NETWORK VISUALIZAITON WITH ℘Gephi

TABLE OF CONTENTS

Gephi

GEPHI RESOURCES

Note: at present, many resources on the Gephi website are not updated for v. 0.9 and higher.

- Gephi, The graph visualization platform: www.gephi.org
- Gephi Quick Start Tutorial: www.gephi.org/users/quick-start
- Gephi Visualization Tutorial: www.gephi.org/users/tutorial-visualization
- Gephi Layouts Tutorial: www.gephi.org/users/tutorial-layouts
- Gephi Development on GitHub: github.com/gephi/gephi

- Download the example data files for this manual here.

INITIAL SETUP

You can download the latest version of Gephi (0.9.1) from www.gephi.org.

For more detailed setup information and troubleshooting, go to gephi.org/users/install.

As of April 2016, version 0.9 is fairly new. It still has bugs that you may or may not encounter depending on your machine and operating system. If you would like to revert to the previous and more stable version of Gephi, you can still download and install it from GitHub: www.github.com/gephi/gephi/releases/tag/v0.8.2

Gephi has a number of community-contributed plugins available from the Tools → Plugins menu, as well as online at marketplace.gephi.org/plugins. At present, many of the plugins do not yet work with Gephi 0.9.1.

One final note about your Gephi settings. The etc subfolder of the Gephi folder on your machine contains a file named gephi.conf. You can edit this file to change the startup settings of Gephi – for instance the maximum available memory that the software can use. To increase that, modify the value that follows the stirng -Xmx in the configuration file to the desired number of MB.

Note that if you are working with very large graphs and want to increase Gephi's memory to several GB, you need to have a 64-bit OS and a 64-bit version of Java.

DATA FORMATS

Gephi projects have their own file 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 going to File → Open.

When you import GraphML, GDF, or GEXF files, Gephi should be able to import the position, color, and size of nodes when those are present in the data.

In order to open a **matrix** data file stored as CSV (from File → Open), you need to be sure that the file is semicolon-separated. If you have a comma-separated CSV file, simply open it in a text editor and replace all commas with a semicolon.
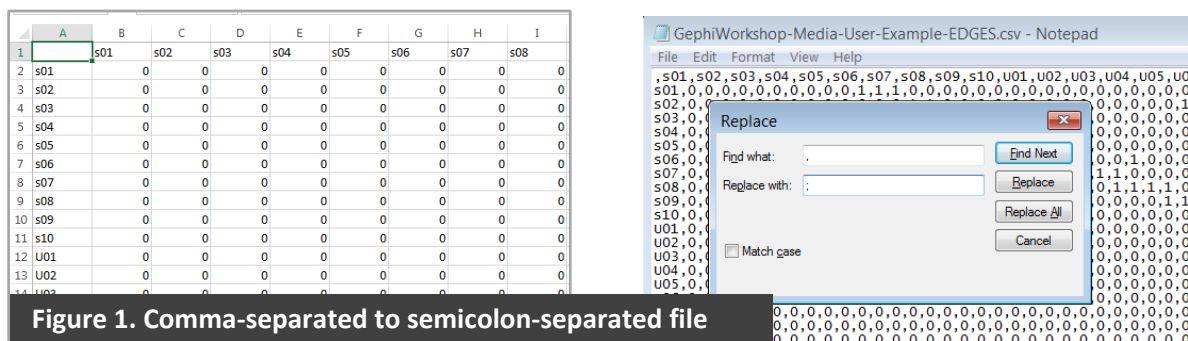

Figure 1. Comma-separated to semicolon-separated file

Gephi can also import Excel and CSV files with data in an **edge list** format. To import data that way, you will usually need to prepare two files – one containing nodes and their attributes, and another containing an edge list and edge attributes (see Fig.2 below).

Source	Target	Weight	Type
n0	n1	70	Directed
n0	n109	73	Directed
n0	n110	53	Directed
n0	n111	47	Directed
n0	n118	23	Directed
n0	n142	29	Directed
n0	n209	22	Directed
n0	n211	133	Directed
n0	n212	25	Directed

ID	Visits	SiteType	Rank
n0	1820	Politics	33
n1	2089	Politics	39
n2	1834	Media	28
n3	3678	Media	42
n4	162	Politics	3
n5	535	Politics	11
n6	4067	Media	43
n7	267	Politics	6
n8	91	Politics	2
n9	25691	Media	70
n10	2640	Media	46
	6105	Media	46
	16199	Media	54

Figure 2. Edge list and node attribute CSV files.

The CSV file containing nodes needs to include a column named **"ID"** containing unique node identifiers – as well as any other node attributes you would like to include. The edge list CSV should include columns titled **"Source"** and **"Target"**, containing the node IDs of the start and end node for each edge, as well as any other edge attributes you would like to include.

Gephi can recognize two more columns if you include them in your data: **"Type"** indicating the type of each edge (*Undirected* or *Directed*), and "**Weight**" containing the edge weight.

To import the files, first create a new project from File → New Project. Then go to the Data Laboratory tab → Import Spreadsheet button. Select your file, and 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.
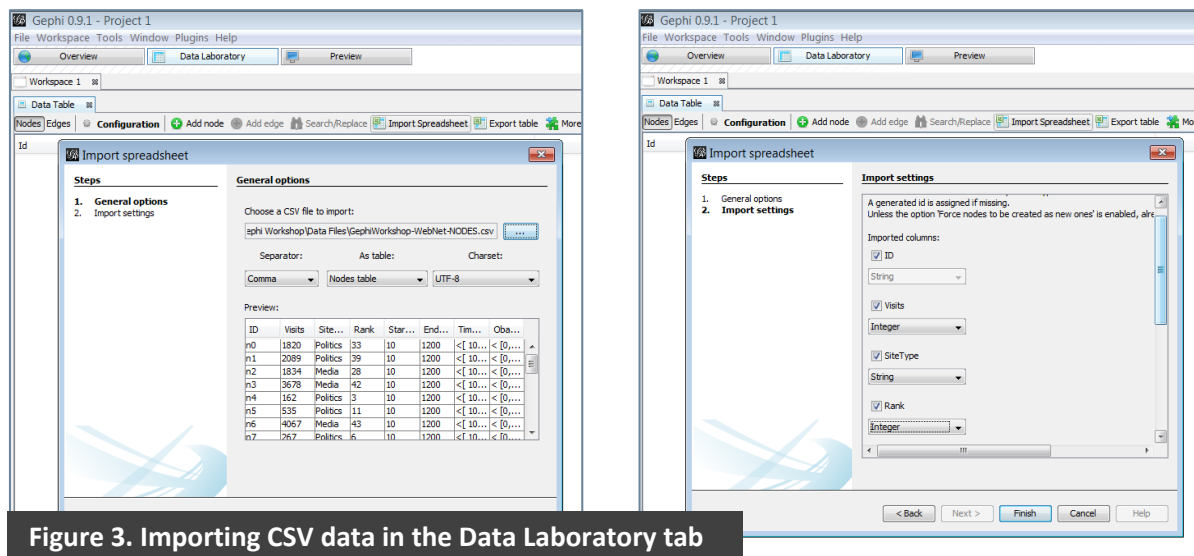

