COMM 645 HANDOUT – GEPHI BASICS

Gephi Resources:

- Gephi: The Open Graph Viz Platform. Download from <u>www.gephi.org</u>
- Gephi Quick Start Tutorial: www.gephi.org/users/quick-start
- Gephi Visualization Tutorial: <u>www.gephi.org/users/tutorial-visualization</u>
- Gephi Layouts Tutorial: <u>www.gephi.org/users/tutorial-layouts</u>
- Gephi Wiki: <u>wiki.gephi.org</u>
- Gephi Community Portal: <u>wiki.gephi.org/index.php/Community_portal</u>

Data Formats

Gephi's own files have an extension .gephi.

The software can read many other file formats, including GEXF, GDF, GML, GraphML, Pajek NET files, and UCINET DL files. You can open any of those by simply going to *File -> Open*.

Gephi can also read Excel and CSV files. To import data from CSV, you will usually need to prepare two files – one containing nodes and their attributes, and another containing an edge list and edge attributes (see the example CSV files used in the class demonstration for Lab 9).

The CSV file containing nodes needs to include a column named "ID" containing - as you might have guessed – unique node IDs. The edge list CSV should include columns titled "Source" and "Target", containing node IDs of the start and end node for each edge. In addition, you can include a column called "Type" indicating the type of each edge ("Undirected" or "Directed").

To import the files, go to the *Data Laboratory* tab -> *Import Spreadsheet* button (3 on Fig.3). Select your file, make sure to select the appropriate table type from the "*As Table:*" drop-down menu (i.e. *Nodes Table* or *Edges Table*). Import both files, preferably starting with the node table.

Click "Next" and make sure that each column of your data is assigned an appropriate format (e.g. you can import text fields as "String", categorical numerical data as "Integer", continuous attributes as "Float", time interval variables for dynamic networks as "TimeInterval", etc.). Click "Finish".

Major sections within Gephi:

- **Overview Tab:** Calculate network measures & set up the network visualization.
- Data Laboratory Tab: Examine and edit the network data (node & edge tables).
- **Preview Tab:** Configure the rendering settings (colors, sizes, line width, opacity, etc.) and preview the visualization.





Fig.1 \rightarrow **1**, **Partition Tab:** Change node/edge color based on a categorical attribute (select the *Nodes* or *Edges* tab, select the attribute from the drop-down *Type* menu).

Fig.1 \rightarrow **2**, **Ranking Tab:** Change the color or size of a node/edge/label based on a continuous attribute (You can select which visual property to change using the buttons in the top right corner of the tab).

Fig.1 \rightarrow 3, Layout Tab: Select and customize one of the available layout algorithms.

Fig.1 \rightarrow **4:** Interactive selection of nodes/edges; change size/color manually; add nodes/edges, etc.

Fig.1 \rightarrow **5**: Reset the node size, color, label, or label size.

Fig.1 \rightarrow **6:** Change color, size and other characteristics applying to **all** nodes, edges, and labels.

Fig.1 \rightarrow **7a, Filters Tab:** Apply filters to select specific nodes and/or edges from your network. The filters are applied through dragging and dropping onto Fig.3 \rightarrow 7b: Queries.

The filters based on *attributes* include "Equal" (select elements with particular attribute values), "Partition"(select different levels of categorical attributes), "Range" (select nodes/edges with attribute values in a particular range – for instance out-degree between 1 and 4), Inter-edges (select edges with particular attributes – for one-mode networks), Intra-edges (select edges with particular attributes, going across the modes of multimode networks), and so on. Filters based on *edges* allow you to select ties with different properties – e.g. a particular range of weights. *Topology* filters allow for selection based on network structure – components, K-cores, degree range, etc. *Operators* allow you to combine other filters in different ways (find their intersection, union, complement, etc.).

Fig.1 \rightarrow 8, Statistics Tab: To calculate a network/node/edge statistic, click the "Run" button next to it. Once calculated, many measures will be available in the Data Laboratory – and can be used in the visualization (e.g. computing eigenvector centrality will allow you to resize nodes based on that attribute). Note that to get betweenness and closeness centrality, you need to calculate the average path length (hit "Run" next to Avg. Path Length).

Gephi 0.8.1 Project 1	DI			- 8
Overvier Dols Window	Plugins Help pratory Preview	(3)		XX
Data Table				
Nodes Edges 😔 Configuration 😂 Add n	ode 🕣 Add edge 🏙 Search/Repl	ace 🖭 Import Spreadsheet 🕮 Export table 🎇 More act	ions Y Filter:	Nodes
Nodes	Id	Label	Type TypeLabel	Size
Wall Street Journal	3	Wall Street Journal	1 Newspaper	
CNN	7	CNN	2 TV	
LA Times	5	LA Times	1 Newspaper	
Google News	13	Google News	3 Online	
FOX News	9	FOX News	2 TV	
ABC	10	ABC	2 TV	
Yahoo News	12	Yahoo News	3 Online	
Reuters.com	14	Reuters.com	3 Online	
BBC	11	BBC	2 TV	
USA Today	4	USA Today	1 Newspaper	
NYTimes.com	15	NYTimes.com	3 Online	
AOL.com	17	AOL.com	3 Online	
NY Times	1	NY Times	1 Newspaper	
MSNBC	8	MSNBC	2 TV	
Washington Post	2	Washington Post	1 Newspaper	
New York Post	6	New York Post	1 Newspaper	
WashingtonPost.com	16	WashingtonPost.com	3 Online	
		S	4	
Add column	Merge column × column	Copy data to Fill column Duplicate other column v with a value v column v	값 값 공 sate a boolean column Create column with list of regex matching groups ~ Negate boolean value	s ×]
				Workspace (

Figure 2: The Gephi interface: Data Laboratory

Fig.2 \rightarrow **1**: Data table for nodes and node attributes.

Fig.2 \rightarrow **2**: Data table for edges and edge attributes.

Fig.2 \rightarrow **3**: Import spreadsheets: node and edge data from Excel/CSV/etc.

Fig.2 \rightarrow **4**: Manipulate data columns (you can also change individual values directly by clicking on them and typing their new values).



Figure 3: The Gephi interface: Preview Tab

Fig.3 \rightarrow **1:** Configure rendering settings: size, color, and other properties of nodes, edges, and labels. Those apply to the visualization in general. If you want to change specific properties (e.g. the color of a particular node subset, the width of some edges), you need to go back to the *Overview* tab (Fig.1).

Fig.3 \rightarrow **2:** If you are visualizing a very large network, rendering it may be time- and resource-intensive. While you're tweaking the visualization properties, you may want to preview a smaller subgraph rather than the full network. This option is controlled by the *Preview Ratio* setting.

Fig.3 \rightarrow **3:** After you change the rendering settings you need to refresh the network preview.

Fig.3 \rightarrow **4**: Change the preview background color.

Save and Export

Once you are satisfied with your network visualization, you can save your project in one of multiple supported formats. To save as a graph file, go to *File -> Export -> Graph File* and select your preferred format. To save as an image or PDF file, go to *File -> Export -> SVG/PDF/PNG File*. If installed, several optional Gephi plugins will also allow you to export to an interactive web project that can be easily embedded into your website.