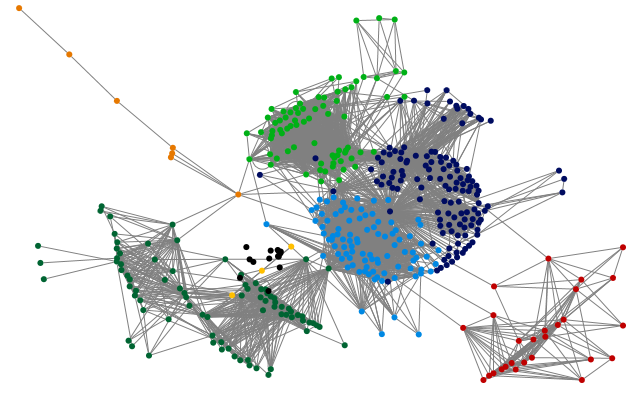
**Complex Data Analytics-Case Problem 10: Facebook Network Analysis**

**Team member names:**

# **Analyzing and visualizing Facebook ego networks**

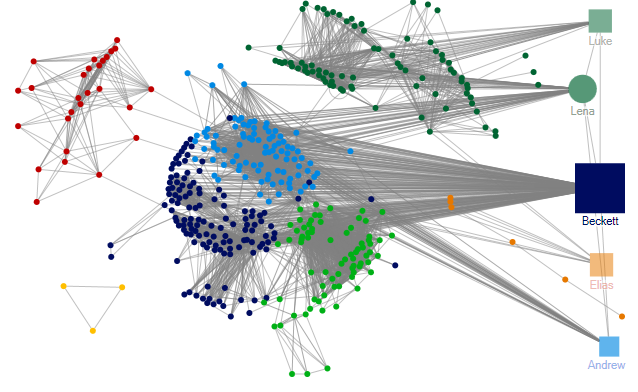
Below is a visualization of the network in ***INFO3400-Sample\_Facebook\_Egonetwork.xlsx***. You do not need to work on the xlsx file.



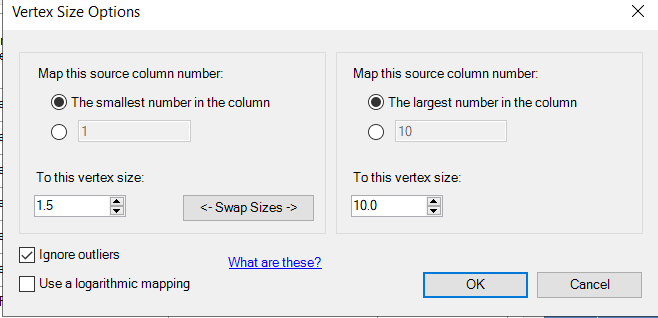
* Network clustered by the Clauset-Newman-Moore method

More visualization is customized below:

* The network is undirected
* Duplicate edges were counted and merged, and the duplications were captured by edge weights in the last column in Edges worksheet
* The edge opacity (ranging from 50%-100%) represents edge weight
* Vertex color represents strength of degree centrality
* Vertex size (ranging 5 – 100) represents strength of betweenness centrality
* Vertex opacity (ranging from 50% - 100%) represents the strength of eigenvector centrality
* The shape of vertex represents sex (Note that the shape does not represent group but sex)



**Task 1)** By observing this graph, can you pick some individuals who are mostly connecting other groups in the network? If you want to look for individuals who are important by connecting to popular individuals in the network, what feature of the node you should look at? (Note: the node with the highest degree should be the ego in the network. In the above graph, no special treatment was taken to “Beckett”. But since Beckett is the ego, you can consider hiding this node, or checking “ignore outliers” when using “Autofill” so that the high degree/betweenness centrality of Beckett will not skew the data and scale. )