Figure 3. Importing CSV data in the Data Laboratory tab

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

Keep in mind that it is important to select the correct variable type, because Gephi associates those variable types with certain actions. For example, the "Appearance" menus will not let you set the size of a node in your graph based on a *String*-type attribute, so you need to make sure that the attribute used for node size is of type *Float*. Once you are done setting the correct attribute types, click "Finish".
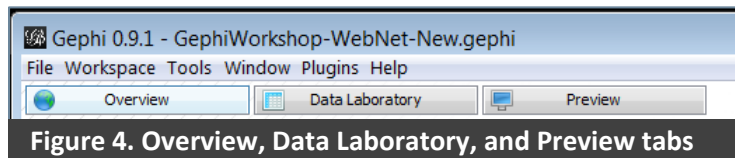
GEPHI INTERFACE: MAJOR SECTIONS

Figure 4. Overview, Data Laboratory, and Preview tabs

- **Overview Tab:** Calculate network and node measures, change visual properties and generally set up the network visualization. This is where you do most of the work: resizing, coloring, and laying out the nodes and edges.

- **Data Laboratory Tab:** Examine and edit the network data (node & edge tables). This is also where you can import CSV and Excel files.

- **Preview Tab:** Configure the rendering settings (colors, sizes, line width, opacity, etc.) and preview the visualization.
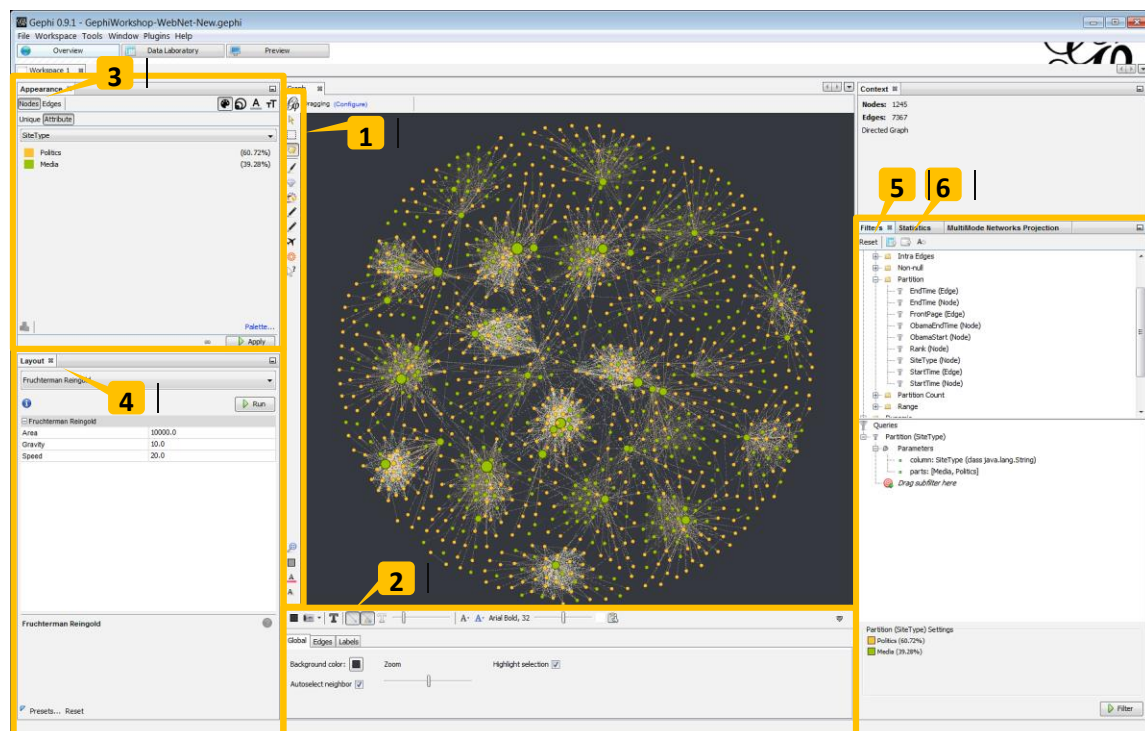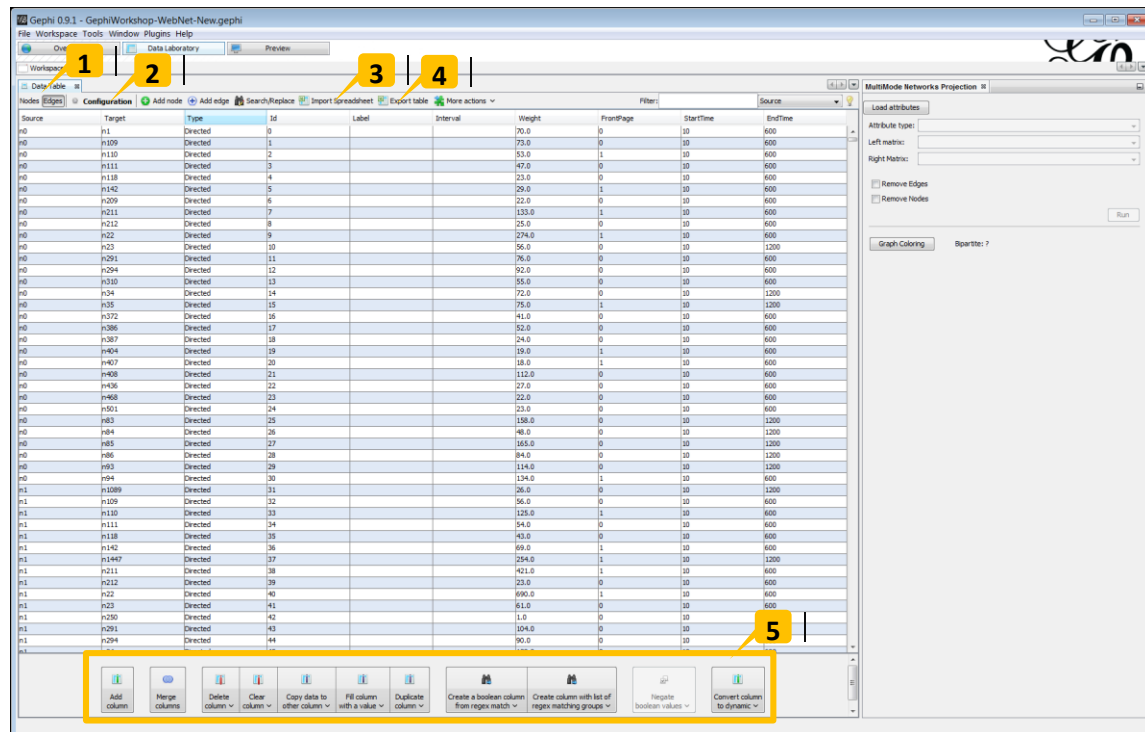

