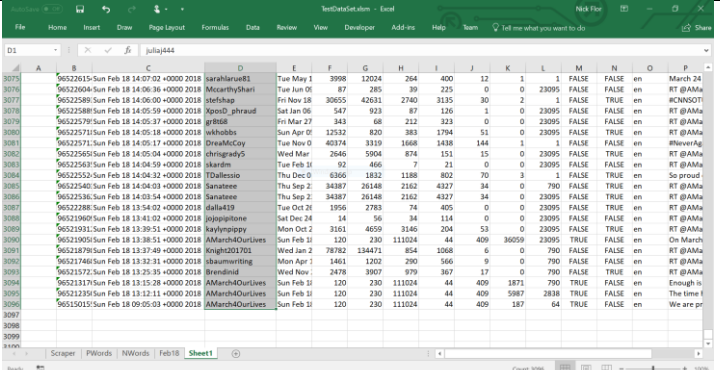
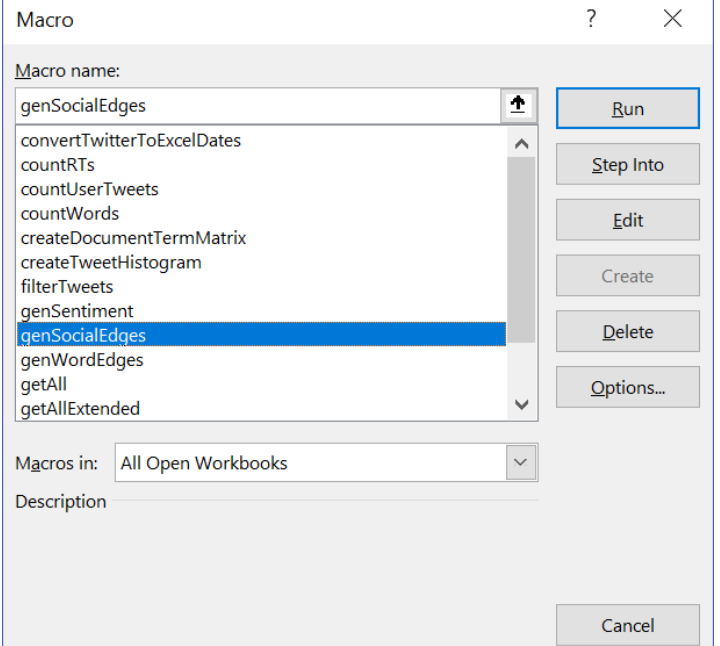


# Social Network Analysis: Clustering in SMEDA & NodeXL

Nick V. Flor, University of New Mexico (nickflor@unm.edu)

**Assumptions.** You have scraped a hashtag using SMEDA and have copied a spike into a separate worksheet.

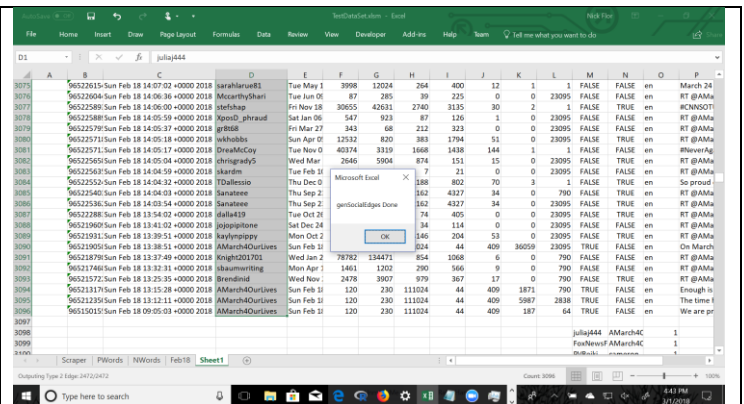
ACTION	REACTION
<b>SMEDA: Generate Social Edges</b>	
<ul style="list-style-type: none"> <li><b>Highlight Column D</b></li> </ul> <p><u>Explanation:</u> The <i>genSocialEdges</i> macro requires that you first highlight the tweeter column (D).</p>	
<ul style="list-style-type: none"> <li><b>Select Developer (menu) &gt; Macros &gt; genSocialEdges</b></li> </ul>	

(CONTINUED ON THE NEXT PAGE)

- Click the **Run** button

Note: This can take a while. Wait for the done box to pop up.

