Social Network Analysis Tutorial

Dr. Thanacahrt Ritbumroong
Assist. Prof. Dr. Satidchoke Phosaard
Contents
1. Social Media Mining Concepts (10 Minutes) .......................................................................................... 3
2. Social Network Analysis using NodeXL ......................................................................................... 5
   2.1. Installation ................................................................................................................................. 5
   2.2. Getting Started with NodeXL ................................................................................................. 6
   2.3. Visualizing Social Network Data with NodeXL ................................................................. 11
   2.4. Calculating and visualizing metrics ..................................................................................... 16
3. Import data from Social Media ......................................................................................................... 19
   3.1. Installation ................................................................................................................................. 19
   3.2. Import data from Facebook Personal Network ........................................................................ 21
1. **Social Media Mining Concepts (10 Minutes)**

**Exercise No. 1** Test your basic knowledge

1) What are vertices? Give example

…………………………………………………………………………………………………………

2) What are edges? Give example

…………………………………………………………………………………………………………

3) Transform the following undirected graph into a matrix and an edge list

![Graph Diagram]

No. of vertices ......................

No. of edges ......................

- incidence matrix

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- edge list

<table>
<thead>
<tr>
<th>Vertex 1</th>
<th>Vertex 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4) Transform the following directed graph into a matrix and an edge list

No. of vertices ......................

No. of edges ......................

- incidence matrix

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- edge list

<table>
<thead>
<tr>
<th>Vertex 1</th>
<th>Vertex 2</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2. Social Network Analysis using NodeXL

2.1. Installation

**Step 1**  Download NodeXL at [www.codeplex.com/NodeXL](http://www.codeplex.com/NodeXL).
Choose the correct version according to the version of your excel.

**Step 2**  If you are asked to accept the "Microsoft Visual Studio 2010 Tools for Office Runtime (x86 and x64)", click the Accept button.

**Step 3**  sdfdsf
2.2. Getting Started with NodeXL

**Step 1** Get started by opening NodeXL. To begin analyze social network data, fill in the columns in the Edges worksheet with an edge list consisting of vertex pairs that are related to each other.

![Step 1 Image]

**Step 2** An easy way to use NodeXL is to type in an edge list.

![Step 2 Image]
**Step 3**  
*Display the graph.* Click on the Read Workbook button directly above the graph pane, a network showing the party invitations is displayed.

![Network showing party invitations](image1)

**Step 4**  
*Highlighting an edge:* Click on one of the workbook rows will highlight an edge and the two vertices in the graph.

![Highlighting an edge](image2)
**Step 5**  *Manual Layout:* Click and drag the vertices one at a time to create arrangements that emphasize structures or create a more orderly display.

**Step 6**  *Automatic Layout:* NodeXL also offers several automatic Layout Types that can be selected from the control in the graph pane.
**Step 7**  
*Undirected Graph Type:* In NodeXL, the default Graph Type is Directed, which means there is a relationship between Vertex 1 and Vertex 2 that may not exist between Vertex 2 and Vertex 1. Select the Graph Type on the left-hand side of the NodeXL ribbon as *Undirected* for undirected relationships.

![NodeXL Graph Interface](image)

**Step 8**  
*Updating the Graph Pane:* Any time you change the underlying data or features that affect layout (e.g., directed versus undirected), you must click on the Read Workbook button to update the graph. If you just want to change the layout you can select a new layout type and click on Lay Out Again to reduce processing time.

**Step 9**  
*Zooming and Scale:* To get a closer look at a subsection of a graph you can use the Zoom slider (or a mouse scrollbar after first clicking somewhere in the graph pane). Once you are zoomed in you can pan across the graph by holding down the Spacebar, clicking the mouse button, and dragging the cursor in the direction you want to pan. You can also use the Scale slider to change the size of the canvas, which makes the vertices and edges get larger or smaller.
Exercise No. 2  Display a basic network using NodeXL
Key in data into NodeXL to display the graph as shown in the following
2.3. Visualizing Social Network Data with NodeXL

**Step 1**  
*Node Colors:* You may want to change the colors of nodes. For example, in the friendship graph, you might want to color nodes that represent men with blue and the women with pink. Look at the tabs on the lower left and click on the Vertices tab, which will bring up the list of 8 vertices (also called nodes). You can select colors for the 3 women and 5 men, then click on Read Workbook to redisplay the Graph Pane.

