UNIT – II
NodeXL

By

Ranjith Kumar M

Assistant Professor, CSE

SRITW

Content

- **NodeXL**
- **Layout**
- **Visual Design**
- **Labeling**
- **Calculating and Visualizing Network Metrics**
- **Preparing Data and Filtering**
- **Clustering and Grouping**

NodeXL

- The NodeXL Template for Microsoft Excel is an add-on to the spreadsheet application that provides a range of basic network analysis and visualization features.

- It is maintained by the Social Media Research Foundation (https://www.smrfoundation.org), a non-profit organization dedicated to the provision of tools and research to aid in "mapping, measuring and understanding the landscape of social media."

- NodeXL's central goal is ease of use, making it ideal for students and professionals to learn the concepts and methods of social network analysis with visualization as a key component.

NodeXL

- NodeXL supports the calculation of frequently used network
- metrics, grouping (i.e., clustering) of vertices, and textual analysis, though it does not support all of the advanced analysis techniques available in research-oriented tools such as Pajek, UCINet, or R.
- As a network visualization tool NodeXL is unparalleled, supporting a variety of network layouts, manual control of vertex placement, labeling of vertices, edges, and groups, and visual properties such as color, opacity, size, tooltips, images, etc. NodeXL is also unique in its ability to help analysts capture a range of social media network datasets without any programming

NodeXL

- **Getting Started with NodeXL**

- Download NodeXL from
  https://www.smrfoundation.org/nodexl/installation/

- open a new NodeXL file by clicking on the NodeXL Excel Template in the Start Menu.

- This will open a blank NodeXL file. NodeXL files end in the .xlsx extension just like Excel files, but when you open them, they will load the NodeXL add-on features like the graph display pane and custom menu items.

- Display the NodeXL menu ribbon clicking on the NodeXL tab in the Excel menu ribbon.

NodeXL

- The NodeXL menu provides access to all of the NodeXL features from a single place.

NodeXL

- **Spreadsheet and graph pane**
- NodeXL is composed of two major sections: the spreadsheet where data is stored and the graph pane (called Document Actions) used to display network visualizations.
- The spreadsheet workbook includes multiple worksheets, each of which contains a different type of network data.
- For example, the Vertices worksheet contains a row for each vertex (e.g., each person in a network), while the Edges worksheet contains a row for each edge (e.g., each connection between 2 vertices).
- When NodeXL starts up, the graphpane shows a default splash screen until a network is visualized.

NodeXL

NodeXL

- **Manually entering data**

- One way to begin using NodeXL is to manually type in an edge list.

- Navigate to the Edges worksheet and enter in the names.

- Each row of the edge list represents a single edge between the two people specified in the Vertex1 and Vertex2 columns.

- The fictional data represents LinkedIn connections between a subset of employees of a large company we'll call Analyzing Big Complex Data (ABCD for short).

- These connections are undirected; that is, the first row showing that Ben is connected to Ava also implies that Ava is connected to Ben.
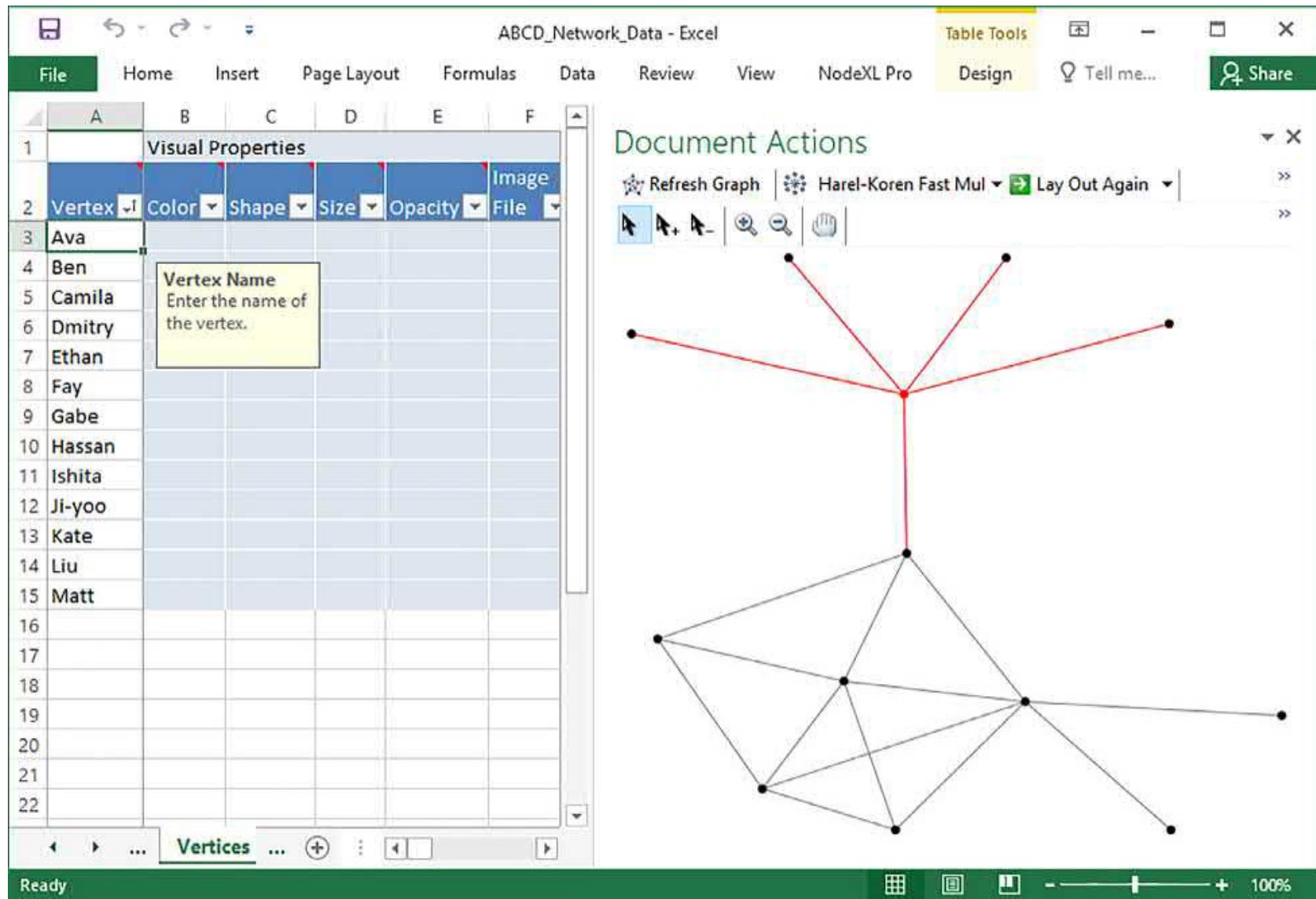
NodeXL
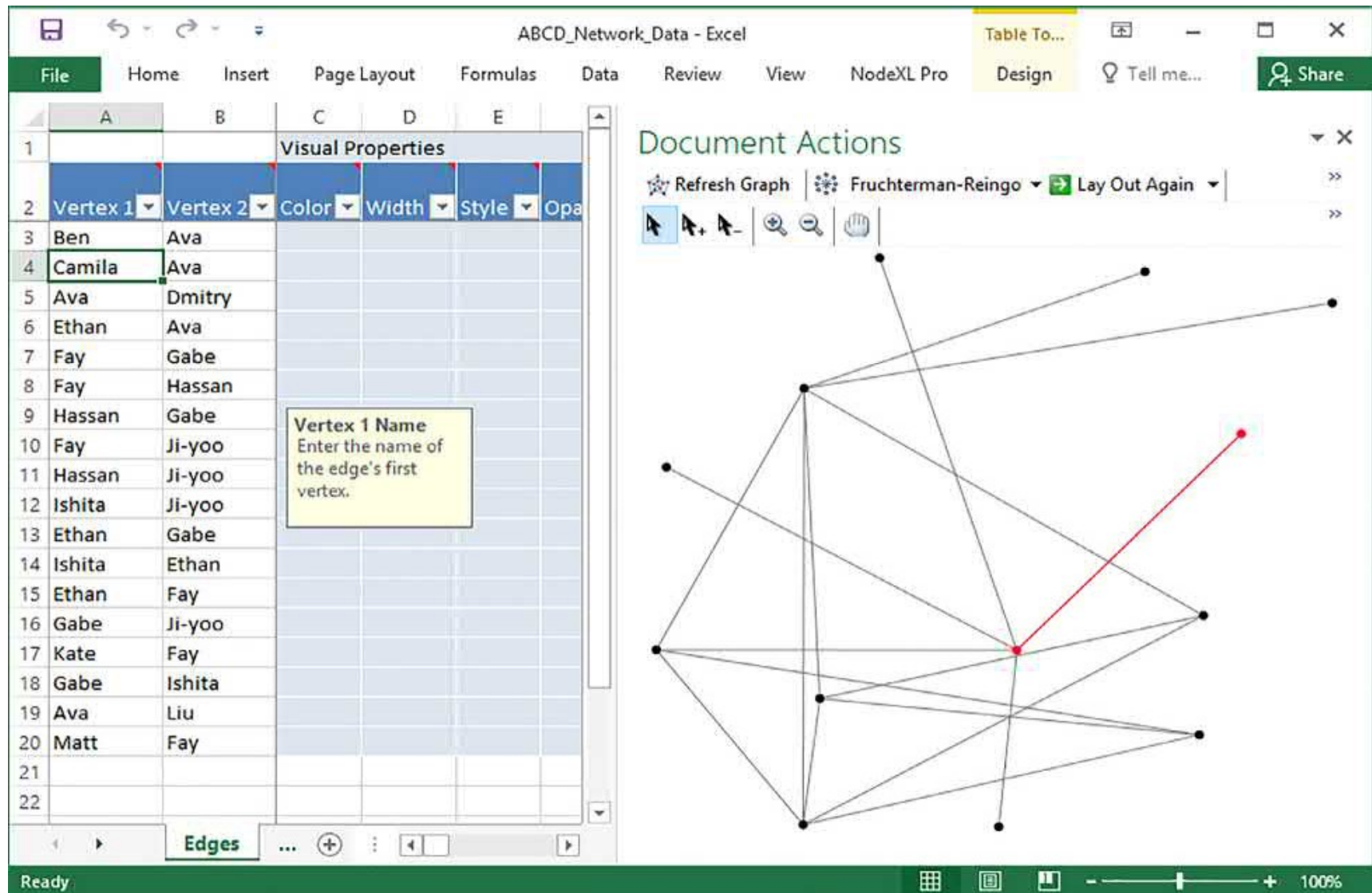
NodeXL

