**Testing Hypotheses and Research Questions through Social Network Analysis**

This project will use NodeXL, a social network analysis tool, to analyze a Twitter dataset related to an interest of yours. For the project, you will pose your own research questions and hypotheses, collect your own data, and use the software to analyze the data and answer your questions.

Before using the software, a class period will be devoted to 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, particularly in the context of understanding influence and social capital online. Time will also be spent explaining how to create research questions and hypotheses applicable to this project.

Then, we will spend some class periods in a computer lab learning how to use the software and interpret the date. 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 (3-5 page) research report that presents your results both in text and in network visualizations. You will need to present your RQs and Hs, “show” your data pertaining to these, and offer educated explanations about what the data supported.

Topic Selection and Research Questions

For this project you need to select a logically-bounded group of people of interest to you who use Twitter. For example, this could be all the players on a professional sports team, or all of the actors on a particular television show, or all of the Oscar-winning directors of the past 25 years. The important thing is that there is a reason for who is on the list and who is not. Ideally, the list should include 15-25 individuals. Instructor approval is required before you proceed with the project.

After you decide on the group to investigate, briefly familiarize yourself with the Twitter accounts of the individuals involved. 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 think about who seems “most central” to this group. Who is at the heart of this social network? Who is the biggest “connector”? These questions will form your first research question (RQ). Fill in the blank:

    RQ1: Within \_\_\_\_\_\_\_\_\_\_\_\_\_ Twitter network, who is the most central actor?

                 [your group’s]

Based on your previous knowledge of these Twitter users, make a prediction on who you think is most central (while this won’t be a formal scientific hypothesis, we’ll call it one for now)

    H1: Within \_\_\_\_\_\_\_\_\_\_\_ Twitter network, \_\_\_\_\_\_\_\_\_\_\_\_\_ is the most central actor.

      [your group’s]                            [your prediction]

Then, think about other questions you might investigate about this group’s Twitter usage. The NodeXL software will calculate both descriptive data (i.e. counts of tweets, retweets) and network data (i.e. measures of centrality and connectedness), so you should create questions that address both. These questions will be your additional research questions (RQ) for the project. If you have a prediction on how the data will answer these questions, you can work with your instructor on creating the proper hypotheses. You should create at least 4 additional RQs and at least 1 H that tests a relationship:

    RQ2:

    RQ3:

    RQ4:

    RQ5:

    H2:

DataPreparation

You will construct a Twitter list for the group you identified above. Ideally, the list should include 15-25 individual Twitter accounts. If you do not know how to create a Twitter list, [please consult the instructions here](https://support.twitter.com/articles/76460).

Once your list is set, you will need to import them into the NodeXL spreadsheet.

1. Open a new NodeXL template spreadsheet.
2. From the Import drop-down on the NodeXL Pro toolbar tab, select From Twitter Users Network.
3. In the dialog box section “Twitter users I’m interested in”, select the “The Twitter users in this Twitter List” option.
4. In the text box, enter the Twitter list location. This can be found in the URL of the Twitter list page on Twitter. Follow the format that the text box asks for. Be sure to eliminate the word “list” in the entry. (For example, “cspan/senators”, not “cspan/lists/senators”)
5. In the “What to import” section, select Basic network.
6. In the “Your Twitter account” section, select the appropriate button. If it’s your first time using NodeXL with Twitter, it will ask you to authorize your account. Do so.
7. In the bottom right menu, make “Limit to \_\_\_\_ recent tweets per user” to 100.
8. Unclick “Expand URLs in recent tweets”.
9. Leave all of the other fields as they are.
10. Select OK. It will take at least a few minutes before the data appears. If another dialog box appears, select OK.

You now have your dataset for the project.

Data Analysis and Visualization

You will use network analysis measures and visualizations to answer your RQs and test your H(s). Various measures will help you do this, and different students may use different measures as evidence. Explanation of the concepts behind the measures will be discussed in class and/or in assigned readings, and students could be asked to justify why they feel certain measures are best suited to address their Hs/RQs (i.e. development of a data analysis plan). The use of some more commonly used measures are described below. Explanations of the concepts behind the measures should be discussed in class and/or in assigned readings.

    Visualizing the Initial Data

1. On the top toolbar, click “Show Graph”. This will render an initial visualization. You are likely to see your vertices (dots) with some edges (lines) connecting them, as well as circular edges on each vertex.

    Adding Labels

1. To add names to your vertices, click “Autofill Columns” on the top toolbar.
2. On the pop-up menu, find “Vertex Label” and select “Name” from the drop-down.
3. Click the “Autofill” button at the bottom of the pop-up menu.
4. Click “Close” to exit the pop-up menu.

    Eliminating Self-References

1. On the spreadsheet, find the “Relationship” column.
2. From the sorting menu for that column, deselect “Tweet”. Keep “Mentions” and “Replies to” checked (you can later test each of those relationships individually).
3. Click “Refresh Graph” on the top toolbar or the visualization window.

Students should now have a visualization they can begin to make sense of. Additional steps can be taken as-needed to tailor the visualization further based on descriptive data or calculated network measures on the “Vertices” tab of the spreadsheet. Student may also explore the various layout configurations from the drop-down menu on the visualization window.

Since Twitter is not an inherently reciprocal platform, students may wish to further explore directed relationships visually by selecting “Directed” from the Type drop-down in the toolbar.

    Calculating Network Measures

1. Click the “Graph Metrics” button on the top toolbar.
2. On the pop-up menu, check all measures to be calculated. The measures likely to be helpful include: degree, in-degree, out-degree, betweenness and closeness, eigenvector, clustering
3. Click “Calculate Metrics” at the bottom of the pop-up.

The network measures will be amended to the “Vertices” tab of the spreadsheet. Students can use the spreadsheet to sort by the various measures and explore who is the most “central” to this network. These measures should enable students to answer/test RQ1/H1 as well as any additional RQs/Hs related to centrality.

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. Research Questions/Hypotheses

In this section, you should discuss the data set you chose to explore (i.e. the Twitter list that you created) and the questions and predictions you made about this data. Explain your choices and decisions in a clear and concise way.
2. Data

In this section, you should present the data that helped you answer your RQs and test your Hs. Please present important data points in text form (i.e. descriptive paragraphs), as well as at least 2 relevant visualizations.
3. Discussion

In this section, you should discuss your results. Two major areas to discuss are:
* How your data helps you to understand the nuances of your Twitter network
* How your data helps 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.