IT 101 Lab: Analyzing Your Facebook or Twitter Network

*Derek L. Hansen, Assistant Professor, Information Technology, Brigham Young University*

## 

## Purpose

The purpose of this lab is to help you learn how to systematically analyze social media data. You will learn to use a free Excel plugin called NodeXL to extract network data from Facebook, Twitter, or email and analyze who is important in the network and what subgroups (i.e., clusters) exist. These techniques are important in the field of “web analytics” and “social computing”.

## Procedures

To complete this lab you will need to install NodeXL (and any special importers such as the Facebook importer that you want to use); import data from your Facebook account, email, Twitter account, or a Twitter account of someone else; use NodeXL to create a meaningful network visualization of your network; and explain the network visualization using the language of social network analysis.

### Step 1. Install NodeXL and importers

If you are using a lab computer you can skip this step. If you want to use your own computer, you will need to already have the Window’s version of Excel 2007 or Excel 2010 (sorry no Mac or Linux version is available). NodeXL is a free plugin for Excel.

* Go to <http://nodexl.codeplex.com/> and download the most recent version of NodeXL

NodeXL already comes with some importers such as an email importer and several Twitter importers; however the Facebook importer must be installed from the following site if you want to use it:

* Go to <http://socialnetimporter.codeplex.com/> and follow the instructions to download and login with your Facebook credentials

### Step 2. Use NodeXL to import the data you want to analyze[[1]](#footnote-1)

Open NodeXL by typing “NodeXL” into the Start Menu search box and selecting it. Display the NodeXL Ribbon (i.e., the NodeXL tab at the top of the Excel file). On the left-hand side of the Ribbon is an Import dropdown from which you can choose the appropriate import form (i.e., Twitter User, Facebook, Email). Fill in the form to capture the data you want. Use the built-in Help feature if you need more details about the built-in importers. It may take a while for the data to download, especially if it is from Twitter, so plan your time accordingly.

### Step 3. Create metrics and calculate clusters

Calculate groups by going to the Groups dropdown and choosing “Group by Cluster Algorithm”. Calculate graph metrics using the Graph Metrics button on the NodeXL ribbon (make sure you calculate all of the metrics by checking all of the boxes). Go to the Vertexes worksheet and sort based on the different network metrics columns to see who shows up at the “top” (i.e., is the most important). Is it who you expect? Note that some people are important according to one metric (e.g., Betweenness Centrality), while others are important according to another metric (e.g., Degree).

### Step 4a. Visualize Your Network

Click on the Show Graph. Try some different network layouts from the dropdown (e.g., Fruchterman-Reingold, Harel Koren Fast Multiscale, Circle). You may want to look at the Layout Options (available via the Layout drop-down) and try some of the advanced layout features (i.e., layout by group). You may also want to use curved edges, available via the Graph Options window. Once you find a layout you like, choose None from the Layout dropdown so it won’t reposition the nodes by accident when you update the graph. Note that you can move around individual nodes by dragging them (try right-clicking and seeing other options for selecting other nodes). If your network is large, you can use the Dynamic Filters function to filter out some of the less important nodes.

### Step 4b. Fine-Tune Your Visualization

Use the Autofill Columns feature (accessible via the NodeXL Ribbon) to map various visual properties (size of nodes, color, opacity, …) to data you’ve collected (e.g., Total Followers, gender) or calculated (In-Degree, Betweeness Centrality). You can also add data directly into the Vertices and Edges worksheet columns. For example, you can use photos for some types of data (e.g., Facebook, Twitter) by choosing Image as the Shape type on the Vertices worksheet. You can also add a Legend by choosing Graph Elements dropdown on the NodeXL Ribbon and selecting Legend.

### Step 5. Create Your Final Visualization and Write-up

Your final 1-page write-up should include the following:

1. Your final network visualization (40% of grade). Save your final visualization by right-clicking on the graph pane and choosing Save Image to File🡪Save Image… To get the full points, your visualization will need to (1) use at least 3 visual properties (e.g., shape, color, opacity, edge width), (2) correctly use either a Directed or Undirected network, (3) visually display the nodes in an organized and meaningful way (e.g., choose a layout algorithm that shows the network well), and (4) appropriately use labels (e.g., group labels or labels of individuals who are important) to highlight your key observations. An example visualization will be provided.
2. An image caption (10% of grade). Add an image caption that succinctly describes what is in your network image (What do the nodes and edges represent? What do the various visual properties such as color and size indicate? Which layout algorithm was used?)
3. A paragraph interpreting your network (50% of grade). Write up a paragraph that describes what you learned from your network including: (1) Your assessment of who some of the most important individuals are in your network. Make sure you accurately describe the metrics that were used to determine who is important (e.g., betweenness centrality, degree, eigenvector centrality). (2) Your assessment of which subgroups (i.e., clusters) exist and what they mean in this context. And (3) any other insights you gained from your analysis.

1. Ideally you’ll be working with a network of between 50 and 300 nodes. Some importers let you limit the number of nodes. Otherwise you can filter some of them out later. [↑](#footnote-ref-1)