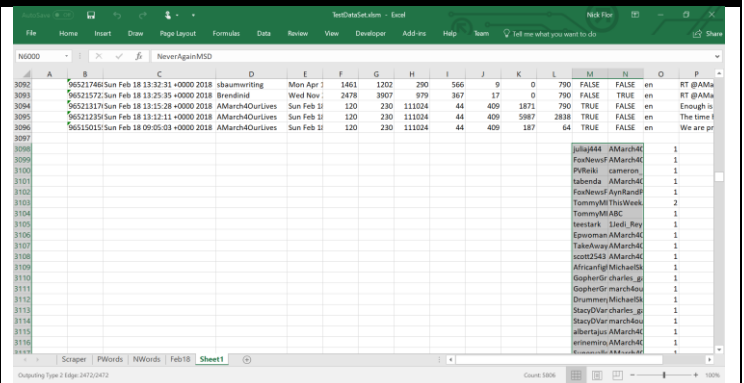
**Explanation:** This macro extracts who is tagging whom in a tweet. This is a social edge, and the macro places them in columns M & N.



### SMEDA: Copy Social Edges

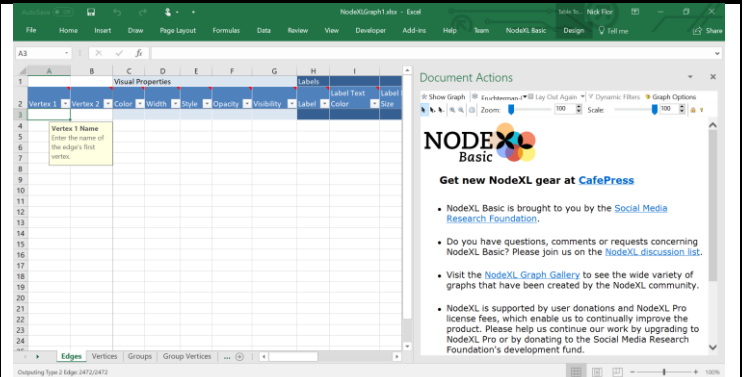
- Select all the edges in columns M & N
- Type **Ctrl-C** to copy

**Explanation:** We are going to copy these edges into NodeXL.



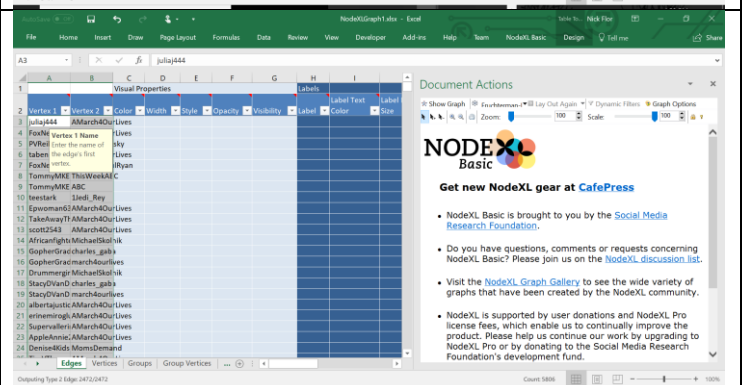
### NODEXL: Open and Paste Social Edges and Initial Graph