Figure 5. The Overview Tab
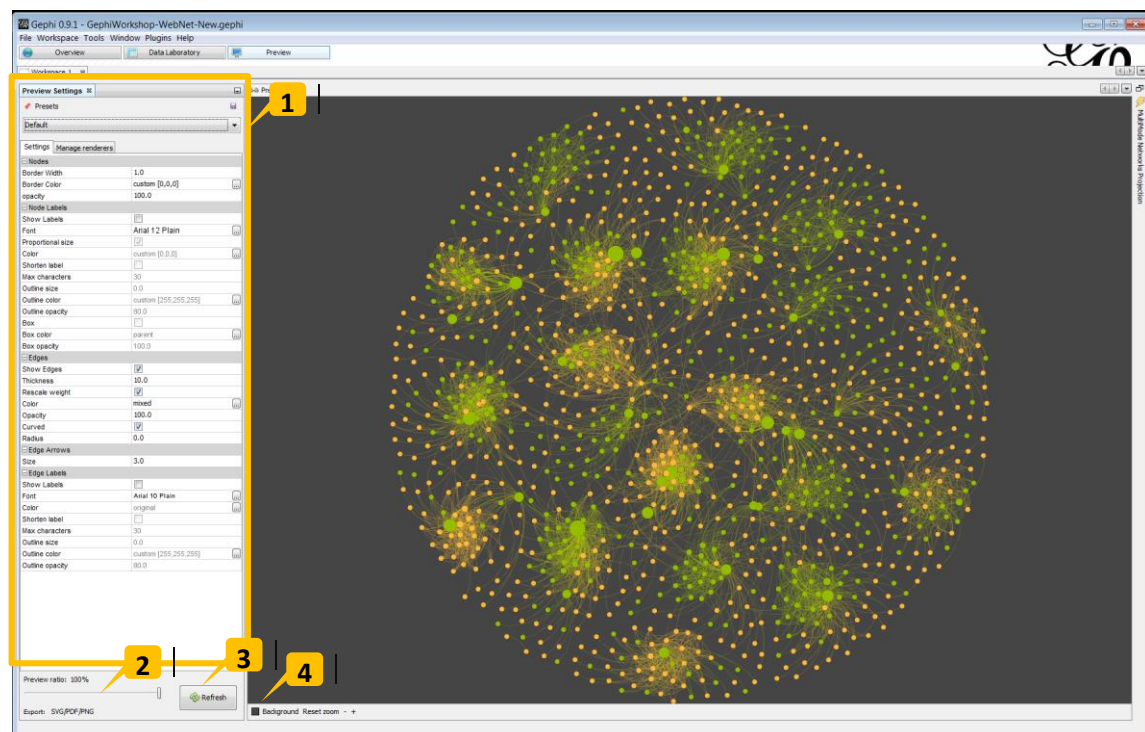
Figure 6. The Data Laboratory Tab

Figure 7. The Preview Tab

Manual actions (Fig.5-1)

There are a number of options to manually interact with a graph – you will find their buttons on the right side of your graph image in the Overview tab.
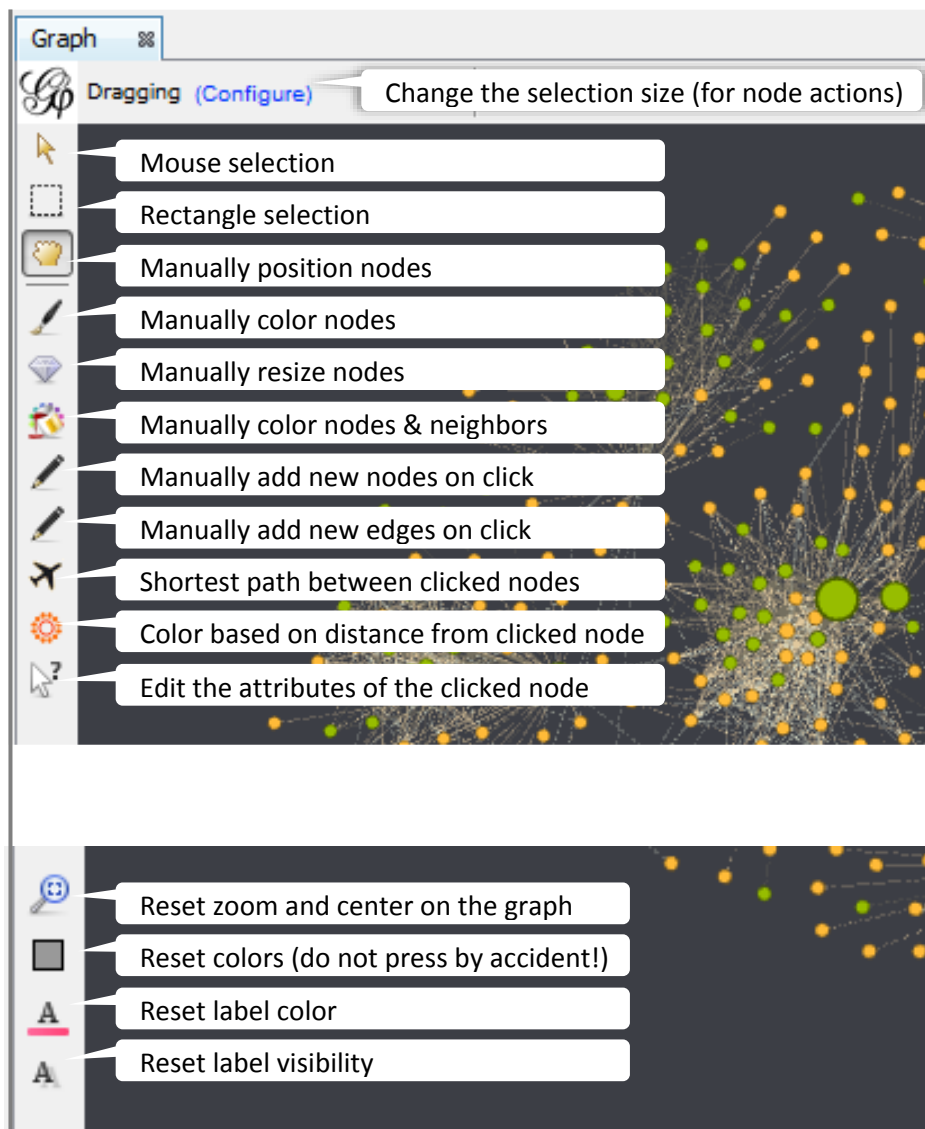

