**Group Network Project**

This project will use NodeXL, a social network analysis tool, to analyze a Twitter dataset related to US Senators. For the project, you will be given research questions to answer based on a data set given to you. You will use the software to analyze the data and answer your questions.

In class we will discuss the method of network analysis, major concepts of networks (nodes, ties, clusters, etc.), centrality and centrality measures (degree centrality, closeness centrality, betweenness centrality, eigenvector centrality), and network visualizations. Application of these concepts to a Twitter network will be explored.

We will spend some class periods in a computer lab learning how to use the software and interpret the data. Please note that you will also need to spend time outside of class working in the lab. I will reserve specific times in the lab for you to drop-in. NodeXL is only available for PCs (not Macs), and the lab computers will have the upgraded version of the program necessary for this assignment.

Finally, the project will culminate in a brief (about 3-5 page) research report that presents your results both in text and in network visualizations. You will need to respond directly to the RQs, “show” your data pertaining to these, and offer educated explanations about what the data supports.

Each group will submit 1 research report, and all group members will receive the same grade for the work.

**Data and RQs / Hs**

The data for this project will be the recent Twitter activity of US Senators. You will be given the data set as an Excel file. You should briefly familiarize yourself with some of the Twitter accounts of the individuals in the data set (i.e. US Senators’ accounts). You don’t need to view all of the accounts, but look at 10 or so. Notice who they interact with and what sorts of activity they utilize on the platform (tweeting, retweeting, liking, replying, mentioning, hashtagging, etc.). You don’t need to formally track this activity, but get a sense for how these individuals use Twitter.

Then, using the data you analyze in NodeXL, investigate the following questions. Keep in mind that you should create visualizations to demonstrate each answer.

RQ1: Within the US Senators retweet network, who has the largest in-degree? What does this mean?

RQ2: Within the US Senators retweet network, who has the highest betweenness score? What does this represent?

RQ3: Within the US Senators retweet network, who has a large Twitter following but seems to be on the periphery of this network? What could this mean?

RQ4: Within the US Senators retweet network, how many connected components are there? What does this mean?

RQ4a: Who makes up the smaller component(s)? What do they have in common?

RQ5: Within the US Senators mentions network, who has the highest out-degree? What does this mean?

RQ6: Within the US Senators mentions network, how many connected components are there? What does this mean?

RQ7: Within the US Senators retweet and mentions network, does a visualization indicate clustering by political party? Please present a visual to support your answer.

(Note: For this one you will need to color the nodes based on the political party variable in the Vertices tab. I’d also suggest making the nodes bigger to see them more clearly)

RQ8: Within the US Senators retweet and mentions network, is anything notable about the Senators who are running for the 2020 Democratic nomination? Explain your findings.

RQ9: Within the US Senators retweet and mentions network, who is the most central actor? Explain your answer.

**Research Report**

You will compose a short (3-5 page\*) research report to convey the findings of your network analysis. This report should follow APA style and page format (please refer to the [Purdue OWL site](https://owl.english.purdue.edu/owl/resource/560/01/), as needed). Use 3rd person voice throughout. Your report should include three sections: 1) Research Questions/Hypotheses, 2) Data, and 3) Discussion.

1. Intro / Research Questions

In this section, you should discuss the network you explored (i.e. the US Senators network). Briefly discuss why it may be important to study Twitter communication from this group, and why examining them as a network makes sense. Then, list the research questions for this project.

2. Data

In this section, you should present the data that helped you answer your RQs. Please present important data points in text form (i.e. descriptive paragraphs) including statistics where relevant, as well as include relevant visualizations.

3. Discussion

In this section, you should discuss your results. Two major areas to discuss are:

* How did the data help you to understand the nuances of your Twitter network?
* How did the data help you understand the concepts of influence and social capital on a social media network?

You should incorporate class materials (i.e. readings on the syllabus and class discussions) into your discussion. You are not required to locate additional outside sources for this project.

\*Note: Because the number and size of network visualizations presented by students may vary, reports may be longer than 5 pages overall.

**Additional Resources**

[Dataset for Group Network Project](https://drive.google.com/file/d/1AL46UyWxD1WfANGRt3AqtuAOmQt88Alo/view?usp=sharing)

[NodeXL Notes for Students](https://docs.google.com/document/d/1s1saoatOxqsmeP5CJqQhyblK42hfcnJctm-GzDhpz2c/edit?usp=sharing)

[Opening NodeXL Files](https://docs.google.com/document/d/1eVwR3qpo2M6if-T1QyyqXaB0ObpOBuaQYytwb7Jtois/edit?usp=sharing)

[Visualizing Data in NodeXL - The Basics](https://docs.google.com/document/d/1FyZphnWD23Ob2XaaZo_fzIKMXfkeGINM5oYTayp7eEk/edit?usp=sharing)