- Run NodeXL and make sure the **Edges** tab is selected



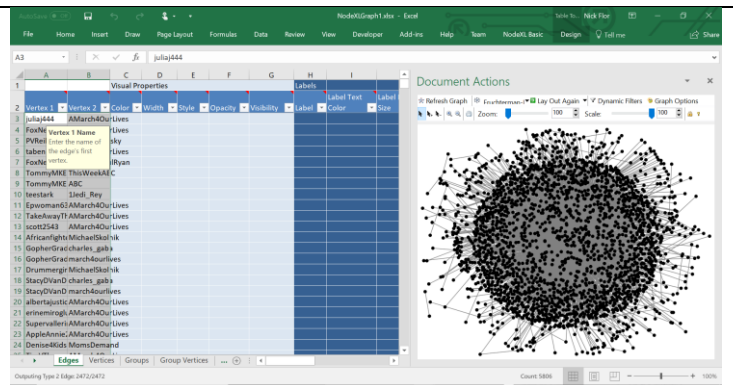
- Paste the social edges from SMEDA into the **Edges** worksheet (e.g., click on cell A3, Ctrl-V to paste)

**Explanation:** Before you can do any social network analysis you need edges.



- Click the *Show Graph* button in the right panel

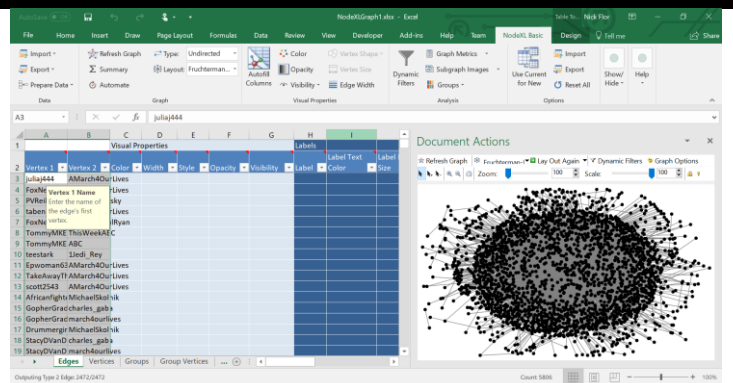
Explanation: Without any clustering or other options you get a big blob of data.



## NODEXL: Set Graph Options

- Select the *NodeXL Basic* tab

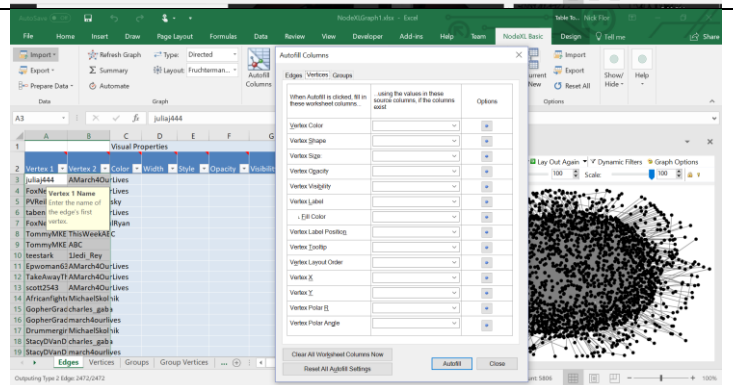
Explanation: This brings up all the NodeXL options.



- Set the *Type* dropdown to *Directed*
- Click on the *Autofill Columns* icon then **Select** the *Vertices* tab

Explanation: *Directed* draws an arrowhead to the receiver. *Autofill Columns > Vertices* allows you to set vertex properties, e.g., labels, size. We will just set the labels for now.

ICYDK: Vertices are the “circles” in the network.



(CONTINUED ON THE NEXT PAGE)

- Set the *Vertex Label* dropdown to *Vertex*

Explanation: This is where you set all the vertex properties. Later on, you should set the vertex size to the Betweenness Centrality, but I'll let you do that on your own.