- **Importing data**
- NodeXL can import data from a variety of formats and data sources.
- To import data, choose the data source on the Import drop-down available in the Data section of the NodeXL Ribbon.
- The options at the top allow you to import data from other network analysis programs such as UCINET and Pajek, as well as files from the generic GraphML network data format.
- You can also open data from another opened excel file that may include an edge list (From Open Workbook…) or an adjacency matrix (From Open Matrix Workbook…)
- A variety of built-in or third party data importers can be used to dynamically download data from social media sites such as Twitter and YouTube or from email collections

NodeXL

- **Showing the graph**

- Next, click on the Show Graph button (directly above the graph pane) to visualize the network.

- This will change the text on the button to Refresh Graph.

- It will also automatically create a row on the Vertices worksheet for each of the 13 unique people.

- This operation can also be manually performed without displaying the graph visualization by selecting the NodeXL>Data>Prepare Data>Get Vertices from Edges Worksheet function.

- Use this approach when working with very large datasets that need to be filtered before they are visualized.

NodeXL

NodeXL

- **Highlighting an edge or vertex**

- The data in the spreadsheet and the graph pane are coupled so that clicking on a row in the Edges worksheet will highlight the corresponding edge in the graph.

- Vertices can be highlighted by selecting a specific row in the Vertices worksheet.

- And visa versa, clicking on a specific vertex in the graph pane will automatically highlight the corresponding row in the spreadsheet. You can even click on multiple rows to highlight all related edges or vertices.

NodeXL

- **Resizing and moving the graph pane**

- The graph pane can be resized by dragging the lefthand side of the pane to the left or right.

- It can also be undocked from the spreadsheet by clicking on the Document Actions title and dragging it outside of the Excel window.

- This is recommended when working with multiple monitors.

NodeXL : Layout

- **Layout: Arranging vertices in the graph pane**

- The location of vertices in a network visualization can make a significant difference in how easy it is to understand and gain insights into a network. For example, compare the visualizations of the ABCD network.

- These display the same network, but the layout in Figure 4.4 is far more readable.

- Creating readable graphs can be challenging, particularly for large networks.

- However, there are some useful heuristics to help improve your graphs, NodeXL provides numerous techniques to manually and automatically adjust network layouts.

NodeXL : Layout

- **Manual layout**

- You can manually position vertices to create arrangements that emphasize structures and create a more orderly display.

- You can even select multiple vertices by drawing a box around them or clicking on additional vertices while holding down the Control key.

- Multiple selected vertices move together when selected.

- Manually adjusting vertices is typically done to fine-tune existing networks that have been arranged automatically according to a layout algorithm.

NodeXL : Layout

- **Automatic layout**

- NodeXL includes several automatic layout algorithms that can be applied to position all of the vertices according to a set procedure.

- In our example of the ABCD network, the Fruchterman-Reingold layout with its default settings is not particularly useful as it includes unnecessary edge crossings and a general sense of disarray that makes it difficult to identify important individuals.

- Use the drop-down menu directly above the graph pane or in the NodeXL Ribbon and select the Harel-Koren Fast Multiscale algorithm.

- Experimenting with different layout types (e.g., Circle, Sugiyama) can reveal useful patterns, relationships, or unusual features when you analyze a dataset.
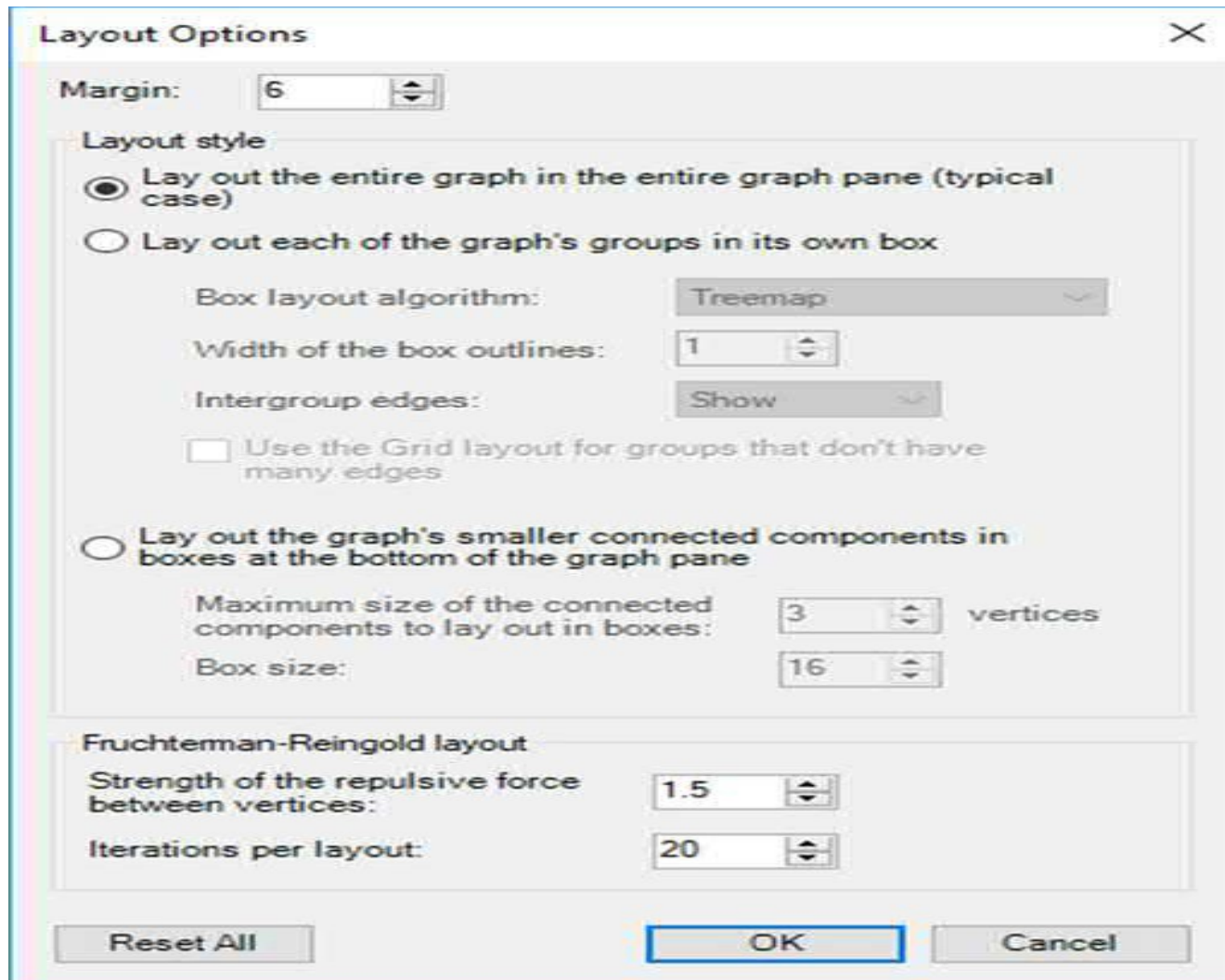
NodeXL : Layout
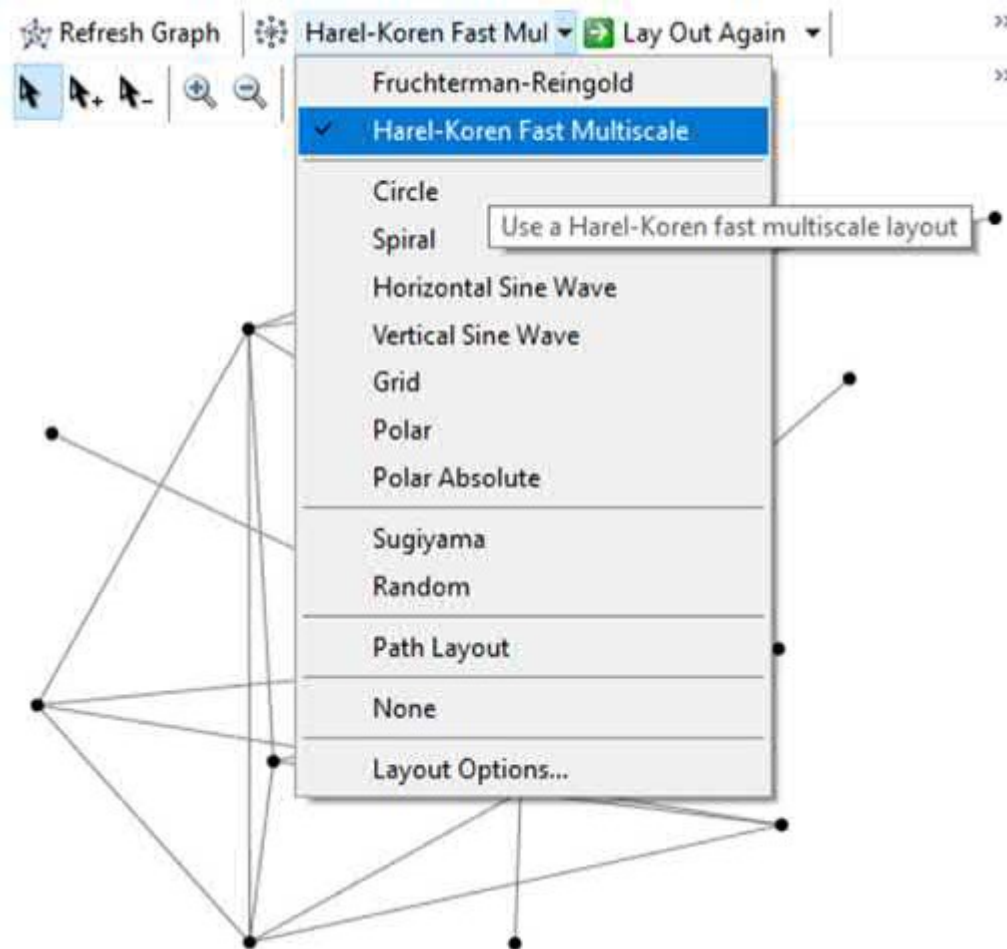
NodeXL : Layout

