Using Gephi to Visualize a Network

Gephi is a leading network visualization tool. It is free open-source software that has been download more than a million times. It is sometimes called the "Photoshop for networks". Many data journalists and data scientists use this tool to produce professional visualizations of networks.

Below we give some instructions for using Gephi to visualize a followers-of-followers subgraph for two Twitter users: TechnologyMindz (https://twitter.com/technologymindz) and ChrisGrant360 (https://twitter.com/ChrisGrant360). A followers-of-followers subgraph is subnetwork of the followers of the user and and the followers of the user's followers. Note that a network is a graph, and a subnetwork (or subgraph) is a part of a large network. For more background on networks, see Abert-Laszlo Barabasi's online book called *Network Science* at http://barabasi.com/networksciencebook/. We will refer to sections in this book in certain places below.

Gephi Setup

Make sure you have the most recent version of Java (https://www.java.com/en/download/) installed on your machine. Download and install Gephi from https://gephi.org/. (Mac instructions are available from https://gephi.org/users/install/.) Increase the memory allocated to Java. This increased memory will help Gephi visualize a larger network. Control Panel > System > Advanced System Settings. Add a New System Variable.

Variable name: _JAVA_OPTION
Variable value: -Xmx512M –Xms1400M

Then restart your machine.

Next, open Gephi. Set antialiasing to 16x. Antialiasing is a computer graphics technique that will make your edges look smoother. From Gephi's menu, Tools > Options > Visualization > OpenGL > Antialiasing: 16x. Close Gephi and reopen it. Try to keep most other programs closed when you use Gephi. It is fine to have a few other programs, such as Juypter or Excel, open. Close the "Welcome to Gephi" dialogue box. We will open our edges files directly from the Gephi's Menu.

Network Analysis

Let's explore the two edges files (TechnologyMindzEdges.csv and ChrisGrant360.csv) in Excel. The first column in each file contains a list of source nodes and the second column contains a list of target nodes. The rows describe edges or links from source to target nodes. In this case, each row contains a source (a user) who has the target (another user) as a follower on Twitter. In network terminology, an important property of a node is its degree – the number of links the node has to other nodes. In a directed network like Twitter, there is also a distinction

between incoming degree (the number of followers that a user has) and outgoing degree (the number of other users that a user follows). Read Sections 2.1-2.3 in *Network Science* (http://barabasi.com/networksciencebook/chapter/2).

Note that, unlike the video we ask you to watch below, we will not use a nodes file for this exercise. Also, we will open our edges files directly from File > Open, rather than create a New Project and use the Data Laboratory. In addition, when opening edges files directly, there is no need to label our two columns "Source" and "Target" as suggested in the video.

Open the two edges files (TechnologyMindzEdges.csv and ChrisGrant360.csv) in Excel. For each file, create a pivot table from the first column. How many followers do these two users have? How many followers do each of their followers have? Before moving to the next step with Gephi, save these files with the pivot tables under new names.

Network Visualization

Open the edges files in Gephi, e.g., File > Open > TechnologyMindzEdges.csv. From the Import report, choose the Graph Type "Directed". Click OK. If the Graph window is not already open (you will see a randomly laid out square network if it is), then open the graph window. Window > Graph.

Change the Graph to a new layout. Window > Layout. From the Layout tab on the left, choose ForceAtlas 2 and press Run. Let it run for a few minutes. After it looks like it is not changing its shape much, press Stop. Most layout algorithms group nodes into clusters and then push the clusters apart from one another. Now, center the graph. Press on the "Center on Graph" magnifying glass at the lower left of the Graph window. Add labels to the graph. Press on the dark bold "T", or "Show Node Labels" element, at the lower left of the Graph window. What do you notice about these two subgraphs?

Next, watch the Jennifer Golback's Gephi tutorial at https://www.youtube.com/watch?v=HJ4Hcq3YX4k. Skip to the 6:25 mark, unless you want to see her load the data from a spreadsheets using the Data Laboratory. Her data are on a co-citations network of those who have cited Stanley Milgram's paper on "small worlds" and each others' papers.[1] Notice how she applies filtering in Gephi. Filters > Topology > Degree Range. She drags "Degree Range" to the Queries section and then filters out the singleton nodes. From Statistics, she runs Network Diameter to get some centrality measures.[2] She runs it on an undirected network. She then changes the color and size of the nodes based on betweenness and closeness centrality.

---

[1] Stanley Milgram conducted a seminal experiment on networks in 1967 (https://en.wikipedia.org/wiki/Small-world_experiment). He showed that all humans are connected on average by six degrees. This experiment is where we get the phrase "six degrees of separation". This phrase means that, on average, we have to go through five other people to get to any other person.

[2] For instance, "betweenness centrality" for a node is the number of pairs of other nodes where the node lies along the shortest path between the pair. This measure is usually applied to undirected networks.

Filtered Network Visualization

Close Gephi to discard the layouts of the two complete subgraphs. We will run through the tutorial's steps on our edges files, but with a filter on outgoing degrees. We will look at only nodes in our subgraphs that have one or more outgoing degrees.

Open Gephi, and then open each edges file. Keep the networks as directed ones. Filter on outgoing degrees: Filters > Topology > Out Degree Range. Drag "Out Degree Range" to Queries, change the bottom end of the range to 1, and press Run. Run the Yifan Hu layout, and then run the Expansion layout a few times until the network fills the Graph window. Add labels to the graph. Increase the edge weights and the label sizes using the slider bars at the bottom of the Graph window. Change your node color and size on the basis of in-degree. Overview > Appearance > Nodes > Attribute > In-Degree > Apply.

Compare the two filtered subgraphs on the statistics "Average Degree" and "Graph Density". Which filtered subgraph has the higher average degree and graph density? What is the significance of this analysis? Assume for a moment that these two users have the same number of total followers of their followers, which of course is not actually the case here. Nonetheless, we can conduct this thought exercise. Under this assumption, which user is likely to get more retweets on any one of their tweets?