# **Analyzing and visualizing user-based Facebook fan pages**

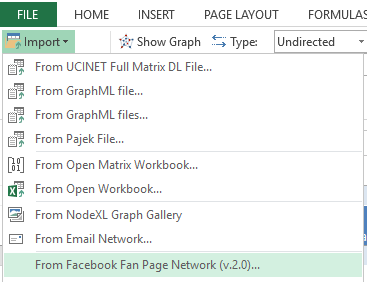
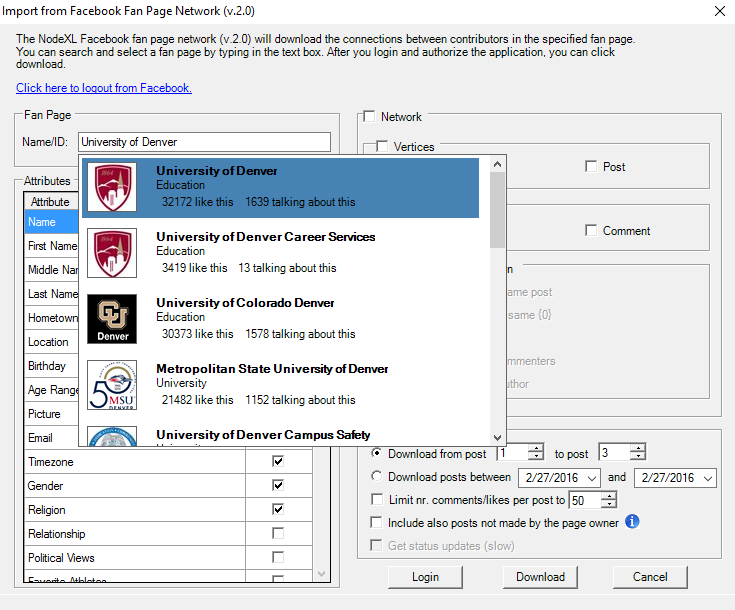
There are two basic options you can use to import the information on fan page

* Facebook User Networks (user as vertices)
* Facebook Post Network (posts as vertices)

**Task 2)** When is it appropriate to construct a user-based Facebook fan page network? When is it appropriate to construct a post-based Facebook fan page network? Think of two different business questions suitable for these two different types of networks.

We picked two Facebook Fan Pages of “University of Denver” and “University of Colorado – Boulder”

* We can create a network (of relationships) based on how a Facebook page owner’s posts are “liked” or “commented“ or “tagged” by its fans.
* By using network properties of “Co-Likers” for their (DU and CU) 5 recent posts, we may identify and compare differences between two universities’ social media campaign.
* NodeXL>Data> Import> From Facebook Fan Page Network> Using your Facebook account, login the importer so that it connects to Facebook API> Type “University of Denver” in Name/ID

--> 

* Let’s assume the follows for this instruction in the importer dialog box (select network>select only user for vertices>select only like for relationship>select only “users who liked the same post”>Options-download posts from 1 to 5.

Assumptions:

1. The network should show usernames
2. We assume that we only investigate relationship based on “Like”, not “comment”.
3. In other words, an edge will be constructed between two Facebook users if both users “***liked****”* the same post. (***Undirected***)
4. The network only considers the recent 5 posts.

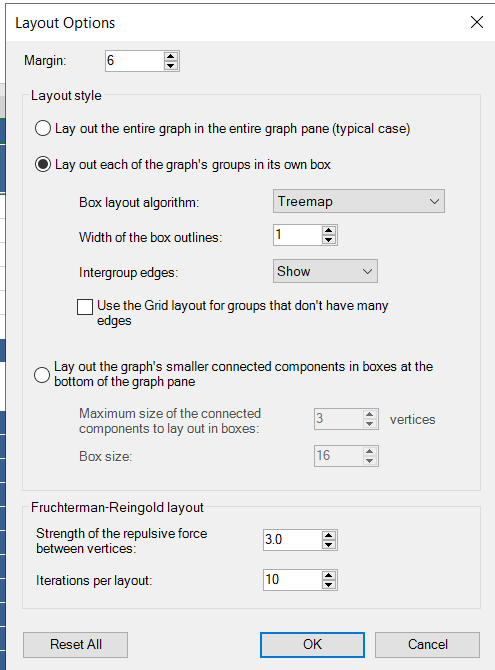
* Click “Download”, click “Yes” to the Text Wrapping notification (it would take about 2-3 minutes and loading to Excel also takes some time, be patient!
* Download “***DU\_Facebook\_CoLiker-5posts.xlsx*** and ***CU\_Facebook\_ CoLiker-5posts.xlsx*** from canvas for this practice)
* To prepare the data (prepare for the duplicate edges), File>prepare data> count and merge duplicate edges> select both count and merge>ok (make sure now you have a new field “Edge weight” in Edges worksheet.
* Perform the same for CU boulder (CU one takes a bit longer because it has a larger network, please be patient).
* For both networks, show only edges with edge weights>=2 (so that we focus on those fans who liked posts together frequently in the network).