![NodeXL Graph Pane](image)

**Step 2**  
*Subgraph Images:* Another NodeXL feature is the drawing of Subgraphs for each node. This is accomplished by going to the Analysis section on the NodeXL ribbon and clicking on the Create Subgraph Images option. Note that this feature can be computationally intensive for large data sets.

![Create Subgraph Images dialog](image)
**Step 3**  **Fixing Node Placement:** If you want to the placement of the nodes to remain fixed in their position, find the “Locked?” column on the Vertices tab and choose “Yes (1)” (or just “1”) for each of the vertices.

**Step 4**  **Adding Descriptive Data:** If you have additional information about the people in the data set, you can add your own columns of data by typing (or pasting it in). You might want to record the age of each person, so scroll the Vertices worksheet to the right until you see the column header “Add your own Columns Here”. Place the cursor on this header to get further instructions. If you select the next free column, you can type an attribute name (e.g., Age) and then enter values for each person. In figure below, we have added two new columns, one for Age and one for number of Prior Parties the individual has attended since the beginning of the year.
Step 5  Changing Vertex Size (and other properties): Another visual property that can be used to encode attribute values is node size, which is controlled by the Radius column in the Vertices worksheet. Put your curser over the Radius column header to show the type of data that must be entered – in this case numbers 1-10. Use this same approach to see what type of data to enter into any of the different fields such as Shape, Color, and Opacity.

Step 6  AutoFilling Columns: Click on the AutoFill Columns button in the NodeXL ribbon. The resulting Dialog box offers a set of drop-down boxes to allow you to select data you have entered in as additional fields. Click on the downward pointing arrow next to Vertex Radius to see all of the data columns you have entered in and choose Prior Parties (instead of Age). Note that the column data won’t show up until after you click on Read Workbook.
Each attribute has an associated Options page that allows you to fine-tune some of the attributes. In our example, we want to assure that the Nodes are large enough to view well, so we can click on the “…” button in the Options column for the Vertex Radius row. Here we can change the smallest number in the column from the default value of 1.0 to 2.0 so the nodes will all be sufficiently large to see well.

![Vertex Radius Options](image)

**Step 7** *Adding Labels:* AutoFill Columns feature can also be used to fill the Primary Labels column with the names from the Vertex column. The color coding remains but the size coding is no longer possible.

![Graph with labels](image)
Step 8   *Adding Tooltips:* You can also add data that only shows up when you mouse over a vertex. This is called a Tooltip.

**Exercise No. 3**   Visualize your network using NodeXL
Use NodeXL to display the graph as shown in the following
2.4. Calculating and visualizing metrics

**Step 1**  *Selecting Graph Metrics:* You must first choose what graph metrics you want to calculate. To do this, click on the downward pointing arrow next to the Calculate Graph Metrics button on the NodeXL ribbon and then choose Select Graph Metrics. This will open up the dialogue box that shows you the available graph metrics. Select the ones you want to calculate by checking in the boxes next to them. Clicking on the Details link next to them will explain what the metrics mean. For now, click on the Select All button and choose OK.

![Select Graph Metrics Dialogue Box]

**Step 2**  *Calculating Graph Metrics:* To calculate the graph metrics you will need to then click on the Calculate Graph Metrics button in the NodeXL ribbon. Some of the graph metrics can take a while to calculate when working with large networks. When all of the graph metrics are done calculating, NodeXL adds new columns will be added to the Edges and Vertices tabs. NodeXL also adds content to the Overall Metrics tab.
**Overall Metrics:** This tab summarizes some of the key properties of the entire network including the graph type (Undirected vs Directed), the number of unique edges, edges with duplicates, and total edges. For some networks the Edges with Duplicates is an important measure since two vertices may be joined by more than one edge.

The Self-Loops row shows the number of vertices that link to themselves. A self-loop occurs when the edge list includes the same exact name in the Vertex 1 and Vertex 2 columns on the Edges tab (i.e., a person is connected to themselves).

The tab also shows the total number of Vertices, the NodeXL version that it was created using, and the Graph Density. The Graph Density is a number between 0 and 1 that is an overall measure of the interconnectedness of the vertices.

For a undirected graph where all vertices are connected to all others through at least one edge, the Graph Density is calculated by dividing the number of Total Edges by the maximum number of possible edges, which equals 0.4 for the example. A more dense graph (e.g., 0.6) would include more Total Edges, while a more sparse graph (e.g., 0.2) would include fewer.

