# Network Analysis in Gephi Instructions

Mackenzie Brooks

July 2021

Gephi website: <https://gephi.org/>

Gephi tutorial: <http://miriamposner.com/dh101f14/wp-content/uploads/2014/11/Creating-a-Network-Graph-with-Gephi.pdf>

## Preparing your data

Gephi requires two separate spreadsheets: an edge list and a node attribute table. Use Excel or Google Sheets to prepare these spreadsheets.

### Edge List

* The Edge List is what tells Gephi about the connections/relationships between two nodes/people. The spreadsheet contains at least two columns and they must be titled **Source** and **Target**.
* Each row in your spreadsheet represents one edge or line between two nodes. You must decide for each project how you will represent relationships. For example, a correspondence network might have multiple rows that look identical if the same person sent letters to the same recipient multiple times. It is your choice how to record your connections, but you should be consistent.
* If necessary for your data, you can include a **Type** column that indicated the direction of an edge. The value of this field can be “Undirected” or “Directed.” An undirected relationship goes both ways (example: parent and child) and a directed relationship goes in one direction (parent sends a letter to a child, but child does not reply).
* You can add other columns to indicate other information about the relationship. For instance, **RelationshipType** might include values like Family, Marriage, Friends. You should be consistent in how you format these terms, as you can use them to filter or color your network graph.