- **Adjusting Fruchterman-Reingold settings**

- The Fruchterman-Reingold Layout works well for many large social networks, though it may require some adjustment.

- It is an example of a force-directed algorithm, which uses an analogy of physical springs as edges that attract connected vertices toward each other and a competing repulsive force that pushes all vertices away from one another, whether they are connected or not.

- Adjust the Fruchterman-Reingold default parameters by selecting Layout Options…, which will open the Layout Options dialog.

- Set the Strength of the repulsive force between vertices to 1.5 and Iterations per layout to 20.

NodeXL : Layout
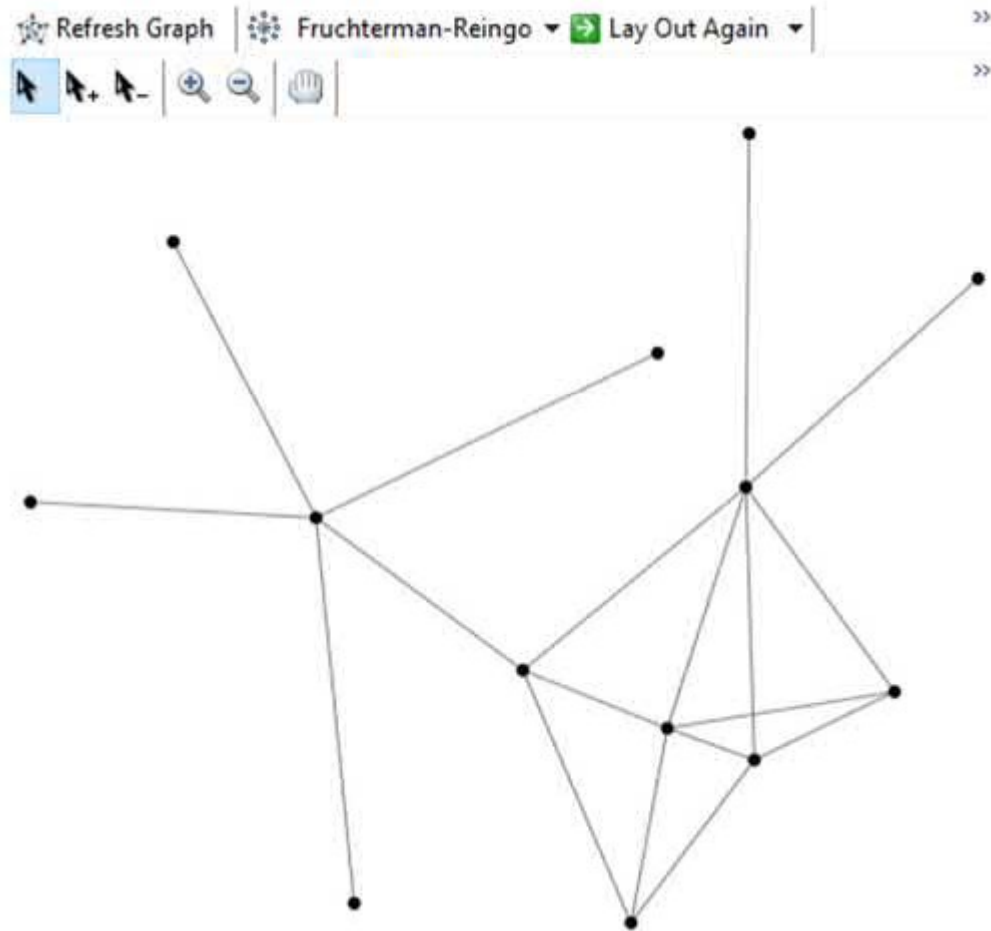
NodeXL : Layout

NodeXL : Layout

- Choose the Fruchterman-Reingold layout and click on Refresh Graph to see an updated version of the graph that looks much more readable than the original, though perhaps still not as clear as the Harel Koren Fast Multiscale algorithm.

- **Updating the graph pane**

- Any time you change the underlying data or features that affect the layout of the network (e.g., directed vs undirected), you *must click on the Refresh Graph 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.

- Note that the Fruchterman-Reingold layout will start from the current layout if you choose Lay Out Again, whereas it will start from an initial seed position if you choose Refresh Graph.

NodeXL : Layout

- **Undirected and directed graph type**

- So far, we have thought of the ABCD network as undirected.

- The edges were said to represent mutual friendship connections on LinkedIn.

- Thus, an edge either exists (for the pair of vertices it connects) or does not exist.

- Many edges are undirected, such as Facebook Friendships, affiliations (e.g., both people were tagged in a photo together), or mutual relationships (e.g., spouse).

- In contrast, other networks are directed, meaning that a vertex "points" to another vertex because of the nature of the link.

NodeXL : Layout

- **Undirected and directed graph type**

- Imagine that the ABCD network instead represents ABCD employees who have "endorsed" other ABCD employees on LinkedIn.

- The ABCD network would then be directed.

- Other directed networks are created when information is passed between people (e.g., a message is sent from one person to another), subscription relationships (e.g., YouTube channel subscriptions; Twitter follower networks), invitation networks, etc.

- Notice that a single social media platform, such as LinkedIn, may have many different types of edges—thus many different networks, even among the same set of people.

NodeXL : Layout

- **Undirected and directed graph type**

- Visually, directed networks are displayed with arrows that point from the source vertex to the destination vertex.

- Additionally, certain network metrics will be calculated differently for undirected and directed networks.

