

Quick notes on NodeXL

Programme organisation, 3 components:

- NodeXL command 'ribbon' – access with menu-like item at top of the screen, contains all network data functions
- Data sheets – 5 separate sheets for edges, vertices, metrics etc, switching between them using tabs on the bottom. NB because the screen gets quite crowded the sheet titles sometimes disappear off to one side so use the small arrows (bottom left) to scroll across the tabs if something seems to have gone missing.
- Graph viewer – separate window pane labelled 'Document Actions' with graphical layout functions. Can be closed while working with data – to open it again click the 'Show Graph' button near the left-hand side of the ribbon.

Working with data, basic process:

1. Start in 'Edges' data sheet (leftmost tab at the bottom); each row represents one connection of the form A links to B (directed graph) OR A and B are linked (undirected).
(Directed/Undirected changed in 'Type' option on ribbon. In directed graph this sheet might include one row for A->B and one for B->A; in an undirected graph this would be tautologous.)
No other data in this sheet is required, although optionally:
 - Could include text under 'label' if the visible label on graphs should be different from the name used in columns A or B. NB These might be overwritten when using 'autofill columns'.
 - Could add an extra column at column L to identify weight of links if (as is the case with IssueCrawler data) each connection might represent more than 1 actually existing link between two vertices, hence inclusion of 'Edge Weight' in blogs data.
2. Prepare data – this button (left hand side of ribbon) has the essential function 'get vertices from edge list' automatically fills column A of the 'Vertices' sheet.
(The other function here allows merging of duplicates in the data (pay attention to directed/undirected here). May not be necessary.)
3. Now check 'Vertices' sheet, which should now have info in column A that makes sense. Won't contain anything else though.
4. 'Graph Metrics' button will fill columns in a number of sheets, including degree counts and centrality measures for individual vertices and (in the 'Overall Metrics' tab) the network as a whole.
5. This is a good point to hit 'show graph' or 'refresh graph' to see if things are making sense.

Working with groups:

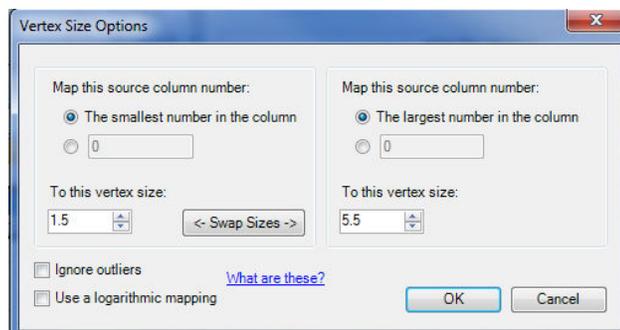
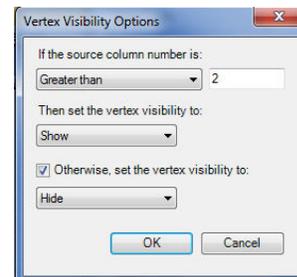
1. First need to add an attribute to vertices by which they will be grouped. Go to column AC of the 'Vertices' sheet, add a heading in row 2 (e.g. Type) and then add value for each vertex. These will (always?) be a result of manual coding.
2. Now use 'Groups' button and select 'Group by vertex attribute', choose the appropriate value in the top drop-down box in the dialogue, hit OK.
3. Now use 'Graph Metrics' button, ensure that 'group metrics' is selected; hit 'Calculate Metrics'.

4. Go to the 'Groups' sheet to see new group data. Labels G1-Gn should correspond to range of values in whatever attribute has been chosen. Group level graph density scores (col. Q) could be useful.
5. The 'Group Vertices' data sheet merely lists every vertex in the network, along with the group that it has been assigned to. Useful to check its worked as expected (Presumably, by manually filling these columns one could make groups without going through the steps of creating an attribute column etc – although it might then be harder to update or use the attribute info elsewhere.)
6. Hitting 'refresh graph' now will probably not display any differences, but may be a useful check that nothing's been broken in the process.

Making graphs with colour coding etc:

The 'autofill columns' button on the ribbon controls the visual properties of vertices and edges in the graphs. It does this by adding data in appropriate sheets. These might be in cols C-G in the 'Edges' sheet, if you want to change the look of connectors; or in cols J-O of the 'Vertices' sheet or in cols B-D of the 'Groups' sheet. These could all be set manually, although in any real, sizable network then it would seem more sensible to use the 'Autofill Columns' button. An example, imagine that we want to make a graph that excludes vertices with a degree of less than 2, adjusts the displayed node size according to degree count and includes a colour coding according to group:

1. Hit 'Autofill Columns' button.
2. In the list, go to 'Vertex Visibility' and choose 'Degree' (or 'In-degree' in directed graph). Click the arrow button in the 'options' column and choose 'Vertex Visibility Options'. Set options to look the image on the right.
3. Now go to 'Vertex Size' and choose 'Degree' (or 'In-degree'). Click the arrow button in the 'options' column and choose 'Vertex Size Options'. Set options to look the image below.



(NB Adjusting the 'to this vertex size:' values basically changes the scale of the displayed nodes, so change these there's too much or not enough difference between node sizes on the resultant graph.)

4. Now go to 'Vertex Colour' and, in the drop-down menu you should find the name of the added attribute (e.g. Type). Select this and in the dialogue box change the option reading 'numbers' to 'categories'. This will should colours automatically.

NB For our data I wanted to use black and white shapes for types. I couldn't do this through autofill, but these properties can be changed manually in the 'Groups' datasheet. Since there's only a few groups this is easy enough. So I changed 'Vertex Color' (col B) to 'Black' in every case and then clicking on the relevant cell in the 'Vertex Shape' column makes an arrow appear for a drop down list of possible shapes.

5. Now hit the 'Autofill' button at the bottom of the 'Autofill Columns' dialogue box.
6. Now hit 'Refresh graph' and all those changes should have taken place.

NB Individual display properties can be adjusted in the appropriate sheet, but I think these would be overwritten if you then used the 'autofill columns' function again (at least it would if you selected that particular property in the autofill columns dialogue).

Kevin Gillan, March 2011