Demo Assignment 1.

The purpose of this assignment is to develop your familiarity working with network data on the NodeXL social network analysis platform. This assignment will make use of your Facebook social network data and allow you to export it into NodeXL for analysis.

These instructions assume the use of Google Chrome on a Windows PC.
You will need to install NodeXL (http://nodexl.codeplex.com/) in Microsoft Excel in Windows.

General Instructions:

This lab has 4 parts. Deliverables that you must include in your final write-up are prefaced with a roman number (I, II, III, IV) in the outline below.

Prepare an itemized report of your data analysis and deliver it to the course instructor or coordinator on paper or by e-mail. You should create one document that includes your responses to all four parts outlined below. For each analysis in your report, you should report and interpret the results as requested. Please insert your network images into your report in the appropriate places. You will be graded primarily based on the completeness and accurateness of your responses and secondarily on the clarity of the prepared report. You are allowed to form a team of the maximum of 3 students and submit a team report. Submit your reports no later than the end of the course.

Instructions for installing NodeXL on your personal machine (separate instructions will be given during the demo session if you choose to use the University PC):

NodeXL is a plug-in for Excel 2007 that creates maps from two columns of data. Each column contains a list of nodes. When two nodes are in the same row, next to each other in two columns, this creates what is called an edge, or a line between the two, in the map that is created. In this case each column will contain a list of friends' names and when two friends are in the same row it is because they are friends.

Follow these steps to install the NodeXL plug-in and save a copy of the template with a new name:

1. Download the Zip file from http://www.codeplex.com/NodeXL
2. Unzip it into any folder. Right-click the Zip file in Windows Explorer and select "Extract All."
3. Close Excel if it is running.
4. Run the "Setup.exe" file. You need to be an administrator to do this. (You don't need to be an administrator to actually use the NodeXL template once it's installed, however.) If this is the first Excel 2007 template you've installed, the setup will install some Microsoft prerequisites.
5. Open the template. In the Windows Start menu, click "All Programs" (Windows 7 or

Vista) or "Programs" (XP), then "Microsoft NodeXL," then "Excel 2007 Template."

6. Save As and give this document a new name so that you still have a blank original.

NOTE: If you are having difficulty installing/running NodeXL, please consult their tech support Discussion Board (http://nodexl.codeplex.com/Thread/List.aspx).

NOTE: If you have other technical questions when completing the steps below, please consult the NodeXL Tutorial. This detailed tutorial is likely to clarify issues and answer many of your questions. Here is the link: http://nodexl.codeplex.com/wikipage?referringTitle=Documentation&ANCHOR#Documentation

PART I: Importing network dataset from Facebook

1. If you do not have a Facebook account, use a friend's account who is not a student in this class.
2. In Chrome, go to http://apps.facebook.com/namegenweb/?fb_dash_section=my_recent and log in to your Facebook account. A notice may pop up requesting permission to access your Facebook account. Grant permission and continue. Go through a series of screens and download your network in the "GraphML" format. The PHP script will run and may take a few minutes to mine your network data. Save the document as "LASTNAME_FIRSTINITIAL.graphml" (e.g., NIKOLAEV_A, etc.) making sure that the "Encoding" option is set to UTF-8 (to ensure non-standard characters such as accents are properly imported).
3. Open the NodeXL Template (see #5 on the previous page for instructions).
4. Import the "LASTNAME_FIRSTINITIAL.graphml" file in NodeXL by going to the NodeXL ribbon in Excel, selecting Import and "From GraphML file". (If you don't see this file where you saved it, make sure the file extension is .graphml, not .txt) Select the file you created on the desktop.
5. Go to the edges tab. When two names exist next to each other in Columns A and B this indicates that these two friends of yours are also friends with each other. This also signifies an "edge" (or link/tie/connection/friendship) between two people. This tells NodeXL to create an edge between those two people in the final network map. Go to the Vertices tab. This is a list of all your friends.
    a. How many edges does your Facebook network have? How many nodes are in your network? What is the number of possible links that can exist in this network?
6. In the "sex" drop-down box, click "Sort Z to A". You likely have a few friends who did not self-report their sex which is currently listed as "None". In cell AE3, paste the following formula: "=IF(AC3="male",0,IF(AC3="female",1,99))" and press enter. Populate this formula through all the cells by double-clicking the black box on the lower right-hand side of the cell. All of the women should now have a value of 1, the men a value of 0, and the "none" a value of 99. Go through the "nones" and manually assign values. (Excel may or may not record these as text instead of integer values. If a Green dot in the upper left appears of the cell, select all these cells, click the exclamation mark floating near by and select "Convert to Number.") Change the column label in AE2 from "Column 1" to "sex2".
7. Save. To save the NodeXL file, save it as you would any other Excel file making sure to select the standard Excel Workbook (with a .xlsx extension). Do not save it as an Excel 97-2003 Workbook, a Macro-Enabled Workbook, or a Binary Workbook.
8. *NOTE: If you return to your work later, open NodeXL first, then open the file within that platform. If for some reason the file opens in a separate workbook, just copy and paste the cells into NodeXL as you did before.*