Autofill Columns ✕

Edges Vertices Groups

When Autofill is clicked, fill in these worksheet columns...	...using the values in these source columns, if the columns exist	Options
Vertex Color	<input type="text"/>	➔
Vertex Shape	<input type="text"/>	➔
Vertex Size:	<input type="text"/>	➔
Vertex Opacity	<input type="text"/>	➔
Vertex Visibility	<input type="text"/>	➔
Vertex Label	Vertex	➔
Fill Color	<input type="text"/>	➔
Vertex Label Position	<input type="text"/>	➔
Vertex Tooltip	<input type="text"/>	➔
Vertex Layout Order	<input type="text"/>	➔
Vertex X	<input type="text"/>	➔
Vertex Y	<input type="text"/>	➔
Vertex Polar R	<input type="text"/>	➔
Vertex Polar Angle	<input type="text"/>	➔

Clear All Worksheet Columns Now
Autofill
Close

Reset All Autofill Settings

- Click the *Autofill* button
- Click the *Close* button

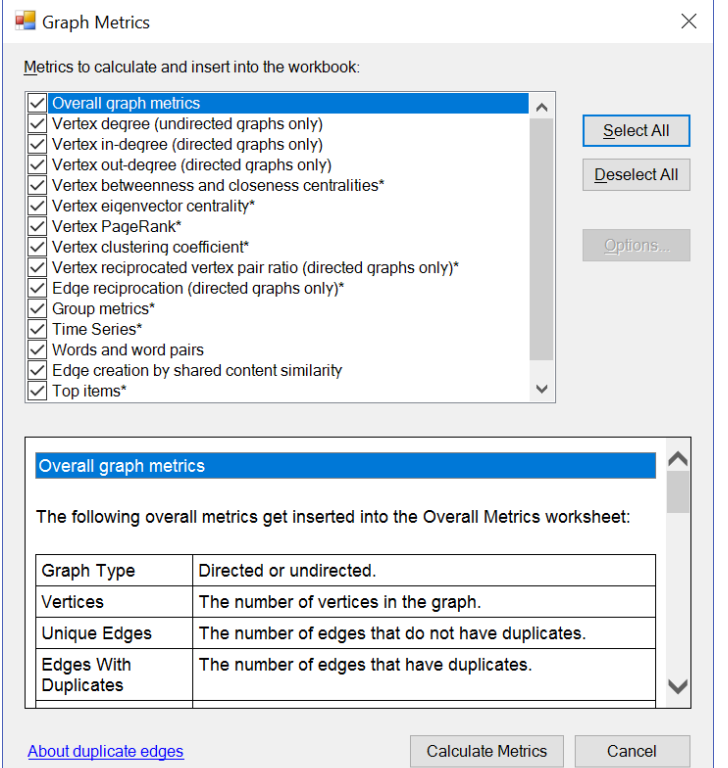
Explanation: The network nodes are now labeled. But you still have a big mess.

The screenshot shows a software interface with a spreadsheet on the left and a network graph on the right. The spreadsheet has columns for 'Vertex 1', 'Vertex 2', 'Color', 'Width', 'Style', 'Opacity', 'Visibility', 'Label', 'Label Color', and 'Label Size'. The network graph on the right is a complex, dense network of nodes and edges, with nodes labeled with names like 'Julia444', 'AMarch80ur Lives', 'PVRel', etc.

(CONTINUED ON THE NEXT PAGE)

- Select *NodeXL Basic* (tab) > *Graph Metrics*
- Click the *Select All* button

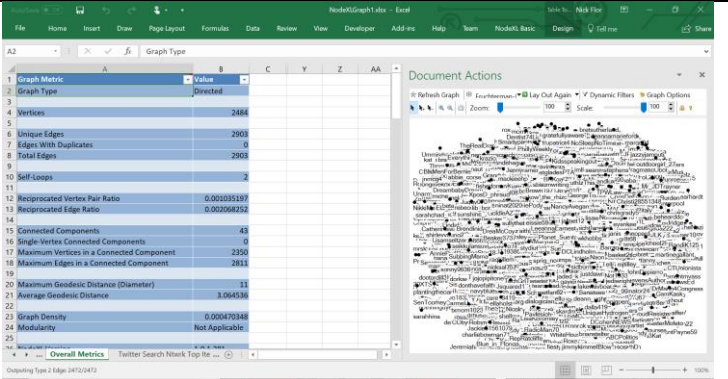
Explanation: Graph Metrics measure properties of the network, the most important of which are the various centrality measures.