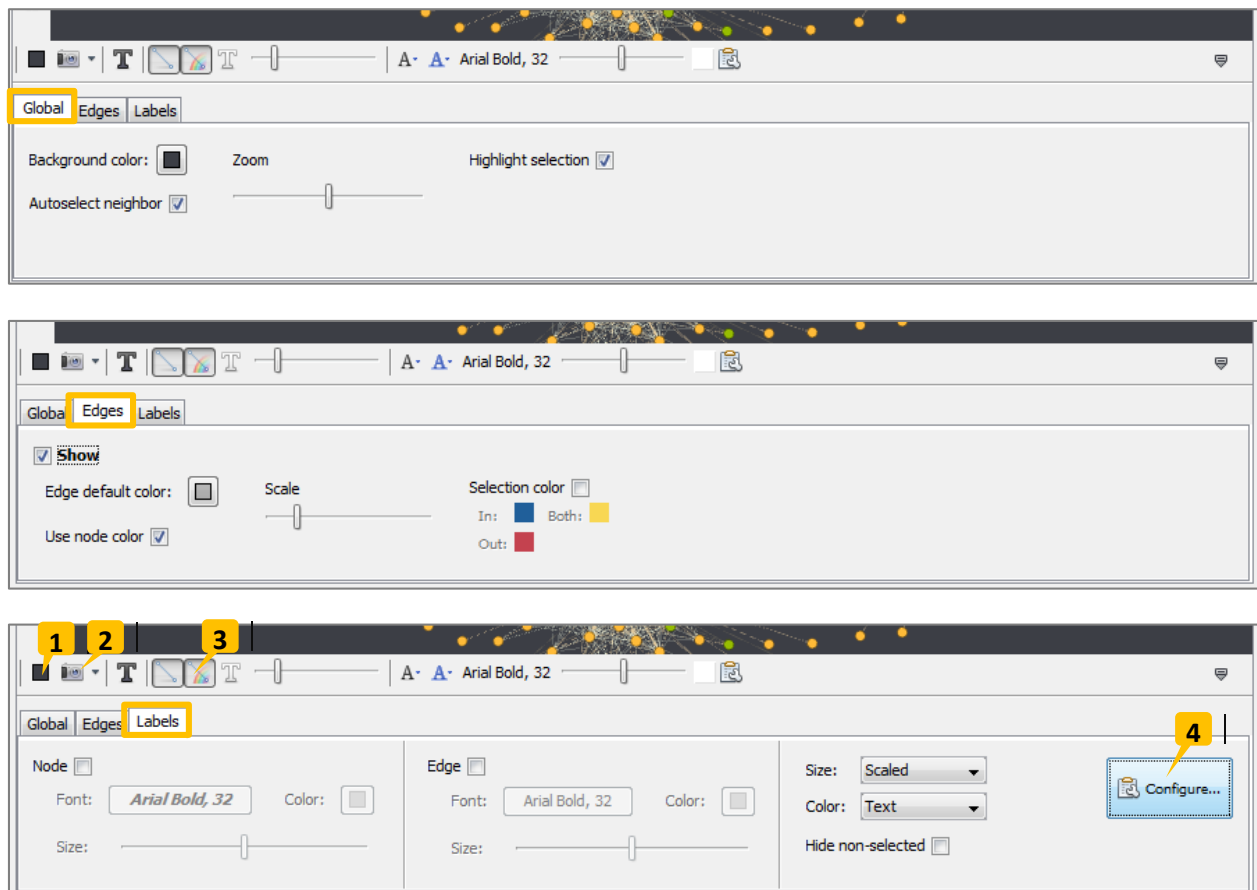**Figure 8. Manual actions**

Graph options (Fig.5-2)

Figure 9. Graph Options

Using the panel below your graph in the Overview tab, you can set global node, edge, and label options. That includes setting default node, edge, and label colors and scaling, as well as quickly showing and hiding all edges and labels in the network overview.

One important setting that should be mentioned here is the selection of a label variable. You can choose which attribute(s) in your data will be used for the label by going to Labels → Configure (Fig.9-4).

This panel is also the place where you can change the background color (Fig.9-1), take a screenshot (Fig.9-2), and change the edge color to the color of its source node. Note that those settings will apply to the Overview tab only, and you will have to set them separately for the rendered image in the Preview tab.
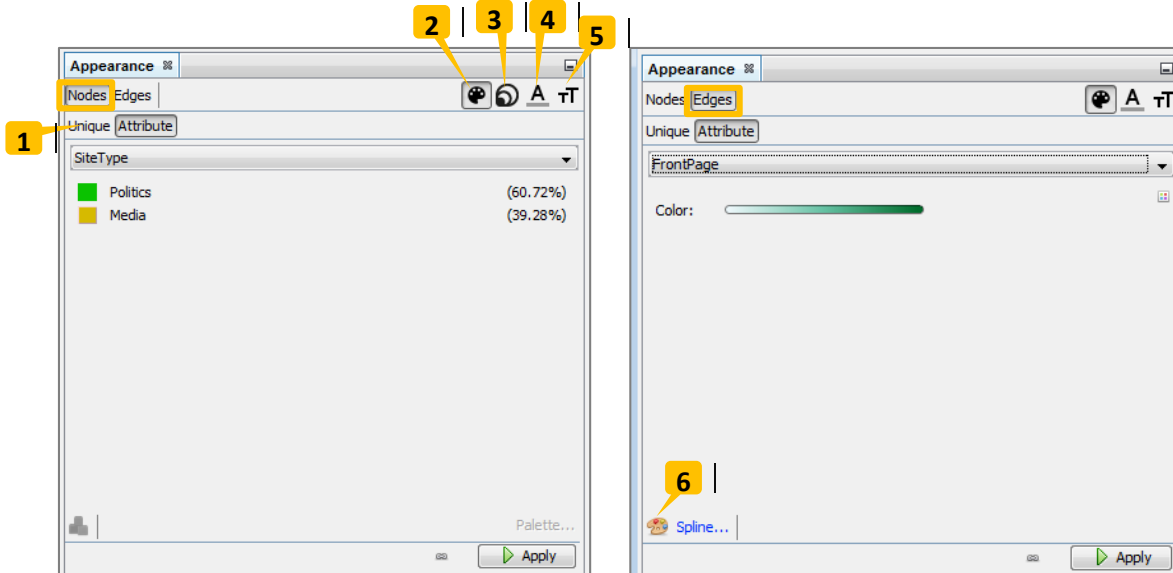
Appearance options (Fig.5-3)

Figure 10. Appearance Options

The Appearance box of the Overview tab allows you to change the visual characteristics of nodes, edges, and labels. The node and edge tabs have two options: Unique and Attribute (Fig.10-1). *Unique* changes all visible elements uniformly to the selected color or size. *Attribute* changes visual characteristics based on a selected attribute from the dataset.