NodeXL : Layout
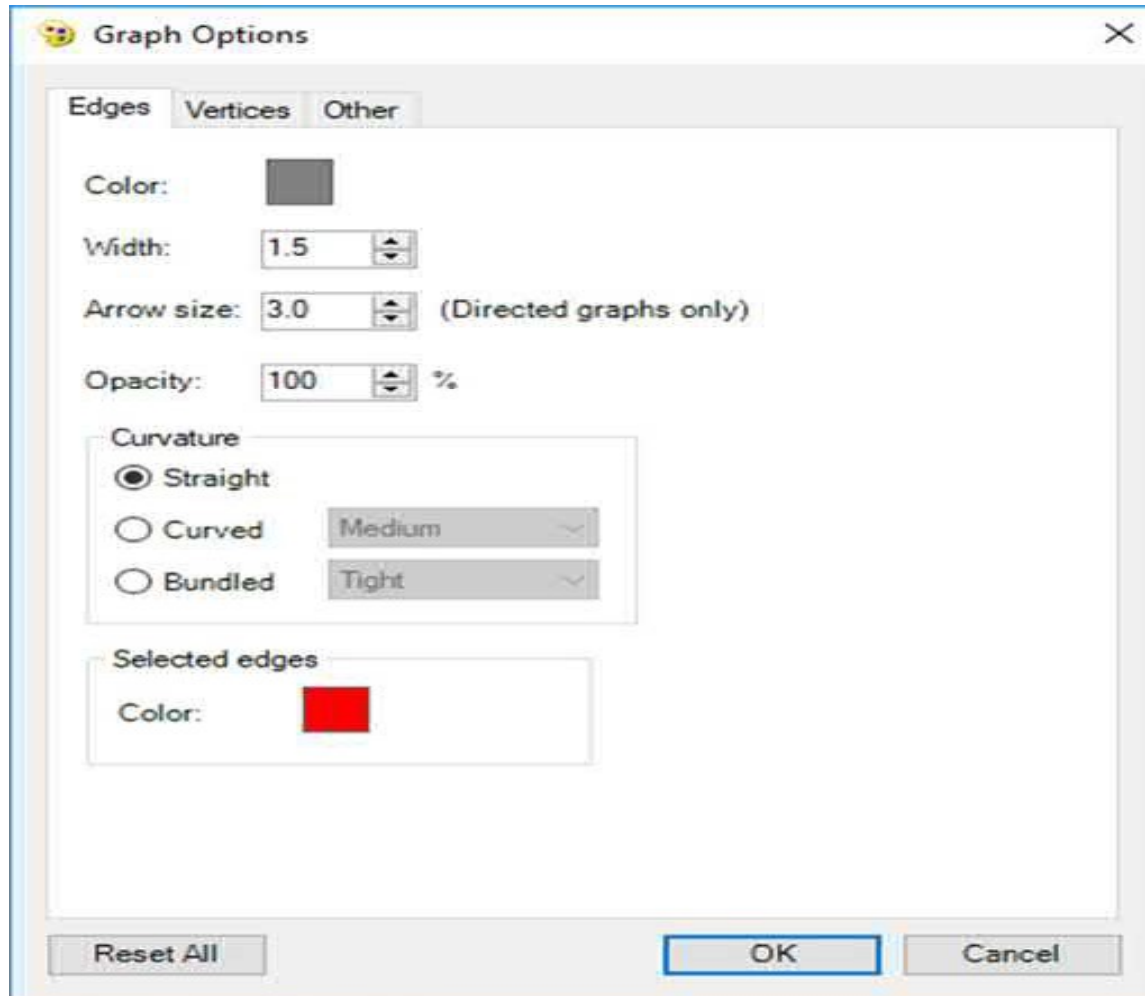
Working with NodeXL files

- **Saving NodeXL files**

- To save the NodeXL file you have been working on, simply save it as you would any other Excel file making sure to select the standard Excel Workbook (with a .xlsx extension).

- Do not save it as an Excel 97-2003 Workbook, a Macro-Enabled Workbook, or a Binary Workbook.

- All data, including the most recent layout positions will be saved. Save your ABCD network file now.

Working with NodeXL files

- **Exporting an existing NodeXL file**

- In addition, you can export NodeXL file data in several ways, accessible from the Export menu found in the Data section of the NodeXL Ribbon.

- Export options for common network file types such as UCINET, Pajek, GraphML, GEXF, and GDF are available.

- You can also export to an email, an automatically generated Power Point presentation, to the online network visualization tool called Polinode, or to the NodeXL Graph Gallery (Advanced topics: NodeXL graph gallery).

- Use the Export Options dialog to add metadata about your file.

Working with NodeXL files

- **Opening an existing NodeXL file**

- You can open a NodeXL file in the same way you would normally open any Excel file.

- If NodeXL is installed, the NodeXL Ribbon and graph pane will automatically open when the file is opened.

- **Opening a NodeXL file created on another computer**

- When opening a file that was created on an older version of NodeXL, you should open a blank NodeXL file and then use the Import>From NodeXL Workbook Created on Another Computer feature available via the NodeXL Ribbon.

Working with NodeXL files
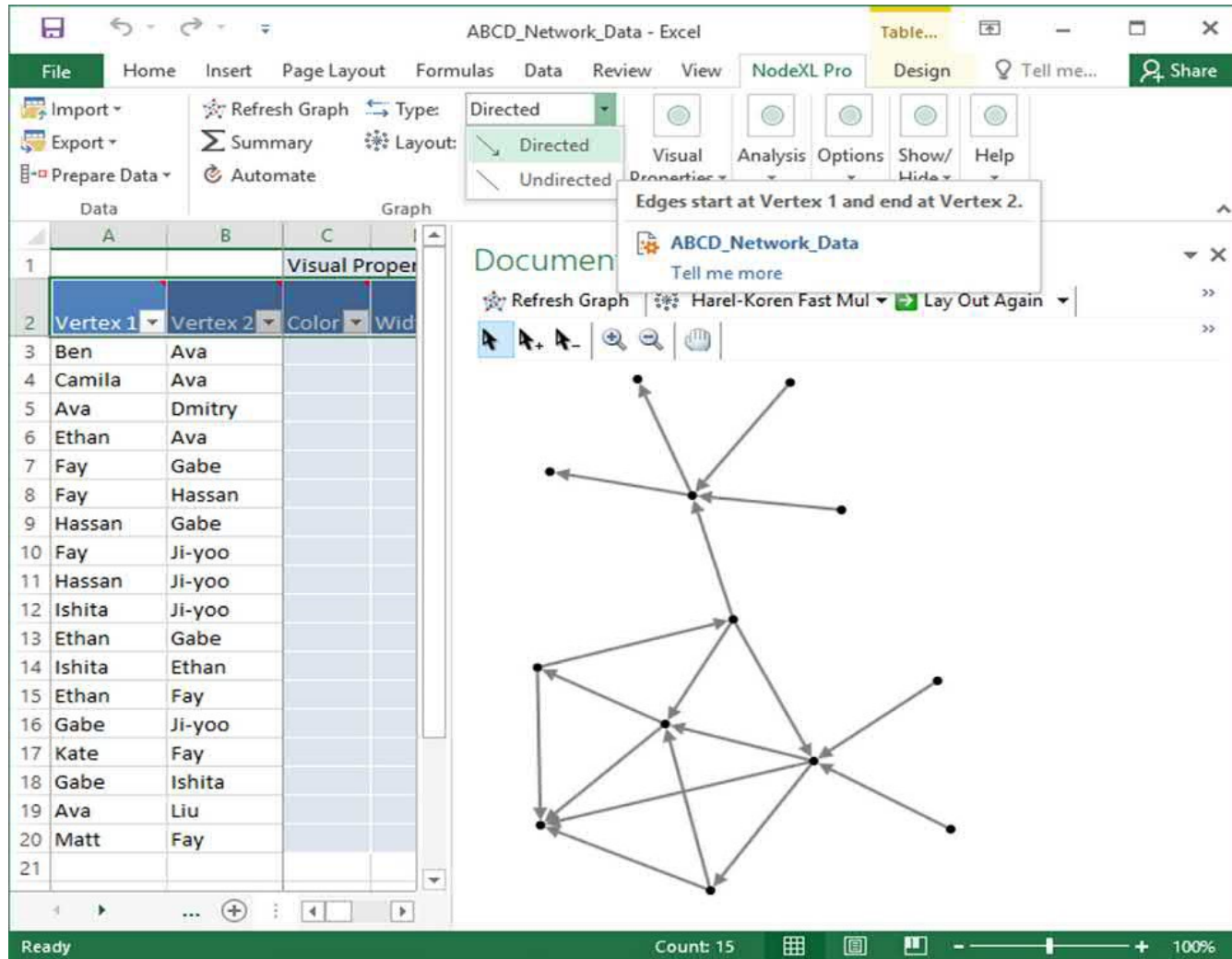
Working with NodeXL files

- **Creating a trusted location for NodeXL files from the Internet**

- Some Excel files can include malicious code.

- To prevent such files from causing trouble, recent versions of Windows require that you specifically give permission to open Excel (and thus NodeXL) files downloaded from the Internet.

- Permission can be given to a single file or to a folder, which can include as many files as desired.

- Instructions for creating a trusted location for various Window's versions can be found on Microsoft's support pages.

Working with NodeXL files

- **Creating a trusted location for NodeXL files from the Internet**

- For a single file you can right-click on the file, go to Properties, and choose Unblock if you have proper permissions as an administrator.

- To create a trusted folder try searching for "create, remove, or change a trusted location for your files" along with the name of your operating system.

- You are encouraged to create a NodeXL folder and make it a trusted location where you store and access all your NodeXL files

Labeling and Visual Attributes

- **Labeling**

- Labeling vertices and edges is essential to creating readable graphs.

- **Viewing attribute data in the ABCD network file**

- The data file includes additional attribute data describing the edges and vertices, which will be mapped to labels and visual attributes.

- If you'd rather enter the data manually, the data values are available.

Labeling and Visual Attributes

Labeling and Visual Attributes

- **Labeling vertices**

- Navigate to the Vertices worksheet of the ABCD network file.

- Scroll over to the Labels columns. Right-click on the first cell in the Label column, choose Format Cells…, and change it from Text to General.

- Then enter = [Vertex] into the cell and press enter so that it will reference the names that are found in the Vertex column (i.e., Column A) on the Vertices worksheet.

