

Social Media Analysis CA

15% of the final module

Constructing and analysing a Twitter Network

Due Date: 24th November 2020

Purpose

The purpose of this CA is to help you learn how to systematically analyse social media data. **You will use NodeXL to extract network data from Twitter and identify important aspects of the network.**

Practical Nos. 5 and 6 will help you prepare for this assignment.

The quality of the analysis and visualisations you can produce depends on the quality of the dataset you construct, and you may find that you need to vary the search terms and topics in order to create a suitable dataset. **Because of this, you must confirm (with your professor) that your dataset is adequate, before starting the analysis.**

I am available to give you advice and feedback on (i) the dataset construction, and (ii) a draft of the report.

Step 1. Use NodeXL to import the data you want to analyse

On the left-hand side of the NodeXL 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.

Ideally, you'll be working with a network of between 300-1000 nodes (users), and 1,500-2,000 edges (tweets). Your biggest danger will be attempting to import networks which are too big (and therefore take a long time to process). If you fail to retrieve enough data, you can try a different search term or hashtag (there is no problem combining the results of multiple searches). **Remember that the dataset that you use should be agreed with me BEFORE you start your report.**

Step 2. 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 Vertices 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 3. Visualise 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 Autofill or Dynamic Filters function to filter out some of the less important nodes.

Step 4. Fine-Tune Your Visualisations and use these to understand and explain the Network

You can use the Autofill Columns feature (accessible via the NodeXL Ribbon) to map various visual properties (size of nodes, colour, opacity, ...) to data you've collected (e.g., Total Followers, gender) or calculated (In-Degree, Betweenness Centrality). You can also add a Legend by choosing Graph Elements dropdown on the NodeXL Ribbon and selecting Legend.

The different worksheets and columns produced by computing the Graph Metrics are explained in Practical 6. You should use your graphs to explain the network, eg, who are the most important people, what are the main clusters, how the clusters are different from one another and what are people talking about in each of the main clusters.

Step 5. Create Your Final Visualisations and Write-up

Your final write-up should not exceed 1500 words and should include the following:

1. A description of how you constructed your network. Describe
 - a) the type of data that you wished to analyse,
 - b) your original intentions and expectations,
 - c) how and when you extracted this information from your chosen network, (including the parameters you chose to download the data, and why),
 - d) the main characteristics of the data (eg, % of retweets/mentions etc., % of singletons, % of tweets by the main participants etc.)
 - e) Description of the difficulties/challenges you encountered during this process.

(30%)
2. Your network visualisations, with a description of the network, describing any interesting features.

Your visualisations should

- a) use at least 3 visual properties (e.g., shape, colour, opacity, edge width) that signify the characteristics of the network,
 - b) visually display the nodes in an organised and meaningful way (e.g., choose a layout algorithm that shows the network well), and
 - c) appropriately use labels and/or filters to highlight your key observations.
 - d) include image captions. 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 colour and size indicate? Which layout algorithm was used?) (30%)
3. Give an interpretation of your network. What have you learned from your network including:
 - a) 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).
 - b) Your assessment of which subgroups (i.e., clusters) exist, what conversations are taking place in the principal clusters and what they mean in this context.
 - c) Identify at least two other insights you gained from your analysis.

You can support your statements by reference to the graph metrics in the spreadsheet (but be careful not to simply present these in the report without a clear purpose).

(40%)