**Task 3)** For edges whose eights are smaller than 2, will they be used in the visualization and matrices calculation?

All the followings should be done for both networks:

* Set the opacity of edges by edge weight (the opacity should range from 50%-100%).

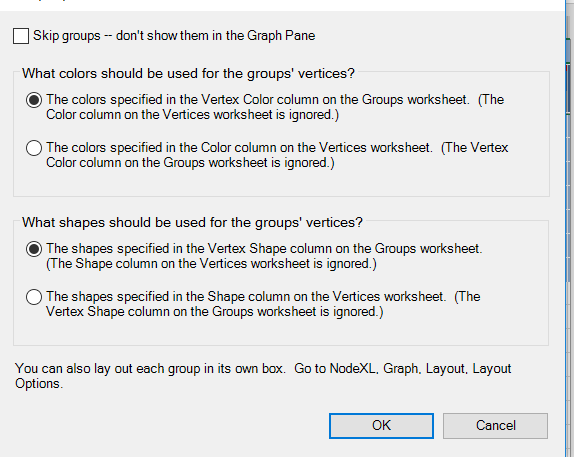
**Task 4)** In the “Autofill columns”, clear the Vertex label column (“Clear vertex visibility column now” ). Using one of grouping methods in NodeXL, visualize groups (use different colors and shapes) for both networks. In the graph, click the “Select the algorithm” dropdown list (the second icon in the top row), change the “Layout options” to “Lay out each of the graph’s groups in its own box (see below).



Copy and paste the clustered graphs for both networks below. How many groups did you find? Do you think they are quite distinguished groups or not?

* **Calculate degree, betweenness, closeness, eigenvector centralities. Then, change your network visual to reflect the follows using Autofill Columns:**

1. The strength of degree of vertex by Vertex color
2. The strength of betweenness of vertex by Vertex size (ranging 5 – 100)
3. The strength of closeness of vertex by Vertex shape (Solid Diamond if closeness>=the average of closeness in the network).
4. Set the vertex tooltip to “vertex” (when you hover over the node, the vertex name will show up)
5. Now that the color of vertices should be determined by degree, however, in Task 4 NodeXL determines the color for each group based on clustering. To make sure the degree-based color overwrites the group-based color, go to Analysis>Groups>Group options (see below). You can then make changes to the selections.



1. Do the same for the vertex shape.
2. Show only the names of the highest betweenness person (In the “Autofill column”, under “Vertex Label”, clear the column. Go to the “Vertices” worksheet, sort Betweenness centrality column in descending order, then select the first few vertices, and copy their names to the “Label” column.
3. Show only the names of the highest closeness person(s).
4. You can use any other techniques (Autofill, dynamic filter, etc.) to make this graph more meaningful/readable. You can do this by focusing on stronger relationships only (edge weight), skipping actors who only have 1-2 friends, and changing the vertex/edge visual features (size, color, width, etc.).

**Task 5)** Show the graphs for DU and CU here and, answer the following questions

* Who are those individuals (pick several, not all) mostly connecting other groups in your network?
* And who are those individuals DU and CU want to send their campaign in Facebook pages?
* Which one (between DU and CU’s frequent likers networks) has a higher density than the other? Why?
* Which network looks tighter in terms of betweenness, closeness and eigenvector? Why?

**Submit your excel files and report in Canvas.**