

## *Business Data Preparation and Visualization*

# **Network Visualization Assignment**

*Make sure you watch the videos before working on the assignment.*

### **Learning Goals**

1. Network is a whole different approach to visualizing data. It is not just explicit networks such as those you see on Facebook (friends, groups etc.), LinkedIn (contacts), or Twitter (follow/retweet); but rather, it is extremely flexible. It is up to you to decide what each NODE represents, and what each LINK (directed or non-directed) represents. The sky (your imagination) is the limit. Understanding this can allow you to explore your data, even apparently mundane datasets, in very interesting ways.
2. Network is also the foundation for some of the visualizations that you will see later in programming, such as the chord diagram and the Sankey diagram. In fact, the optional hub-and-spokes map demo that I gave you in Tableau before, that's also an example of a network.
  - a. Network visualization in Tableau, at the moment, is still quite a pain. Here is an intro: <https://kb.tableau.com/articles/howto/creating-a-network-graph>. It should be noted that unless you are willing to do some coding, you cannot really change the layout of graphs within Tableau like we show in NodeXL. You have to give each dot its coordinates.
  - b. For that reason, it is still difficult to incorporate networks as part of a dashboard, **though you should certainly explore it for your stories** (making your case).
3. Make sure you are familiar with the key concepts we mentioned in class (all of which and a lot more can be found online at <http://faculty.ucr.edu/~hanneman/nettext/>)
  - a. Nodes
  - b. Edges
    - i. Directed vs. non-directed
  - c. Components / clusters ("Groups" in NodeXL)
  - d. Properties of edges or nodes
    - i. Sizes of nodes
    - ii. Size of edges
  - e. Some metrics
    - i. Degree centrality
      1. In-degree centrality; out-degree centrality
    - ii. Density
4. Make sure you know how to visualize simple network data in NODEXL BASIC (or other tools such as Gephi if you prefer, as long as you know how to do what we did in NodeXL there).

- a. If you use a Mac, you need to use MyCloud to use NodeXL. Please follow the instructions in the syllabus on how to download and install it on MyCloud. You should have sufficient privilege to do so. It is an Excel template, so please close Excel before you install it.
- b. If you want to know more about NodeXL, beyond what we discussed in class, there are many resources available online, such as this:  
<https://www.smrfoundation.org/2009/06/23/new-tutorial-available-analyzing-social-media-networks-learning-by-doing-with-nodexl/>.

**\*\* You will submit two files for this assignment – a Word document (Parts I, II and III), and an Excel file (for Part I).**

### **Part I (70%)**

For the first part of this exercise you will need the data from <http://bit.ly/netexample>. You need to create a Word document (for those that I need you to keep screenshots or where you need to answer) and the final version of the Excel file that you have after finishing these questions.

*Naming convention: LastName\_FirstName\_Network.docx/xlsx*

1. Import the data into NodeXL
2. Try at least four layout algorithms. Take a screenshot and paste into your Word document. Make sure the labels show the name of each person. Remember that you can manually move nodes around to create your own visualization; they look different, but they are the same network.
3. Use a different shape (other than the default) to show the nodes. Take a screenshot and paste it into Word.
4. Format the visualization so that the size of the nodes corresponds to the in-degree of each person. Keep a copy of that graph.
5. Highlight Alam in the graph using a different color. Take a screenshot and paste it.
6. Add the following edges (tip: just add rows to the EDGES sheet), and then create another visualization. Be sure to try different algorithms and find the best way to show the graph. Take a screenshot and paste to Word.
  - a. Ali to O'Neal;
  - b. Jones to Abrams
  - c. Powell to Clemence;
  - d. Cohen to Smith
7. Remove the four new ties created above. Then, remove Alam from the graph and all of Alam's ties. Create a graph to show the disconnected two components. Describe what you did and paste the graph in the Word file.

**Part II (30%): Your answers to the following questions should be added to your Word document from Part I above.**

1. Think about any one of the datasets that you are using for your group project. Define/envision a network that may be useful in showing a pattern:
  - a. What would you define as a node?
  - b. What would you define as a link (edge or arc – directed or non-directed)?
  - c. What would this graph show, if you have the full graph?
2. Enter at least 8 of those edges in NodeXL Basic (or other tools of your choice) by hand, based on your data. Feel free to enter more than that if you'd like. Show a graph with the label on, and paste that network graph into the Word document (no need to submit the Excel file).

**Part III (not graded --- recommended exercise). Feel free to share your thoughts.**

Explore other network datasets to see the variety of ways that networks can be constructed and used. There are many network repositories online, and some of them may just become your **inspiration** to do something fun with your own data or projects. Some of them can be large, but you can play with part of the data for exploration purposes.

- <https://sites.google.com/site/ucinetsoftware/datasets> (note that network datasets here are in the UCInet format, which is one of the formats that NodeXL can import and visualize.)
- <http://networkrepository.com>
- <https://kateto.net/2016/05/network-datasets/>