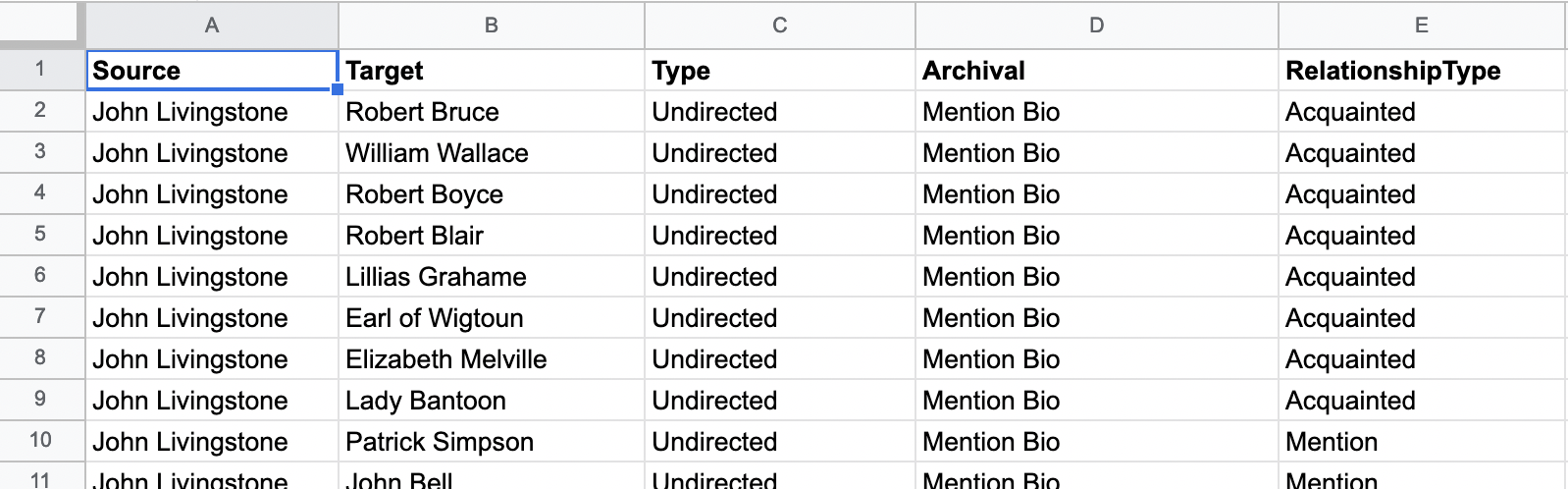


Figure 1. Example Edge list in Google Sheets

### Node Attribute Table

* The Node Attribute Table is what tells Gephi information *about* the people in your network. The spreadsheet must contain at least two fields, one called **ID** and one called **Label**. ID contains an identification code or just the name of the person, however it *must* match whatever is found in the Edge List. Label is a more human-readable version of the name and it will show up in your network graph. Having different IDs/labels is useful when you have people with similar names or titles. Example: John1, John 2 could be IDs for two different people named John Smith.
* Each person should be listed only once in your Node Attribute Table, however you can list as much information about them as you want. Again, be consistent in how you format the information and remember that you can use this information to style your network graph.