To change parameters based on attribute, select the appropriate element (node or edge), select the "Attribute" tab (Fig.10-1) and choose an attribute from the drop-down menu.

The visual parameters that can be changed for **nodes** from this box include: color (Fig.10-2), size (Fig.10-3), label color (Fig. 10-4), and label size (Fig.10-5). The parameters for edges are the same except for size (edge width can be set based on the edge *Weight* attribute).

Note that this also means you can visualize different types of relationships in by coloring their respective edges in a different color. Gephi 0.9 supports multiplex graphs, and you can have multiple ties between the same two nodes. Unfortunately, as of 0.9.1 Gephi cannot visualize these multiple overlapping links – support for that is expected in future releases.

When you set nodes or edge color based on a numeric attribute, the attribute can be treated as categorical (separate colors for each level) or as continuous (a color gradient, usually with darker or more intense colors representing higher attribute values). To switch between categorical and continuous, press the cubes/palette button at the bottom left corner (Fig.10-6)
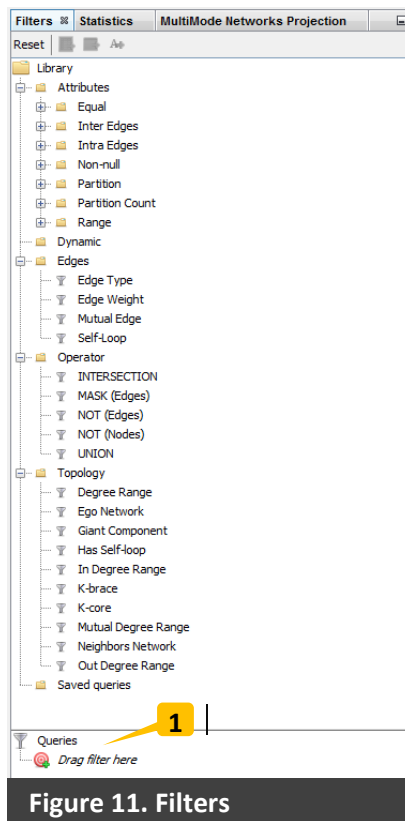
Filter options (Fig.5-5)

Figure 11. Filters

You can apply a variety of filters to select specific nodes and/or edges from your network. The filters are applied by dragging and dropping them onto the Queries window below (Fig.11-1).

Filters based on *attributes* include **Equal** (select elements that have 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 going between nodes that have the selected attribute values), **Intra-edges** (select edges going among nodes with a particular attribute value), and so on.

Filters based on *edges* allow you to select graph edges 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.).
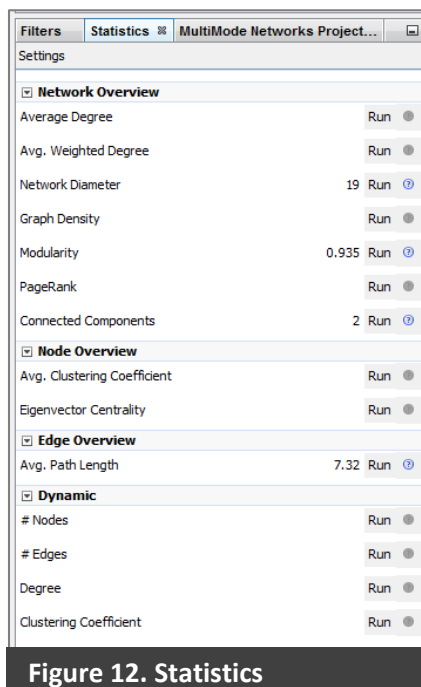
Statistics (Fig.5-6)

Figure 12. Statistics

To calculate a network/node/edge statistic, simply 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). You will see the newly generated attributes in the drop-down Attribute menu of the Appearance box (Fig.5-3).

Note that to get betweenness and closeness centrality, you need to calculate the average path length (hit "Run" next to *Avg. Path Length*).
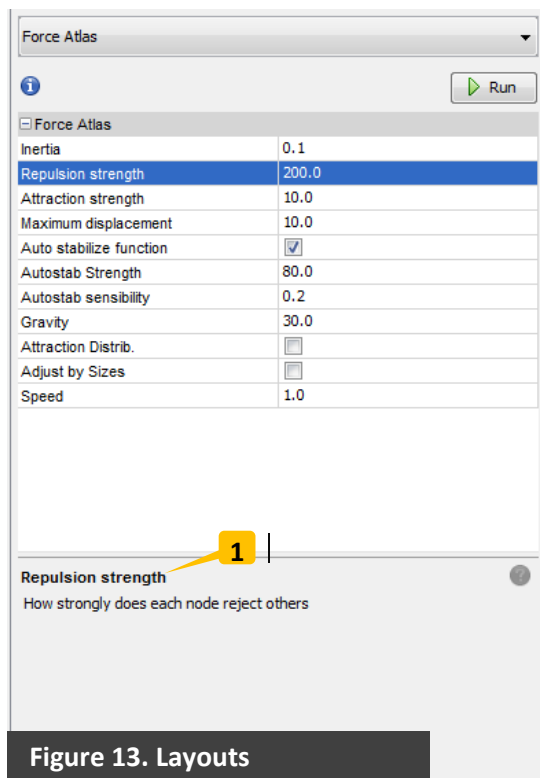
Layouts (Fig.5-4)

Figure 13. Layouts

Gephi offers a variety of layouts, including several relatively fast home-grown force-directed options (*Force Atlas,* and the faster *Force Atlas 2*).

The classic force-directed algorithm of *Fruchterman & Reingold* nicely distributes the nodes in the available area, but is too slow for large graphs.

Yifan Hu provides a multilevel force-directed layout that performs well for larger graphs.

For undirected graphs, *OpenOrd* also provides a nice layout that aims at separating clusters.

The *Geo Layout* allows you to place nodes according to geographic coordinates in the projection of your choice. In order to use it, you need to have node attributes for geographic latitude and longitude in your data. If you do not see the *Geo Layout* in the layout menu, go to Tools → Plugins and install *GeoLayout* from the Available Plugins tab.

Several available algorithms are meant to help further improve your layout – preventing node overlap (*Noverlap*) and label overlap (*Label Adjust*); rotating the network (*Rotate*), and scaling the network (*Expand* and *Contract*) to make it more spread out or condensed.