PART II: Using NodeXL to Visualize your Facebook Network

9. A Facebook network should be an undirected network. Click the "NodeXL" tab at the top of the Excel window to reveal the NodeXL toolbar. In the "Graph" box, there should be a drop-down menu next to "Type." Make sure it shows "Undirected."
10. To the right of your data there should be a blank window labeled "Document Actions." Click the "Show Graph" button in this window. If your Excel window is not maximized, you can also drag this window outside of the Excel window onto a separate part of your screen\monitor if you wish.
11. The network should load and you should see a dense jumble of links and nodes. The default layout type in NodeXL is called Fruchterman-Reingold, which is a force-directed layout algorithm (the greater the force, the more the nodes are pulled away from each other). Click the drop-down menu next to "Fruchterman-Reingold," choose "Layout Options" and increase the number of iterations to 100. Try the various other layout types from the drop-down menu next to "Fruchterman-Reingold." Press "Refresh Graph" with each new selection and wait for it to load.
12. To get a closer look at a subsection of a graph you can use the "Zoom" slider (or a mouse scrollbar in the graph pane). Once you are zoomed in you can pan across the graph by holding down the Spacebar, clicking the mouse button, and dragging the cursor in the direction you want to pan. You can also use the "Scale" slider to change the size of the vertices and edges. This can make a very dense graph more aesthetically pleasing.
13. Here are some other options to try:
    a. You can click on any row of your spreadsheet and the visualization window will highlight that link in red. Or, you can click on a node in the visualization and the corresponding row in your spreadsheet will be highlighted.
    b. You can highlight all the friendships for one individual. First, click on a node in the visualization. Next, choose the "Vertices" tab. Then, in the "Vertex 1" column click on a name. All the friendship links with that friend should now be highlighted in red.
14. Save the visualizations. Right-click on the image, choose "Copy image to clipboard," paste it into this Word document.
    a. Save two different visualizations of your Facebook network: one Fruchterman-Reingold and another of your choice. If you have zoomed in to any part of the graph, make sure you zoom out to the full graph before saving.
    b. Describe the macro-level structure of your friend graph using the Fruchterman-Reingold visualization. Is it a giant, connected component, are there distinct sub-components, or are there isolated components? What is the common feature of each component or subcomponent?
    c. Identify some of the brokers between your subcomponents. Do these surprise you? Are there any unexpected linked between subcomponents? How might these people have met or what interests could they possibly share?

d. How many isolates are in your network? Choose a few isolates and pendants (individuals with degree=1). How often do you talk or interact with these people? Did you expect them to have more links in your network? What stops you from cutting a tie to this person?

PART III: Visual analysis with NodeXL

Identifying subgroups within a network is of great interest to social network researchers. They have developed a variety of mechanisms to measure and identify subgroups which we will learn about later. However, let's use one of NodeXL's built-in tools to identify subgroups and central individuals for visual inspection.