*Degree, Centrality, and Clustering Coefficients:* Go to the Vertices tab. You will notice there are several new columns that were inserted next to the Vertex column after you Calculated Graph Metrics. Each of these metrics relate directly to one of the vertices.

**Degree:** The Degree of a vertex (sometimes called Degree Centrality) is a count of the number of edges that are connected to it. If we were using an undirected graph (such as the Party Network), the single Degree metric would be split into two metrics: (1) In-Degree, which measures the number of edges that point toward the node of interest, and (2) Out-Degree, which measures the number of edges that the node of interest points toward.

**Betweenness Centrality:** Vertices that are included in many of the shortest paths between other vertices have a higher Betweenness Centrality than those that are not included. In a case where Betweenness Centrality is 0, if this person was removed from the graph everyone would still be connected to everyone else and their shortest communication paths would not even be altered. High Betweenness Centrality indicates that the person acts as a “bridge” in passing information.

**Closeness Centrality:** Another characteristic you may care about is how close each person is to the other people in the network. If information flowed through edges in the network, some people would be able to contact all the other people in only a few steps, while others may require many steps. Closeness Centrality is a measure of the average shortest distance from each vertex to each other vertex.

**Eigenvector Centrality:** In many cases, a connection to a popular individual is more important than a connection to a loner. The Eigenvector Centrality metric attempts to take into consideration not only how many connections a node has (i.e., its Degree), but also the Degree of the nodes that it is connecting to.
**Clustering Coefficient:** In some cases, a person’s friends may be friends with each other, creating a clique. The Clustering Coefficient measures how connected a node’s neighbors are to one another. More specifically, it is the number of edges connecting a node’s neighbors divided by the total number of possible edges between the node’s neighbors.

**Exercise No. 4** Calculate network metrics using NodeXL

Use the following edge list, visualize the network and determine its metrics

<table>
<thead>
<tr>
<th>Vertex 1</th>
<th>Vertex 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andre</td>
<td>Beverly</td>
</tr>
<tr>
<td>Andre</td>
<td>Diane</td>
</tr>
<tr>
<td>Andre</td>
<td>Carol</td>
</tr>
<tr>
<td>Andre</td>
<td>Fernando</td>
</tr>
<tr>
<td>Beverly</td>
<td>Diane</td>
</tr>
<tr>
<td>Beverly</td>
<td>Ed</td>
</tr>
<tr>
<td>Beverly</td>
<td>Garth</td>
</tr>
<tr>
<td>Carol</td>
<td>Diane</td>
</tr>
<tr>
<td>Carol</td>
<td>Fernando</td>
</tr>
<tr>
<td>Diane</td>
<td>Ed</td>
</tr>
<tr>
<td>Diane</td>
<td>Fernando</td>
</tr>
<tr>
<td>Diane</td>
<td>Garth</td>
</tr>
<tr>
<td>Ed</td>
<td>Garth</td>
</tr>
<tr>
<td>Fernando</td>
<td>Garth</td>
</tr>
<tr>
<td>Fernando</td>
<td>Heather</td>
</tr>
<tr>
<td>Garth</td>
<td>Heather</td>
</tr>
<tr>
<td>Heather</td>
<td>Ike</td>
</tr>
<tr>
<td>Ike</td>
<td>Jane</td>
</tr>
</tbody>
</table>

Unique edges ……………………

Total edges ……………………

Graph density …………………

Who has the lowest degree centrality …………………………………………………

Who has the highest Betweeness Centrality ……………………………………………

Who has the highest Closeness Centrality ……………………………………………

Why Ed has higher degree of Eigenvector Centrality than Heather …………………

Why Ed has the clustering coefficient of 1 ……………………………………………
3. Import data from Social Media

3.1. Installation

**Step 1** Go to [http://socialnetimporter.codeplex.com/](http://socialnetimporter.codeplex.com/)


**Step 3** Close NodeXL and copy these files to the NodeXL Plug-ins Directory (C:\Program Files\Social Media Research Foundation\NodeXL Excel Template\PlugIns)
**Step 4** Restart NodeXL: you should see the Facebook Import option in the NodeXL>Data>Import menu.
3.2. Import data from Facebook Personal Network

**Step 1** In excel, Click NodeXL ➔ Import ➔ From Facebook Personal Network. A small window will pop-up.

![NodeXL Import from Facebook Personal Network window](image)

**Step 2** Click login and login to your Facebook account. Then, click Download.