- Click the *Calculate Metrics* button

Explanation: The results are placed in the Vertices tab (not shown).

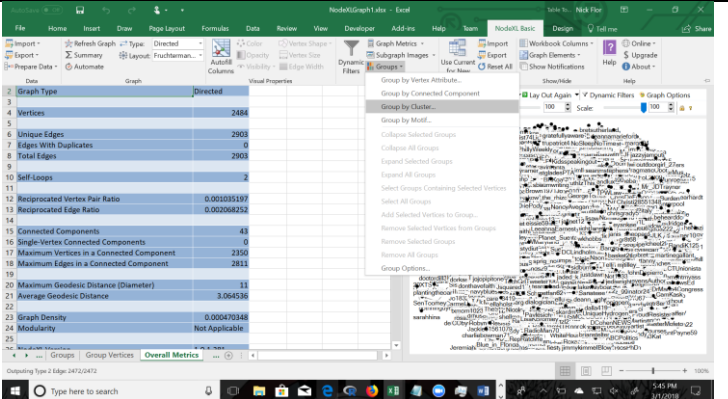
Now you're ready for clustering and more detailed analyses.



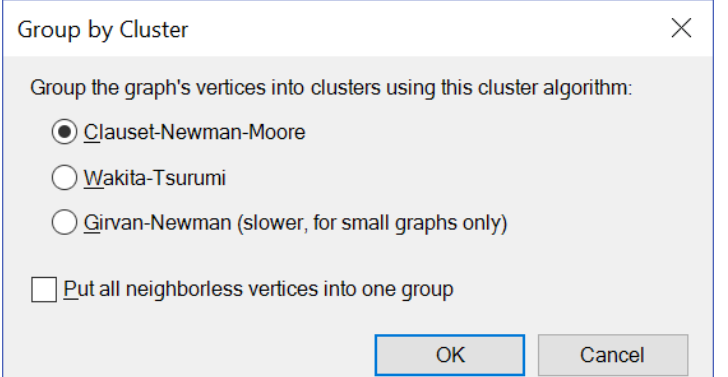
## NODEXL: Cluster

- Hover over *NodeXL Basic* > *Groups* > *Group by Cluster* ...

Explanation: This tab gives you various grouping options.

