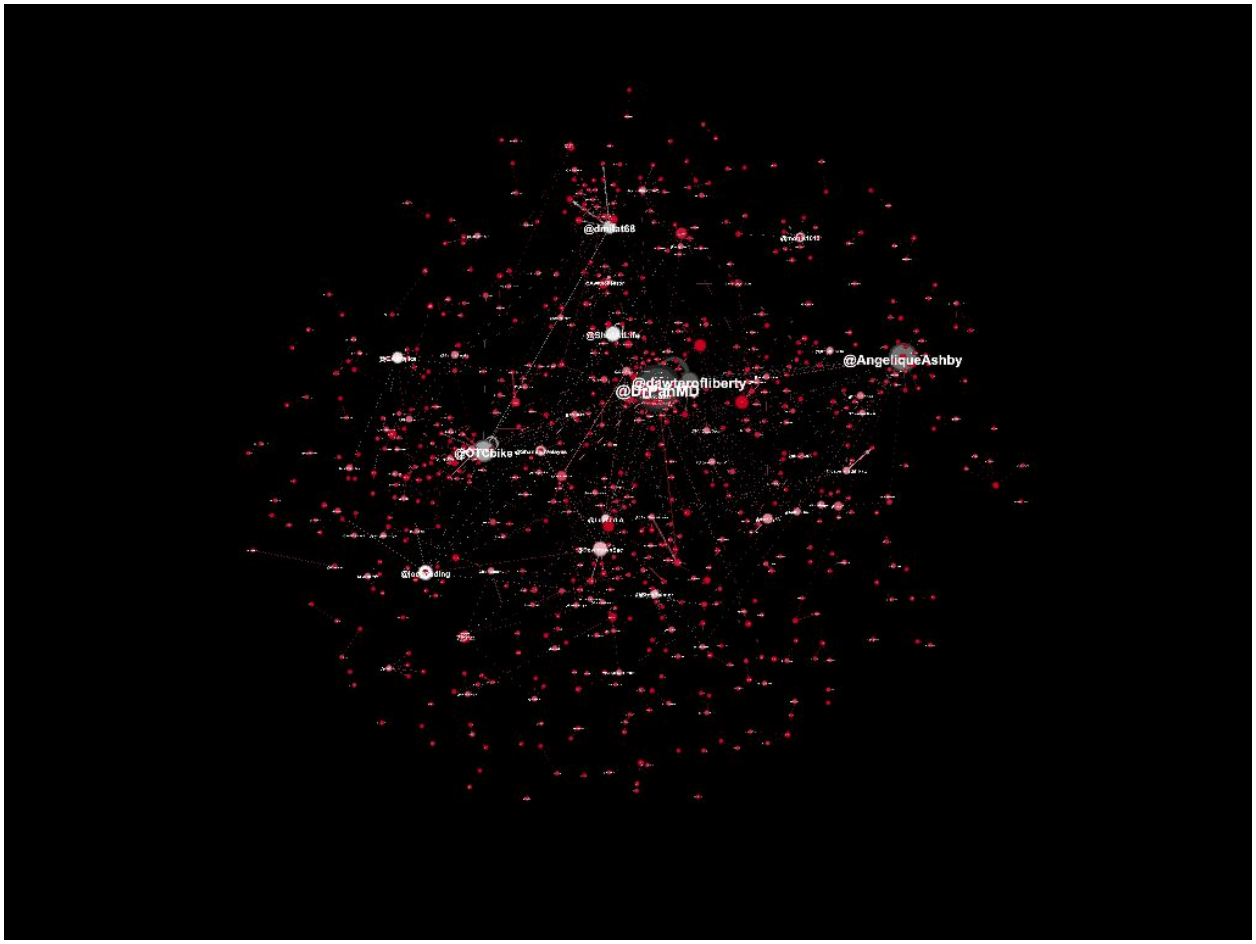
- **Edge Cut**: 1.0
- **Num Iteration**: 1000

These changes were made to emphasize node clustering. The following changes to the node size and color can be seen below:

- **Node size**: Ranking > Degree (Min: 20, Max: 200)
- **Node color scheme**: Ranking > Out-Degree
    - **Red**: Low number of out-degrees
    - **White**: Medium number of out-degrees
    - **Black**: High number of out-degrees

The node size in the out-degree network applies the overall degree (in-degree and out-degree). The colors scheme applied is a red/black value to emphasize the number of out-degrees. Thus, the nodes with higher number of out-degrees were shown in black and the nodes with the lower number of out-degrees were red. In between, the nodes were white. In my figure, most of the nodes had a lower number of out-degrees since they were red. A few nodes that had a higher number of out-degrees stood out with a black color. They also stood out because their node size reflected their overall degree too.

SNA-OpenOrd-OutDegree.png



3. Explain the meanings of “Network Diameter”, “Graph Density”, and “Modularity”. What are the values of these parameters in the lab-06 example network? Is this example network similar with random networks, or small world networks, or scale-free networks? WHY?

Network diameter is the shortest distance between two of the farthest nodes. In the Graph Distance window when running the statistics, the following results are displayed:

- **Betweenness Centrality**: How often a node appears on the shortest path between nodes.
- **Closeness Centrality**: Average shortest distance from a node to all other nodes.
- **Harmonic Closeness Centrality**: A derivation to closeness centrality. It would solve the original formula when dealing with unconnected graphs.
- **Eccentricity**: Measures the distance between a node and the node furthest from it. The largest distance between nodes is the network diameter value reflected in the Graph Distance window.

For **graph density**, it measures how tightly interconnected a network is by calculating the current network's edges relative to the possible number of connections. As for **modularity**, it is an algorithm that detects clusters/modules. In the modularity report, the graph displayed is the size distribution of all the clusters/modules in a network. From the the lab-06 example network, the calculated network characteristics is listed below:

- **Network Diameter**: 11
    - Radius: 0
    - Average Path Length: 4.07920393199581
- **Graph Density**: 0.001
- **Modularity**: 0.830
    - Modularity with Resolution: 0.830
    - Number of Communities: 97

Based on this exercise, the network appears to be a small-world network. A small-world network has a higher modularity and almost the same average path length of a random network with the same number of nodes and edges. Thus, it makes sense why our exercise network appears random due to the wide degree distribution from one to 65. The only reason it would not be classified as a random network is because of its high modularity value (0.831). Based on the modularity report, 97 cluster communities are formed. This can be observed in the in-degree and out-degree network screenshots when using different base layouts (OpenOrd and Force Atlas 2). A scale-free network would not fit the definition because the exercise network does not have a significant number of nodes with a lot of connections. The average node degree is actually only 1.042.