Each specific layout parameter is nicely explained in Gephi (Fig.13-1) once you click on it or hover over it with the mouse cursor.
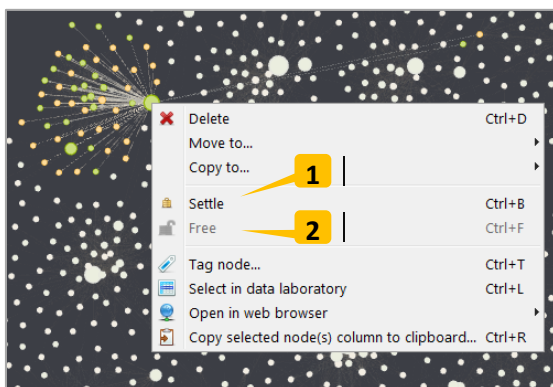
Node options

Figure 14. Node options

Right-clicking on a single node or group of nodes allows you to perform a number of actions including deleting, moving or copying to a new workspace, selecting the corresponding row in the data, etc.

One interesting option is "Settle" (Fig.13-1) – selecting it locks the node to its position, so that applying layout algorithms will not change its position. To release a settled node select the option "Free" (Fig.13-2).
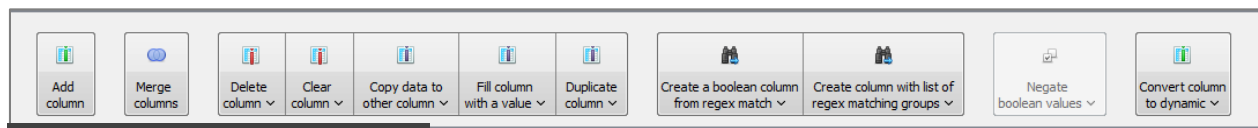
GEPHI INTERFACE: THE DATA LABORATORY TAB

Figure 15. Column manipulation

The data laboratory tab (Fig.6) contains the data tables for node and edge attributes. This is where you can import CSV files (Fig.6-3), export data tables from open network projects (Fig.6-4), modify or delete graph elements.

You can change individual values directly by clicking on them and typing their new values. The Data Laboratory also provides a variety of ways to manipulate data columns (see Fig.6-5). For instance, you can duplicate columns or copy values between them, as well as create new columns by combining existing ones based on certain rules.

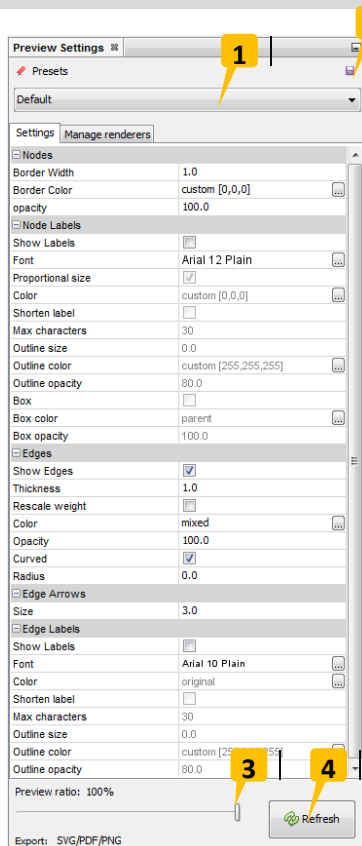## GEPHI INTERFACE: THE PREVIEW TAB

Figure 16. Preview Settings

This is where you configure the rendering settings to polish and finalize the look of your visualization.

Use the right-side Preview Settings menu to change the rendering options: size, color, transparency, 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 size of a particular node subset, the color of some edges), you need to go back to the *Overview* tab (Fig.5).

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.16-3). After you change the rendering settings you need to refresh the network preview (Fig.16-4).

Note that the background color in *Overview* won't match that of *Preview* – you need to change it separately (Fig.7-4)

Once you have tweaked the preview settings to your liking, you can save them for future use (Fig.16-2). Later you can restore them by selecting them from the drop-down menu (Fig.16-1).

SAVE AND EXPORT

Once you are satisfied with your static 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 (e.g. JSON, GDF, GEXF, GML, GraphML, Pajek .net, Ucinet DL, CSV, etc.).

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 easily be embedded into your website. One of those is *Sigma.js*. If you do not see it in the File → Export menu, you will need to install the SigmaExporter plugin from the Available Plugins tab of the Tools → Plugins menu.

The export parameters of the plugin are rather self-explanatory: you can add a title, logo, legend, author, long and short description. You can also allow users who load the Sigma.js webpage to select subgraphs based on a node attribute (Group Selector).

The Sigma.js plugin creates a folder – to view the generated interactive web visualization, navigate to that folder and open the *index.html* file.
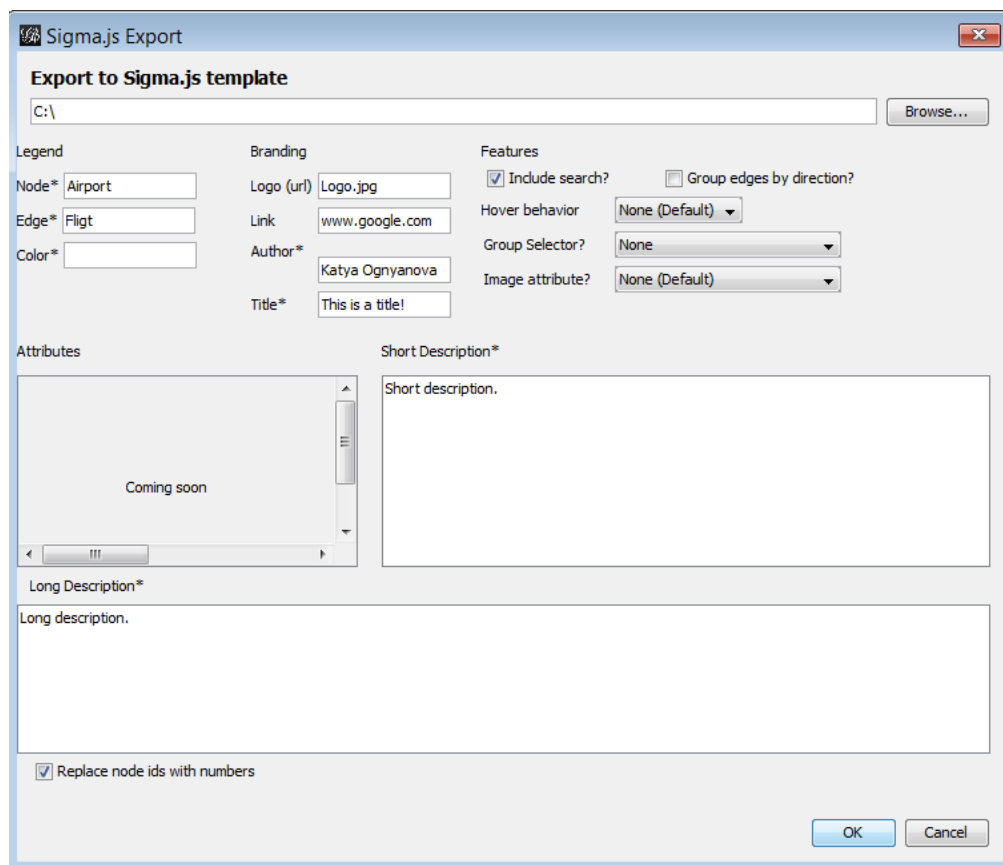