Labeling and Visual Attributes

Labeling and Visual Attributes
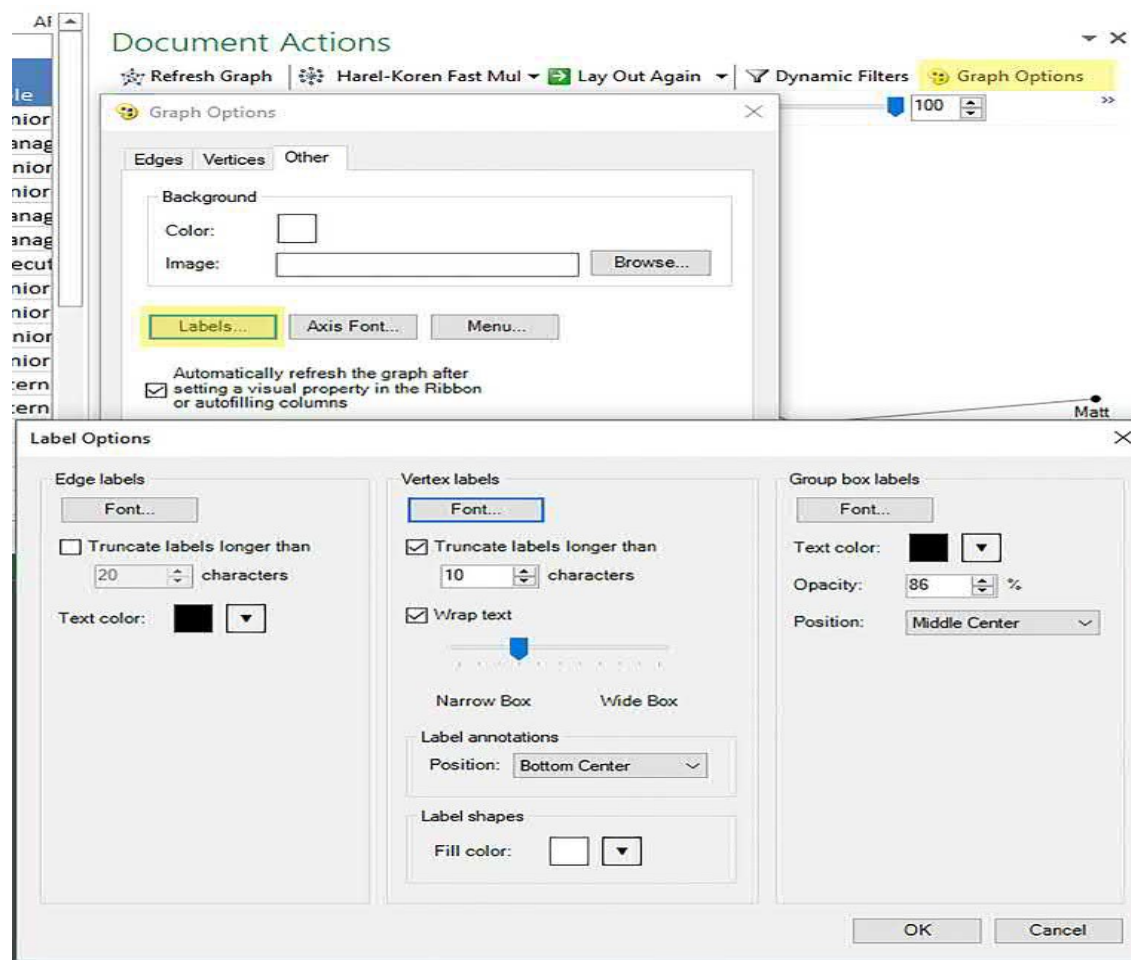
Labeling and Visual Attributes

- **Formatting, positioning, and truncating labels using label options**

- Several additional visual attributes of labels can be modified through the Label Options dialog.

- Click on the Graph Options button on the NodeXL Graph Pane to open the Graph Options dialog.

- Navigate to the Other tab, and click on the Labels… button also

- This will open the Label Options dialog shown at the bottom of where you can automatically truncate labels, change the font and textual properties of labels, set the default position of labels in comparison to the vertices, and more.

- These can be adjusted for edge, vertex, or group box labels.

Labeling and Visual Attributes

- **Label vertex shape**

- To increase readability, it is often useful to turn the vertex into a label rather than the default disk (i.e., filled in circle).

- To do this, navigate to the Shape column on the Vertices worksheet.

- Place the cursor inside of cell C3 and a drop-down menu option will appear next to the cell on the right.

Labeling and Visual Attributes

Labeling and Visual Attributes

- **Labeling edges**

- Edges can also be labeled, though this is less common than labeling vertices, because edge labels are difficult to read on most networks.

- Typically, other visual properties, such as width or color, can represent the value or type of an edge more effectively than edge labels.

- When data is qualitative or unique, and the network size is small, edge labels can be useful.

- Adding label text to the Label column on the Edges worksheet is similar to adding it to the Vertices worksheet.

- You can also customize the color and size of the edge label by entering data into the Label Text Color and Label Font Size columns on the Edges worksheet.

Labeling and Visual Attributes

- **Visual properties**

- NodeXL is a sophisticated and flexible network visualization tool, allowing you to map many types of data to a variety of visual properties of a network graph.

- For example, the color of a vertex may be based on demographic data such as gender or age.

Labeling and Visual Attributes

Labeling and Visual Attributes

- **Vertex color**

- To make the ABCD network graph more visually meaningful, change the color of the male students to blue and the female students to a custom color.

- To set Ben's color, type Blue into the Color column on Ben's row.

- You can type in any color from the 140 Cascading Style Sheet color names

Labeling and Visual Attributes

- **Vertex shape**

- When we set the Shape of each Vertex to Label.

- A variety of additional vertex shapes are available: solid shapes (Disk, Solid Square, Solid Diamond, and Solid Triangle), outline shapes (Circle, Square, Diamond, Triangle), and others (Sphere, Label, and Image).

- The Image shape only works if the Image File field is populated with a valid path name to a file on your computer (e.g., C:\MyImages\Image.jpg) or a URL (e.g., http://www.somesite.com/Image.jpg).

Labeling and Visual Attributes

Opacity	Solid Square	Solid Triangle	Circle
Color	Sphere	Solid Diamond	Drop Shadow
Size	Square	Glow	Selected
Image	Diamond	Label	Triangle

Labeling and Visual Attributes

Labeling and Visual Attributes

- **Vertex size**

- Similar approaches can be used to fill in the data for the Vertex Size column.

- When working with numerical data, such as the data in the Years_at_ABCD column, it is often useful to use the Autofill Columns feature of NodeXL to map the raw data onto the visual properties (e.g., Size).

- Open the Autofill Columns dialog, choose Years_at_ABCD from the drop-down menu next to Vertex Size, and then open the Vertex Size Options dialog

Labeling and Visual Attributes
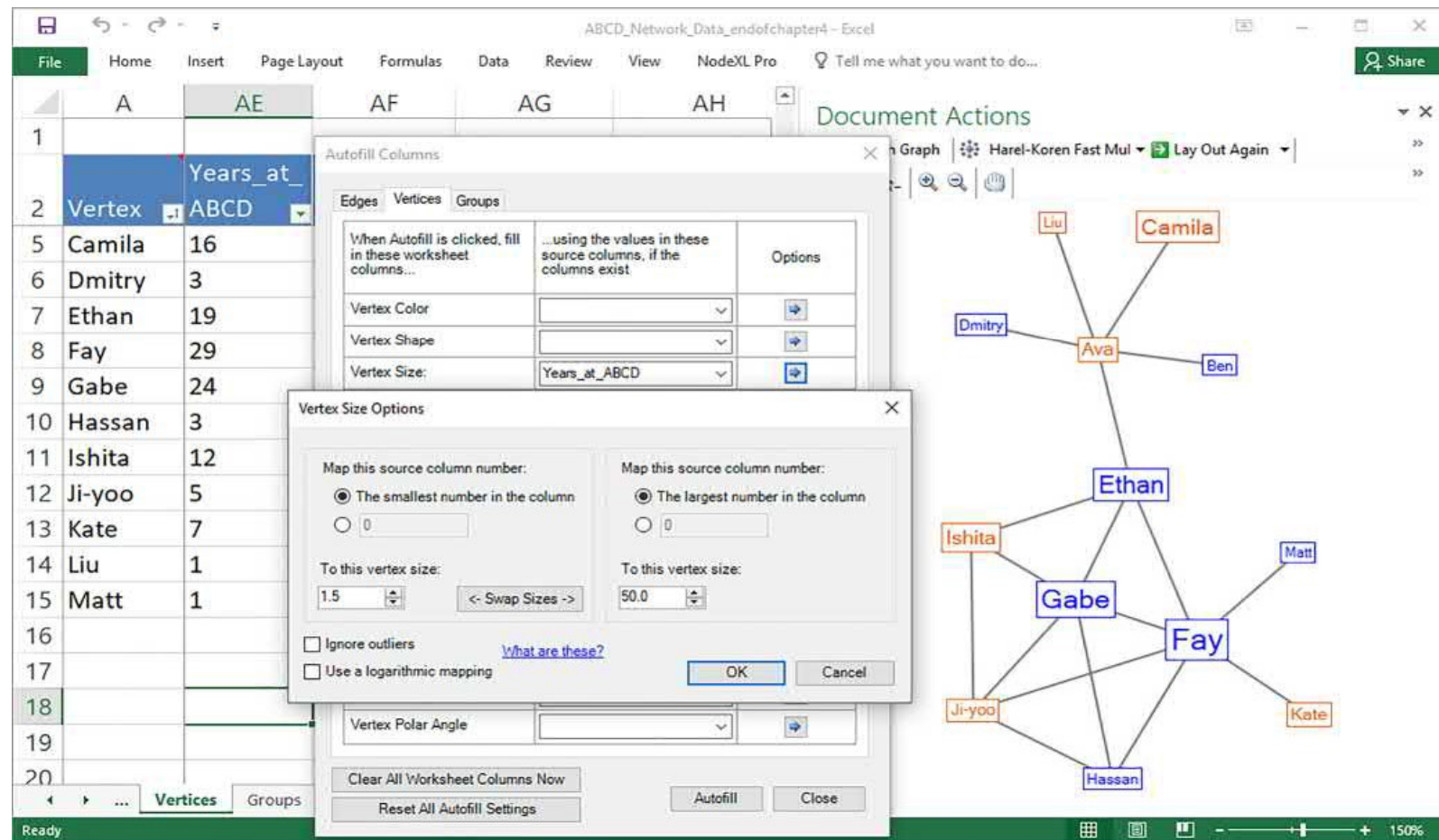
Labeling and Visual Attributes

- **Vertex opacity**

- Vertex Opacity determines the level of transparency (i.e., how see-through) for each vertex.