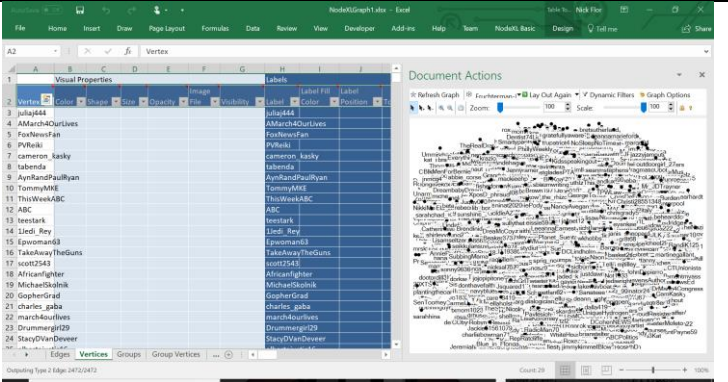


- Click on *Group by Cluster...*



- Click the *OK* button

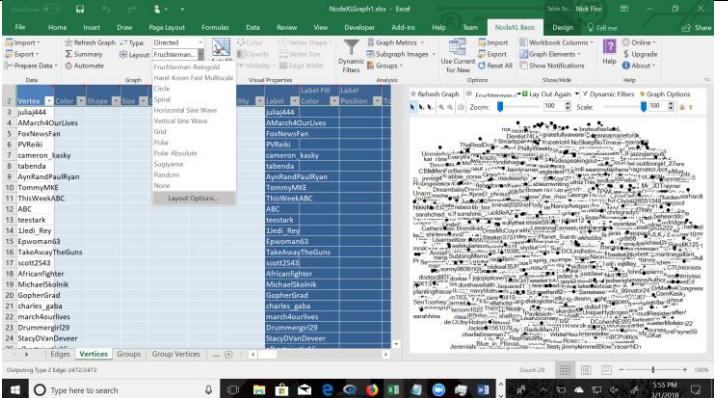
**Explanation:** It takes a while to finish, but you'll be thrown back into the main NodeXL interface. One last set of steps is necessary to see the grouping.



## NODEXL: Graph Clusters

- Hover over *NodeXL Basic > Layout > Layout Options...*

**Explanation:** This is where you set the cluster groupings



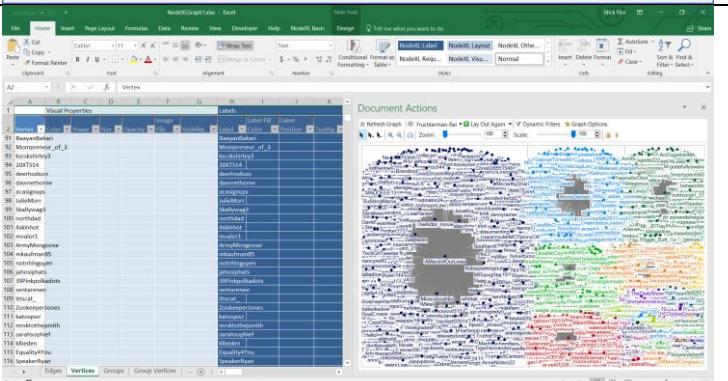
(CONTINUED ON THE NEXT PAGE)

- **Click** on *Layout Options*
- **Click** on *Lay out each of the graph's groups in its own box*
- **Select** *Combine* in the *Intergroup edges* dropdown
- **Set** *Strength of the repulsive forces between vertices* to 10.0 (you should experiment with this until you get a clear set of nodes in the center)

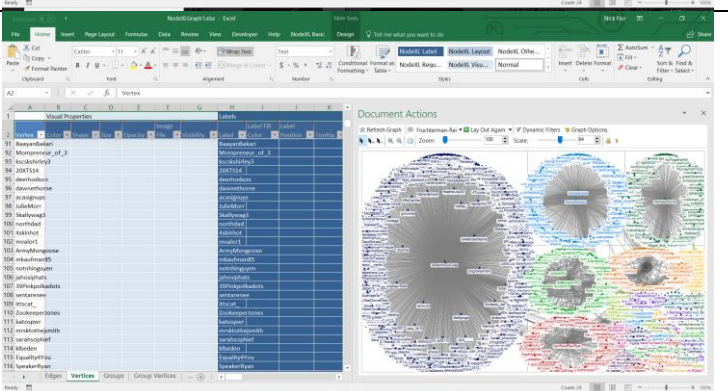
Explanation: Combining intergroup edges reduces line clutter between groups. The repulsive force will push weaker (less central) nodes to the edges.

- **Click** the *OK* button
- **Click** the *Refresh Graph* button in the right panel if you don't see any change

Explanation: You should see each cluster in its own box, with the most central (influential) nodes in the middle. It may still be hard to read, so follow the next step.



- **Click** the *Layout Again* button in the right panel
- **Continue** clicking this button until the weaker nodes are pushed to the edge. You may have to go back to layout options (see top row on this page) and increase the repulsive force)
- **Adjust** the scale slider down until you can read the central labels more clearly



## SAVING THE IMAGE

- **Right-Click** over the image > *Save Image to File* > *Save Image*

Alternatively, you can set the image dimensions by selecting *Image Options* in the pop-up.