Figure 17. Sigma.js export options

DYNAMIC GRAPHS

Gephi allows you to explore longitudinal networks, as well as conduct some basic analyses on dynamic data. You can practice and get comfortable using some of these features by using randomly generated dynamic graphs: File → Generate → Dynamic Graph Example.

Gephi offers two alternative formats for dynamic data: *Timestamp* and *Interval*. Both are implemented in version 0.9 and higher, but only interval data is available in earlier versions.

Timestamps allow you to identify exact times when a node or edge exists. If you are importing this type of dynamic data from a CSV file, each cell in the timestamp column (showing when a node or edge is present in the network) will be formatted as follows:

< [time1, time2, time5] >

This timestamp indicates that a network element exists at times 1, 2, and 5. Timestamped dynamic attributes are formatted similarly:

< [time1, value1]; [time2, value2]; [time5, value5] >

This timestamp shows that the attribute had *value1* at *time1*, *value2* at *time2*, and *value5* at *time5*. The times in each case can be numbers (representing sequential time points), or they can be dates, in a *yyyy-MM-dd* format. The attribute values can be of any type (float, string, etc).

Timestamp	score
<[2000.0, 2001.0, 2003.0, 2004.0, 2009.0, 2010.0]>	<[2000.0, 1]; [2001.0, 2]; [2003.0, 3]; [2004.0, 1]; [2009.0, 2]; [2010.0, 3]>
<[2000.0, 2001.0, 2002.0, 2003.0, 2004.0, 2005.0, 2008.0, 2009.0, 2011.0]>	<[2000.0, 4]; [2001.0, 4]; [2002.0, 4]; [2003.0, 1]; [2004.0, 3]; [2005.0, 4]; [2008.0, 2]; [2009.0, 3]; [2011.0, 2]>
<[2003.0, 2004.0, 2005.0, 2006.0, 2007.0, 2011.0, 2013.0]>	<[2003.0, 3]; [2004.0, 4]; [2005.0, 2]; [2006.0, 3]; [2007.0, 2]; [2011.0, 0]; [2013.0, 0]>
<[2000.0, 2002.0, 2003.0, 2005.0, 2006.0, 2011.0, 2012.0, 2014.0]>	<[2000.0, 0]; [2002.0, 1]; [2003.0, 3]; [2005.0, 4]; [2006.0, 2]; [2011.0, 2]; [2012.0, 4]; [2014.0, 0]>
<[2002.0, 2003.0, 2005.0, 2007.0, 2014.0]>	<[2002.0, 1]; [2003.0, 4]; [2005.0, 0]; [2007.0, 3]; [2014.0, 2]>
	<[2002.0, 1]; [2007.0, 3]; [2009.0, 0]; [2010.0, 3]; [2012.0, 2]; [2014.0, 3]>

Figure 18. Timestamp format

The interval format represents periods of time when a node or edge exists. Interval data in Gephi is formatted as follows:

< [time1, time2]; [time5, time7] >

This time interval shows that an element exists from *time1* to *time2*, then again from *time5* to *time7*. If we use round instead of square brackets – e.g. **[***time1*, *time5***)** – that means the interval is open: it goes up to but does not include *time5*. Dynamic attributes can be represented as:

< [time1, time3, value1]; [time4, time7, value2] >

This indicates that the attribute was *value1* from *time1* to *time3*, and *value2* from time4 to *time7*.

Interval	Visits_dynamic
<[10.0, 10.0]>	<[0.0, 1200.0, 1820.0]>
<[10.0, 10.0]>	<[0.0, 1200.0, 2089.0]>
<[10.0, 10.0]>	<[0.0, 1200.0, 1834.0]>
<[10.0, 10.0]>	<[0.0, 1200.0, 3678.0]>

Figure 19. Time interval format

In order to import the correct type of dynamic data (timestamp vs. interval) from a CSV file, you need to select it from Data Laboratory → Configuration (see Fig.20-1 & 2) before the import.
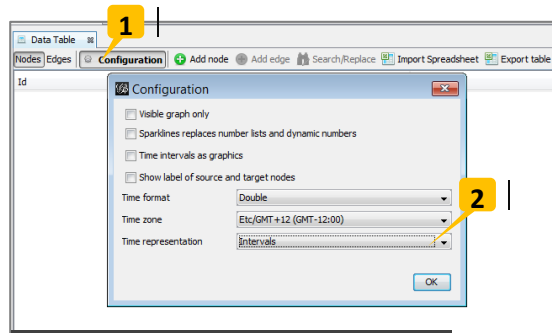

Figure 20. Dynamic configuration

When you import dynamic network CSV files, you also have to remember to select the correct variable type for each dynamic column. The type for timestamps of nodes or edges is *TimestampSet,* and that of dynamic attributes is *TimestampXMap*, where *X* is the attribute type (e.g. *IntervalFloatMap*). The type of interval data for nodes or edges is *IntervalSet*, and that of attributes is *IntervalXMap* where *X* is the attribute type (e.g. *IntervalFloatMap*). In versions of Gephi prior to 0.9, the interval format is named *TimeInterval.*