- Values can be between 0 (fully transparent) and 100 (fully opaque).

- The default value is 100.

- The Autofill Columns options allow you to determine the minimum and maximum value, similar to the Vertex Size Options dialog

- **Vertex visibility**

- When working with large networks, it is often useful to filter out some vertices, so they do not show up in the network.

- The Visibility column allows you to do so without deleting the information from the Excel spreadsheet.

Labeling and Visual Attributes

- **Vertex visibility**

- There are four options available.

- Show if in an Edge will display the vertex on the graph if the vertex is connected to another vertex by at least one edge.

- Otherwise, the vertex row will be ignored. This is the default. Skip will ignore the vertex row and any edges connected to it.
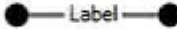
Labeling and Visual Attributes

- **Edge visual properties**

- The Visual Properties columns on the Edges worksheet are slightly different, but work in a similar manner to the Visual Property columns on the Vertices worksheet.

- Below figure presents the many edge visual properties available in NodeXL.

- Color and Opacity work the same way as the corresponding vertex

- attributes.

- Style changes the type of line (Solid, Dash, Dot, Dash Dot, and Dash Dot Dot) and is comparable to the Shape column for vertices. It is best used when working with categorical data.

Labeling and Visual Attributes

- **Edge visual properties**

- Visually, different styles are difficult to differentiate in large networks, so coupling style with distinct colors is often useful.

- Width determines how wide the edge is and is most comparable to the Size vertex property.

- The Visibility column affects the visibility of edges and can be set to Show (always show, no matter what), Skip (act as if the edge does not even exist in the dataset), or Hide (do not display on the graph, but otherwise treat it as if it is present).

Labeling and Visual Attributes

Calculating and Visualizing Network Metrics

- Network graph metrics can describe an entire network, subgroups, or specific actors within a single network.

- Aggregate graph metrics such as network density can be used to systematically compare communities, helping analysts decide which communities are highly connected and which are sparse.

- Tracking aggregate graph metrics over time can determine the effectiveness of interventions on the network as a whole.

- For example, you would expect the total number of edges to grow, increasing the density of the graph, after a photo sharing activity designed to introduce people to those they don't know.

Calculating and Visualizing Network Metrics

- **ABCD network example**

- To better understand the meaning of each graph metric, start by opening the ABCD network visualized in the last chapter (or download it from https://www.smrfoundation.org/nodexl/teaching-with-nodexl/teaching-resources/).

- This network was designed specifically to illustrate the differences between several key metrics.

Calculating and Visualizing Network Metrics

- **Computing graph metrics**

- To calculate graph metrics, first click on the Graph Metrics button on the Analysis section of the NodeXL ribbon.

- This opens the Graph Metrics dialog.

- Select the metrics you want to calculate by checking in the boxes next to them. Details about the metric that is selected (e.g., Vertex clustering coefficient) are shown in the box below.

- Some metrics allow you to customize various options by clicking on the Options… button on the right-hand side. Check the boxes next to the metrics shown in below figure and then click Calculate Metrics.

Calculating and Visualizing Network Metrics

- **Computing graph metrics**

- Some of the graph metrics can take a while to calculate when working with large networks, so a status bar is used to show progress.

- NodeXL will create a new Overall Metrics worksheet and take you there to show summary information for the entire network.

- It also populates a set of Graph Metrics columns on the Vertices worksheet that shows vertexspecific metrics, such as centrality metrics.

Calculating and Visualizing Network Metrics
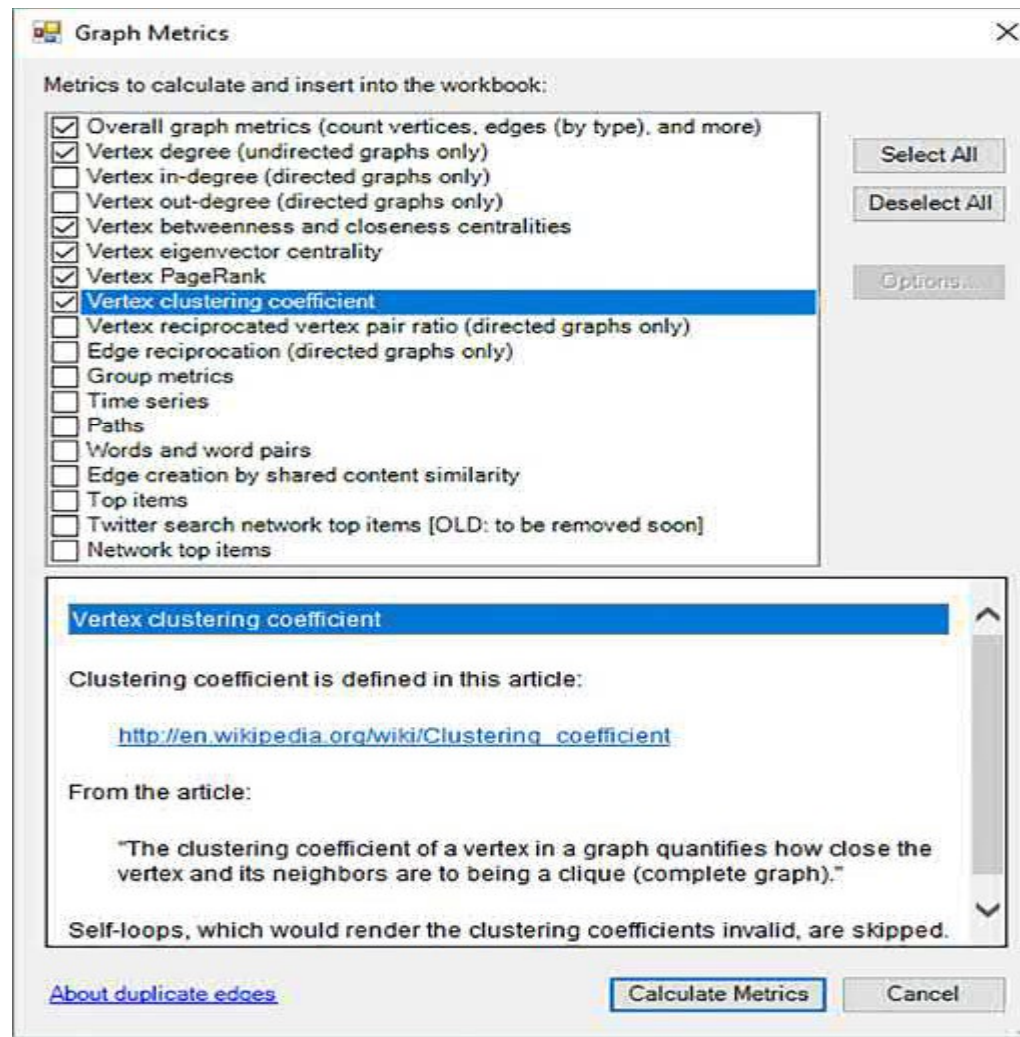
Calculating and Visualizing Network Metrics

- **Overall graph metrics**

- Navigate to the Overall Metrics worksheet, which summarizes some of the key properties of the entire network

- These metrics include the following:

- *Graph type. Undirected or directed.*

- *Vertices. The number of total vertices (i.e., rows on the Vertices worksheet).*

- *Unique edges. The number of unique edges found in the Edges worksheet.*

- *Edges with duplicates. The number of repeated vertex pairs on the Edges worksheet.*

Calculating and Visualizing Network Metrics

- **Overall graph metrics**

- *Total edges. The number of total edges (i.e., rows on the Edges worksheet),*

- *Self-loops. The number of edges that connect a vertex with itself.*

- *Reciprocated vertex pair ratio. This is only applicable for directed networks.*

- *Reciprocated edge ratio. This is also only applicable for directed networks.*

- *Connected components. The number of connected components (i.e.,* clusters of vertices that are connected to each other but separate from other vertices in the graph).