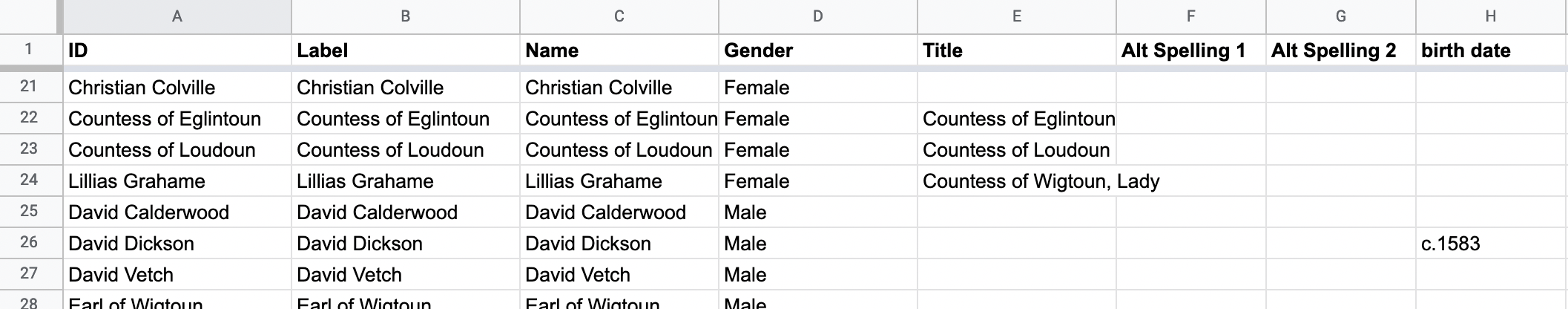


Figure 2. Example Node Attribute Table in Google Sheets

## Getting Started in Gephi

* To start using Gephi, your data should be on your computer. You may have to download it from Google Sheets or Box. If you’re working in Google Sheets, go to **File > Download > Comma-separated values** to download your data in .CSV format.
* Once your data is ready, open Gephi. You can create a New Project to get started – it will function as a Workspace where you work on your graph. You should see a tab that says Workspace 1 at the top left of your screen. (Figure 3)
* To add your Edge List, go to **File > Import Spreadsheet**. (Figure 4).

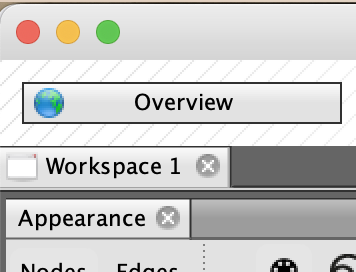
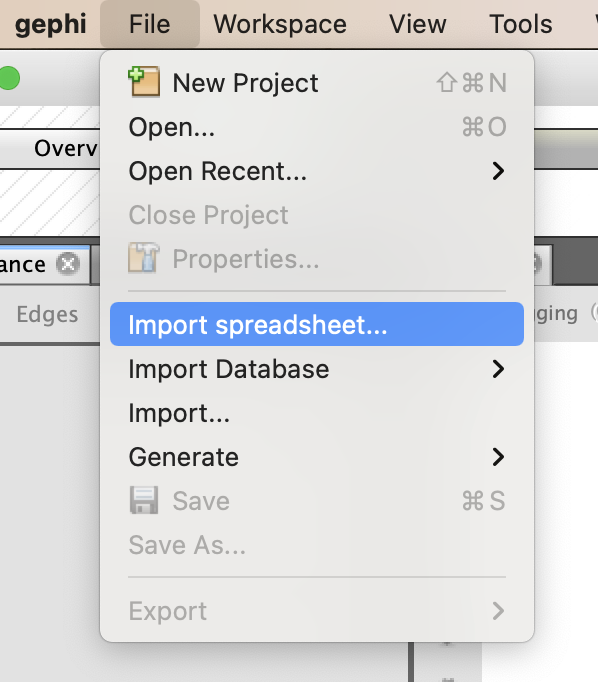
 