You can also generate dynamic variables in the Data Laboratory. One way to do that is to combine two columns with Merge columns (Fig.21-1). One column should contain the time when a node or an edge entered the network, the other the time when it exited the network. Select the columns containing the start and end points, and choose Create time interval from the dropdown menu (Fig.21-2). Attributes can also be converted to dynamic from the Convert column to dynamic button (Fig.21-3)
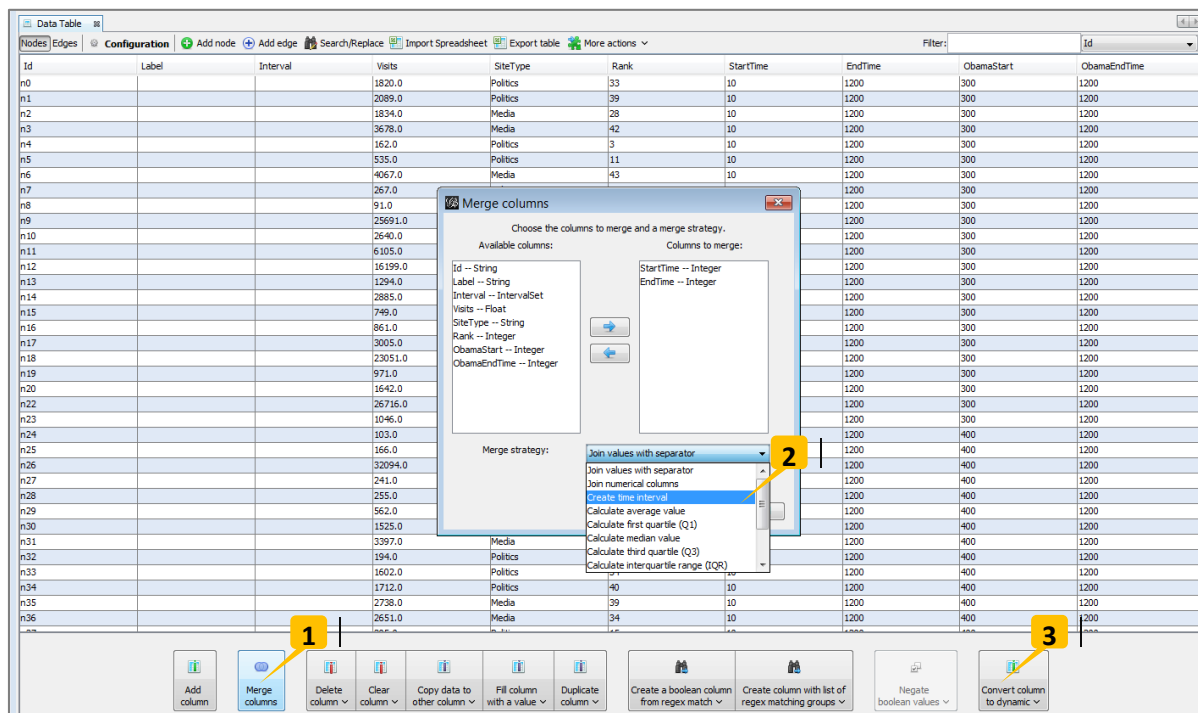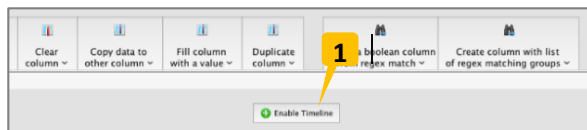

Figure 21. Dynamic options in the data laboratory

Figure 22. Enable Timeline

After the dynamic data has been loaded or generated, you can click on Enable Timeline (Fig.22) to begin using the dynamic features. Once you return to the Overview tab, you will be able to play an animated sequence of your network as it progresses over time (Fig.23-2). You can also calculate dynamic network statistics such as node degree (Fig.23-5). Many of Gephi's features support dynamic settings. For instance, with the Auto Apply appearance button (which needs to be toggled on at Fig.23-3), nodes can change size or color over time as some dynamic attribute of theirs changes. Many layouts can run continuously as well, so that the node positions will also change as edges enter and exit the network.

Additional options for dynamic visualization are available in the Animation Settings menu (Fig.23-1). You can change the speed of the play and whether elements appear cumulatively or only if present within a moving time window (*one bound* vs. *two bounds* mode, Fig.23-4).

One final thing to note here: unfortunately, at present there is no way to export a video of the dynamic visualization you have created. Your best bet is to record your screen while it plays using a screencasting software (for example camstudio.org).
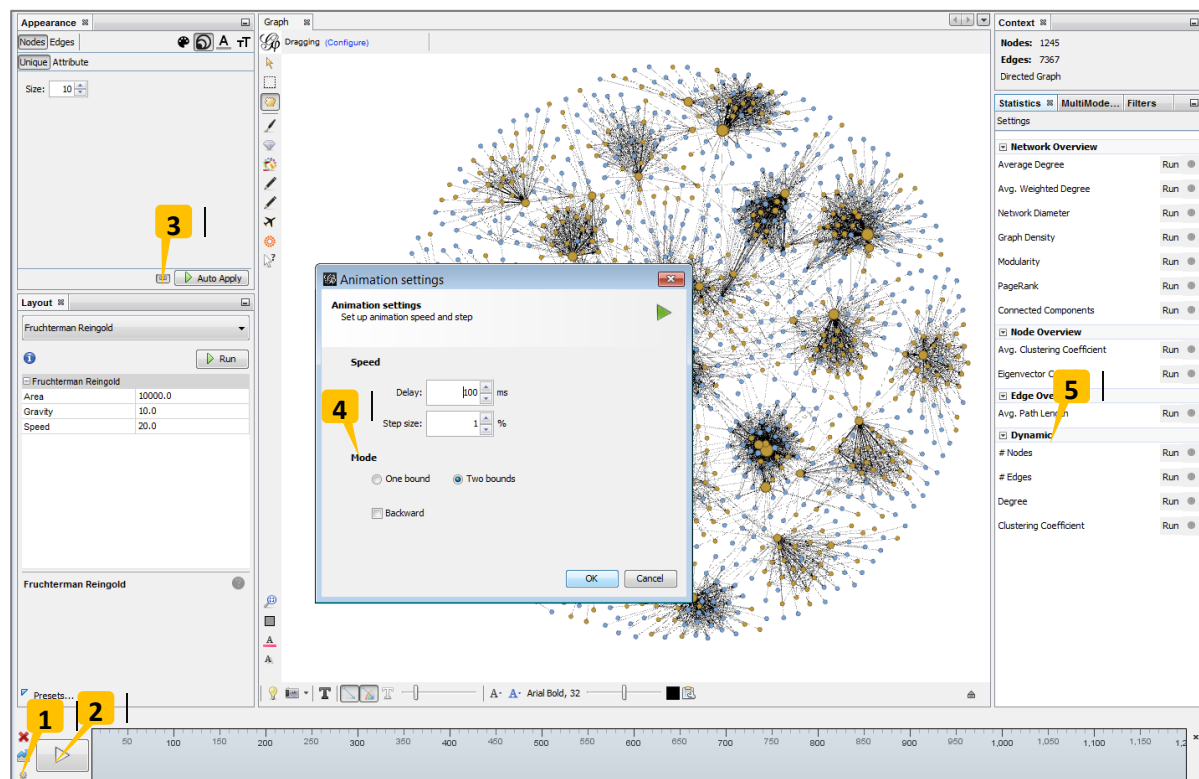

Figure 23. Dynamic network overview