Calculating and Visualizing Network Metrics

- **Overall graph metrics**

- *Single-vertex connected components. The number of* isolated vertices that are not connected to any other vertices in the graph.

- *Maximum vertices in a connected component. The* number of vertices in the connected component with the most vertices.

- *Maximum edges in a connected component. The number* of edges in the connected component with the most edges.

- *Maximum geodesic distance (diameter). The geodesic* distance is the length of the shortest path between two people

- *Average geodesic distance. This is the average* length of the geodesic distances between all pairs of vertices.

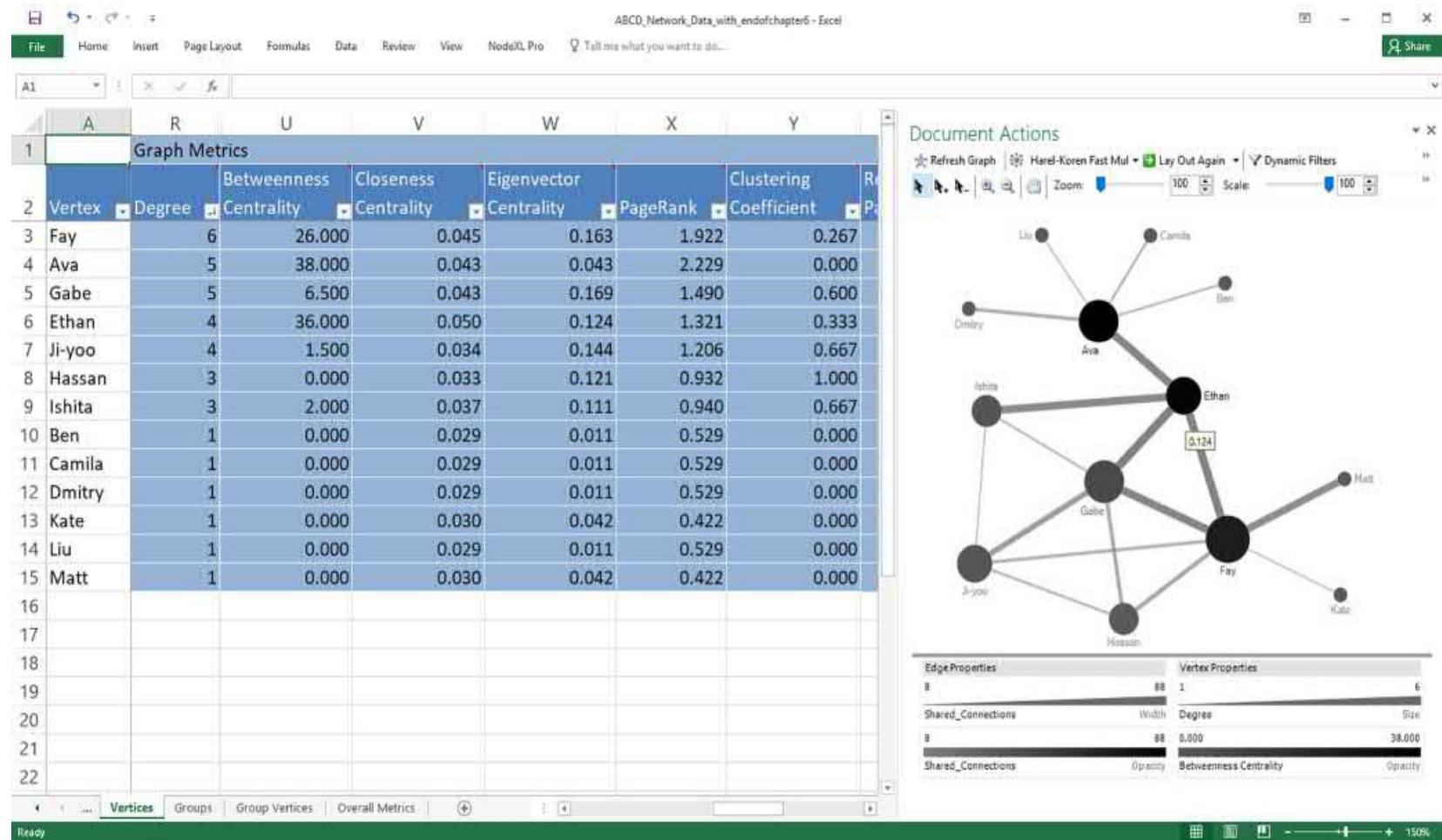# Calculating and Visualizing Network Metrics

- **Overall graph metrics**

- *Graph density. The graph density is a number* between 0 and 1 that indicates the percentage of possible edges that are realized.

- *Modularity. The modularity metric is only* calculated when working with subgroups

- *NodeXL version. Indicates the version of NodeXL in* use when metrics were calculated.

Calculating and Visualizing Network Metrics

	A	B
	Graph Metric	Value
1		
2	Graph Type	Undirected
3		
4	Vertices	13
5		
6	Unique Edges	18
7	Edges With Duplicates	0
8	Total Edges	18
9		
10	Self-Loops	0
11		
12	Reciprocated Vertex Pair Ratio	Not Applicable
13	Reciprocated Edge Ratio	Not Applicable
14		
15	Connected Components	1
16	Single-Vertex Connected Components	0
17	Maximum Vertices in a Connected Component	13
18	Maximum Edges in a Connected Component	18
19		
20	Maximum Geodesic Distance (Diameter)	4
21	Average Geodesic Distance	2.224852
22		
23	Graph Density	0.230769231
24	Modularity	Not Applicable
25		
26	NodeXL Version	1.0.1.407

# Calculating and Visualizing Network Metrics

Calculating and Visualizing Network Metrics

- **Vertex-specific metrics**
- The different vertex-specific metrics, also called centrality metrics, help identify who is "important" or "central" to a network.
  - *Degree*
  - *Betweenness centrality*
  - *Closeness centrality*
  - *Eigenvector centrality*
  - *PageRank*
  - *Clustering coefficient*

Calculating and Visualizing Network Metrics

- **Mapping graph metrics to X and Y coordinates**

- In most layouts, the exact location of the vertices is not meaningful; only their position relative to one another has meaning.

- However, you may want to map network graph metrics, or other attribute data, to X and Y coordinates to visualize how two metrics interact with one another.