15. Select "Autofill columns" and go to "Vertex color". Select "sex2" from the drop down box. Click the right arrow in the Options column in the "Vertex color" row and select "Vertex color options." You can leave it as "Categories" or change it to "Numbers" if you wish to customize the colors so that girls are red and boys are blue, for example. Select "OK" and then click "Autofill" and the changes should populate to your visualization. If you are having difficulty with nodes overlapping each other, try adjusting the "Scale" slider in the visualization window.
    a. Do you observe any gender-based homophily in the subcomponents of your network? Discuss potential processes that might give rise to this homophily.
16. In NodeXL, make sure you are looking at the NodeXL toolbar. In the "Analysis" box, click the "Groups" drop down and select "Group by Clusters". Then, click "Refresh Graph." The nodes in your network should now assume different colors.
17. By clicking "Find Clusters" your spreadsheet should have automatically populated the "Groups" and "Group Vertices" tabs. Now, explore and interpret the different clusters to see how NodeXL detects and divides your network up into clusters/subgroups. Click on the "Group Vertices" tab in your spreadsheet. Column A should list the cluster numbers and Column B should list the people that belong in those clusters.
    a. As you scroll down, look at the people in each group. What do they have in common? Provide reasons why these people would be grouped together? (e.g., Is it their location, school, workplace, family, other relationship, etc?) As you do this for each group, replace the default group numbers (e.g., C1245, C1378, etc.) with meaningful names. Switch to the "Groups" tab in your spreadsheet and do the same. Why did you choose these names? In your write-up, be sure to list your group names and explain your reasons for choosing those names.
    b. Pick one of these groups and look for anyone who doesn't seem to belong? Can you think of any other reason why they might be included in this cluster based on their pattern of links?
18. If your clusters are not well separated, you may need to increase the number of iterations in the Fruchterman-Reingold layout to achieve a desired graph that separates your clusters. Click on the drop-down menu next to "Fruchterman-Reingold," choose "Layout Options," and increase the iterations (try 10 and go up from there until you achieve a desirable layout).Try the repulsive force strength of 30 and adjust if needed.
    a. Save this diagram and paste it into your report.
    b. What information does this layout convey? Are the clusters well separated or is there a great deal of overlap?

    c. Describe the brokers between any components and cliques. What are common features of these brokers? Do any of these brokers surprise you? How many brokers would you have to remove from your network to "shatter" it into two or more disconnected components?

19. Click the “Graph Metrics” button in the “Analysis” box in the NodeXL toolbar. A new window will open. Click “Select All” and then “Calculate Metrics.”
20. Now go to “Autofill Columns” in the “Visual Properties” box in the NodeXL toolbar. A window will open. Click the drop-down menu next to “Vertex Size” and choose “Degree.” Click “Autofill,” wait for the graph to update. Your graph may be “lumpier” than you desire, so change so boundaries for the visualization. In the “Autofill Columns” window, click on the arrow in the “Options” column next to “Vertex Size” and select “Vertex size options”. Also try adjusting the “Scale” slider in the visualization window. Try out a few options to improve the visualization. Repeat this step one of Betweenness, Closeness, and Eigenvector centrality (selecting one at a time from the drop-down menu next to “Vertex Size.”)
    a. Save a diagram for two of the four kinds of centrality.

PART IV: Quantitative analysis using NodeXL

21. In the previous section you computed a number of graph metrics. Doing so automatically created columns in the "Vertices" and "Overall Metrics" tabs with values for these metrics.
22. Click on the "Overall Metrics" tab.
    a. What is the density of your network?
    b. Describe the distribution of degree centrality scores (just called "degree" in NodeXL) in your network? (Overall Metrics—scroll down to frequency distribution of degree.) What is the min, max, average, and median degree?
    c. Describe the distribution of betweenness centrality scores in your network.
    d. Describe the distribution of closeness centrality scores in your network.
    e. Describe the distribution of eigenvector centrality scores in your network.
    f. Describe the distribution of clustering coefficients in your network.
23. Click on the "Vertices" tab.
    a. Click the drop-down menu in the "Degree" column and select "Sort largest to smallest." Identify five people who have the highest degree centrality in the network. Interpret the results in the context of both the methodological definition of degree centrality as well as your relationship with the people with high degree centrality.
    b. Click the drop-down menu in the "Betweenness Centrality" column and select "Sort largest to smallest." Identify five people who have the highest betweenness centrality in the network. Interpret the results in the context of both the methodological definition of betweenness centrality as well as your relationship with the people with high betweenness centrality.
    c. Click the drop-down menu in the "Closeness Centrality" column and select "Sort largest to smallest." Identify five people who have the highest closeness centrality in the network. Interpret the results in the context of both the methodological definition of closeness centrality as well as your relationship with the people with high closeness centrality.
    d. Click the drop-down menu in the "Eigenvector Centrality" column and select "Sort largest to smallest." Identify five people who have the highest eigenvector centrality in the network. Interpret the results in the context of both the methodological definition of eigenvector centrality as well as your relationship with the people with high eigenvector centrality. Do you agree that these individuals are the most central? How do these measures compare to closeness centrality?
    e. Describe members of your network that exhibit high scores on each of the different kinds of centrality (degree, betweenness, closeness, eigenvector). What are common features among people within one measure of centrality? Are there individuals that appear in your top five for more than one centrality measure? Who are they? Given that each of the different centrality types exhibit potentially different roles in the network, do these network theoretic roles match up well with

their actual real life roles? Why or why not? Finally, do these central individuals tend to exist in one cluster or across multiple groups?