- Other metrics can be used to adjust visual properties, making it possible to display additional dimensions.
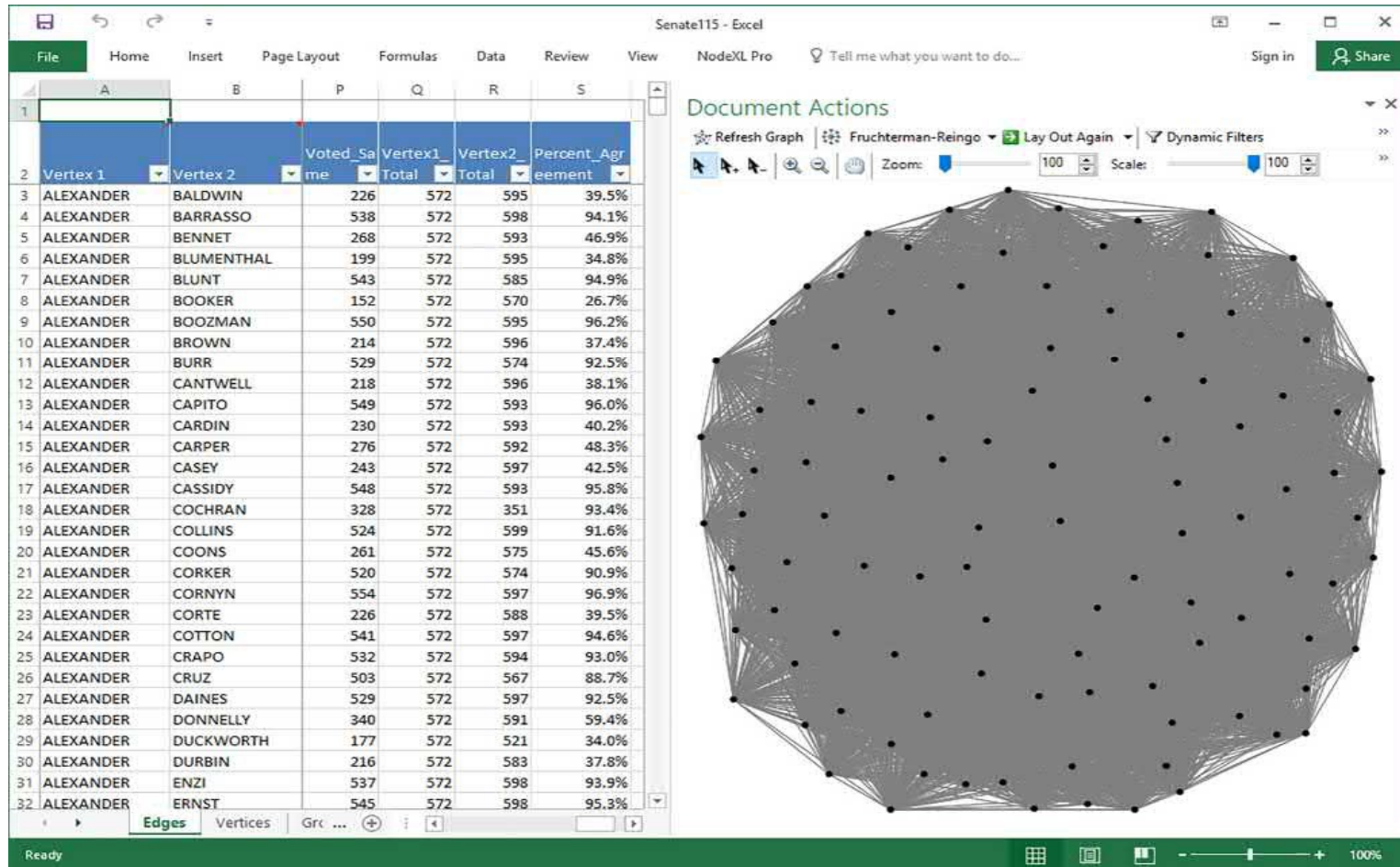
Grouping and Filtering
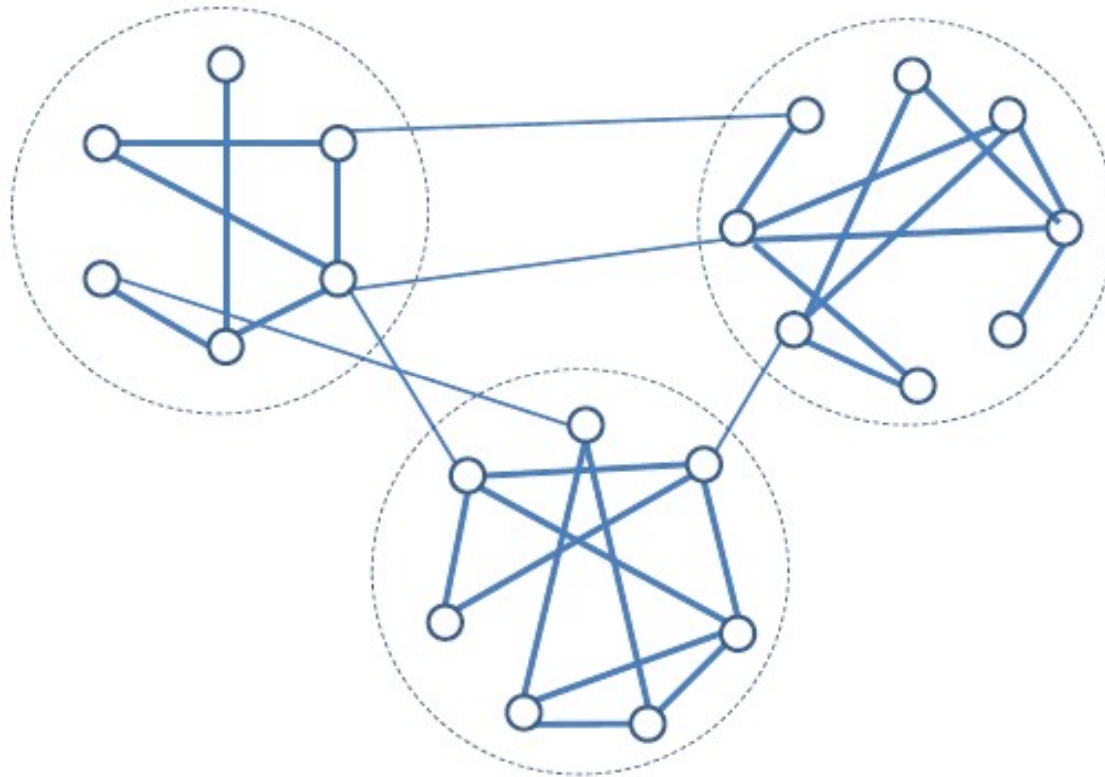
FIGURE 7.1 A network of three densely connected clusters (i.e., groups), each shown inside a dashed circle. Ties between clusters are rare and less dense.

Grouping and Filtering

- **Filtering edges to identify groups within a network**

- When working with weighted networks such as the 115th Senate network, it is often necessary to filter out some of the edges to identify subgroups.

- Because every senator voted the same as every other senator at least once, choosing Show Graph results in an uninformative dark mass of connections
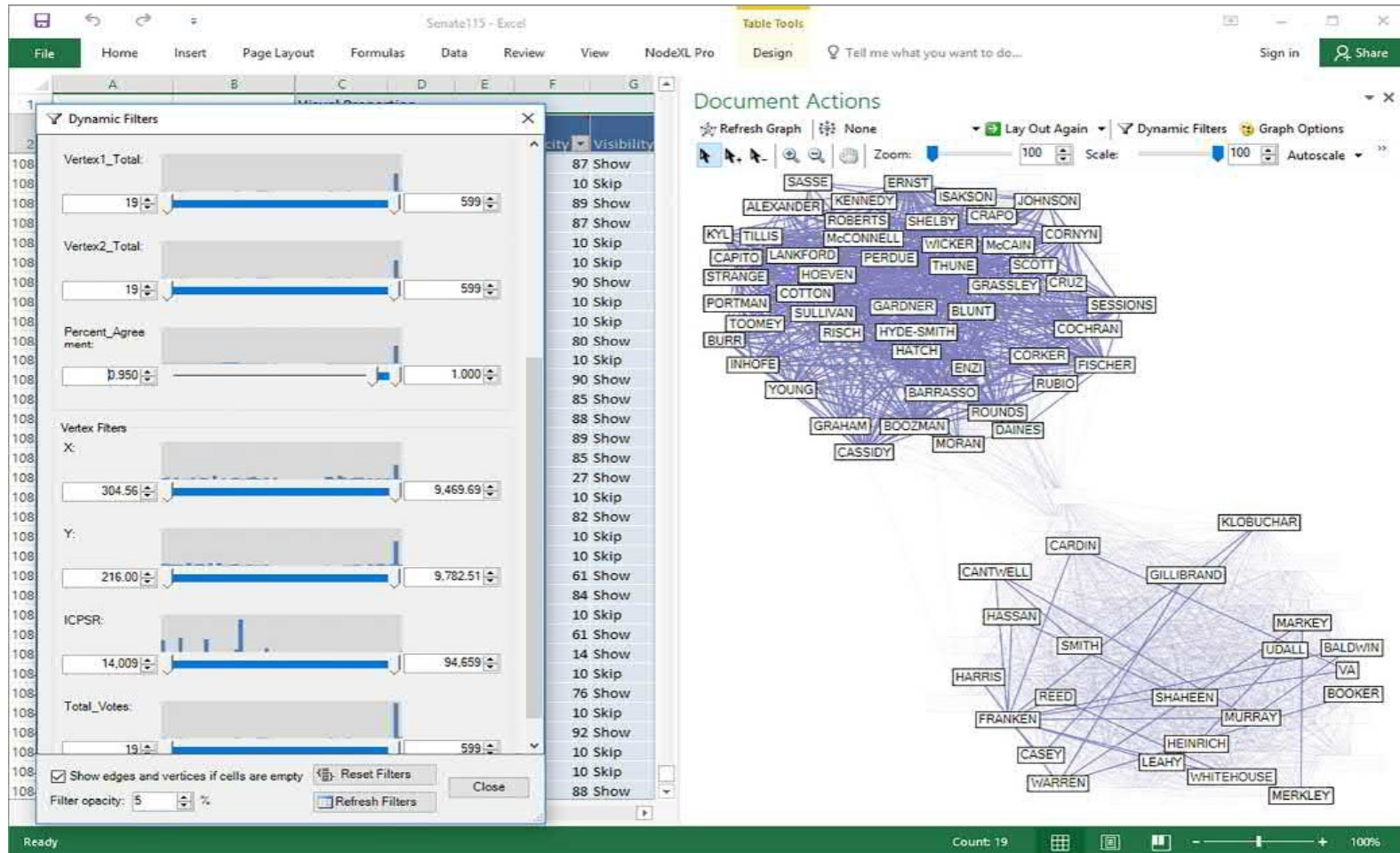
Grouping and Filtering

Grouping and Filtering

- **Using dynamic filters**

- NodeXL allows you to dynamically filter out edges or vertices based on any data fields on the Edges and Vertices worksheets.

- This can be an excellent way of exploring a dataset, without making any permanent decisions about how to display it.

- Click on the Dynamic Filters button above the graph and you will be presented with a new window that lets you set the minimum and maximum values for each variable.

Grouping and Filtering

Grouping and Filtering

- **Creating groups based on vertex attribute**

- Sometimes you will have data that describes attributes of the people in your network.

- For example, our Senate dataset includes information on the political party of each senator (see Party column on the Vertices worksheet).

- Values include D (Democrat), R (Republican), and I (Independent). To visually display this information, click on the Groups drop-down menu in the NodeXL ribbon and choose Group by Vertex Attribute.

- Choose Party from the first drop-down menu, since this is the attribute you want to group based upon.

Grouping and Filtering

- **Creating groups based on vertex attribute**

- Then choose Categories from the second drop-down menu, since the data is categorical in nature.

- Notice, though, that groups can be created based on numerical data or date and time data. When you click OK, NodeXL will create two new worksheets.

Grouping and Filtering

Grouping and Filtering