Figure 3. Workspace tab. Figure 4. File > Import Spreadsheet

* Once you select your Edge List, a wizard will take you through the import process.
  + Gephi should detect that you are importing an Edge List in a CSV format. If the Preview looks like your data, click **Next**. (Figure 5)
  + The Import Settings step will let you change the data type of the fields you’re importing. Most of the time you can leave everything alone, however, if you are importing fields with integers or other special situations, you might want to change the drop down. For example, if you had a column with people’s ages, you might change the type to “integer” so you could make calculations based on age. (Figure 6)
  + Click **Finish** to continue the import process. You should then see an **Import Report** screen. It’s worth taking a look to make sure that everything looks right. The number of nodes and edges should correspond with what you know about your data. The Graph Type should also be Mixed, Directed, or Undirected based on what you’re recording about your relationship types.
  + Finally, if this is the first spreadsheet you’re importing, the **New Workspace** radio button should be selected. Press **OK**.

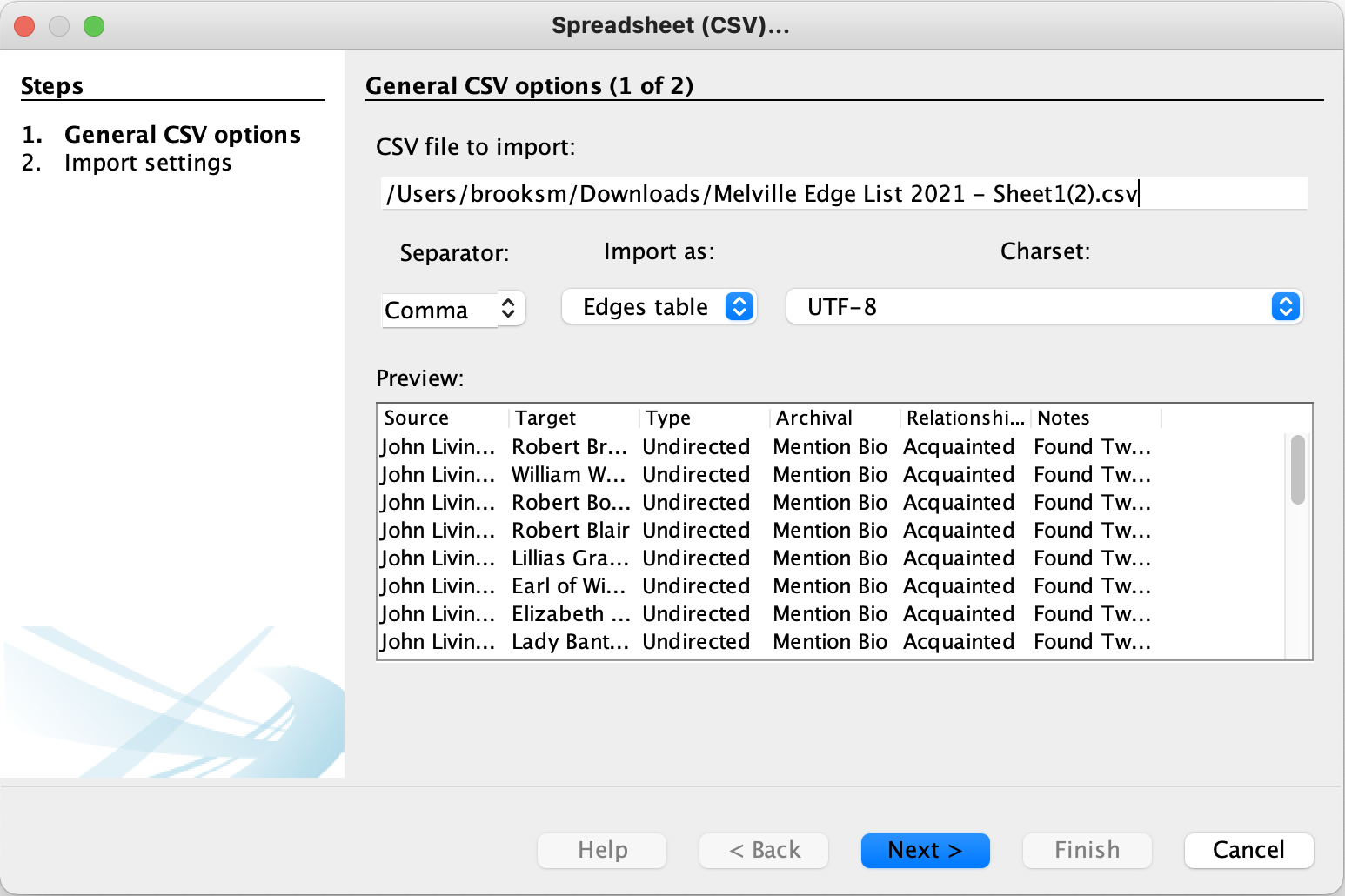


Figure 5. Edge List Import Wizard

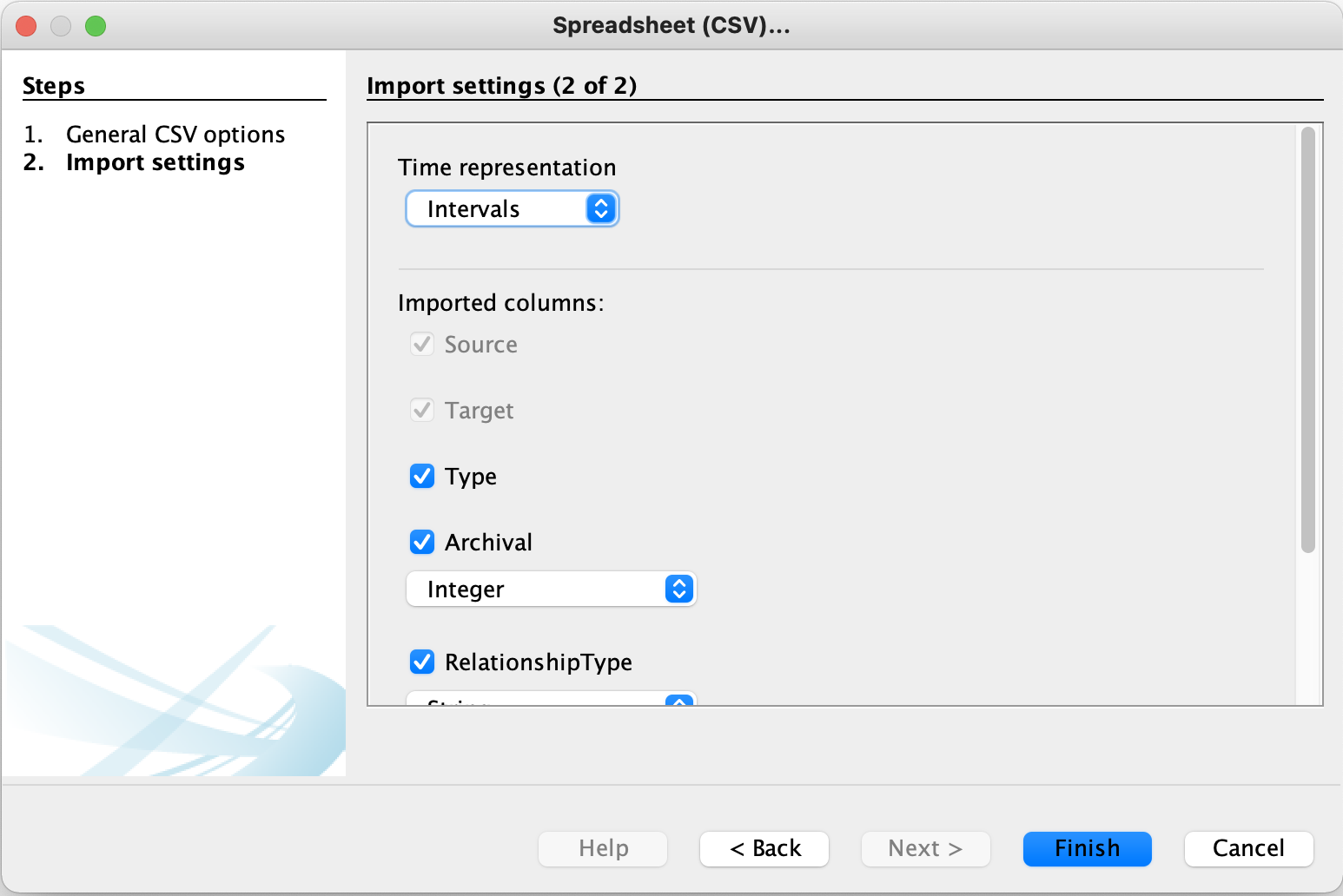


Figure 6: Edge List Import Wizard

* You should see a graph now, but it won’t have any labels. To add labels, we must import the Node Attribute table.
  + Follow the same process as the Edge List import above: **File > import spreadsheet** to select the Node Attribute CSV file.
  + Gephi should detect that you are importing a Node Attribute table.
  + Click next to proceed through the wizard. (Figure 7)
  + Again, in Import Settings (2 of 2) you can change the data types for any fields. For instance, birth year should be an integer. Click **Finish**.
  + **IMPORTANT!** On this screen you must select **Append to existing workpace** in order to connect your Node Attribute Table to your Edge List. (Figure 8)
  + You should also consult the Issues box at top to see if any errors have been detected. In Figure X, you can see that there are duplicate entries for two people. Consult both the Edge List and Node Attribute spreadsheets to look for duplicates.

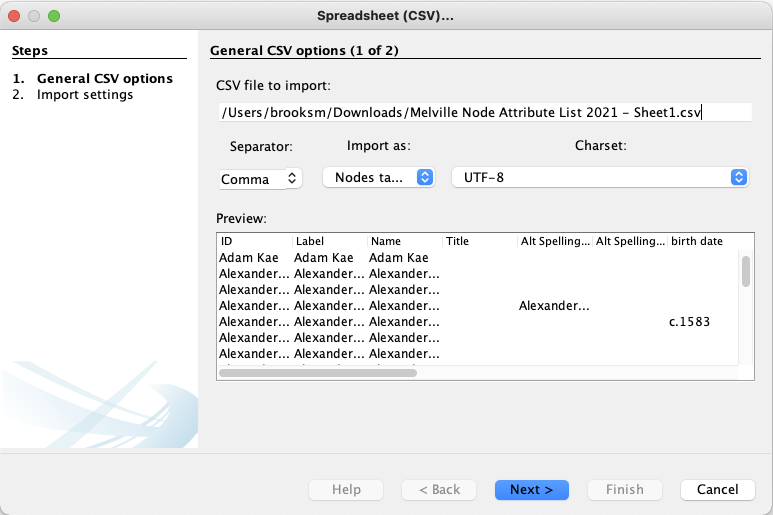


Figure 7. Node Attribute Table Import Wizard

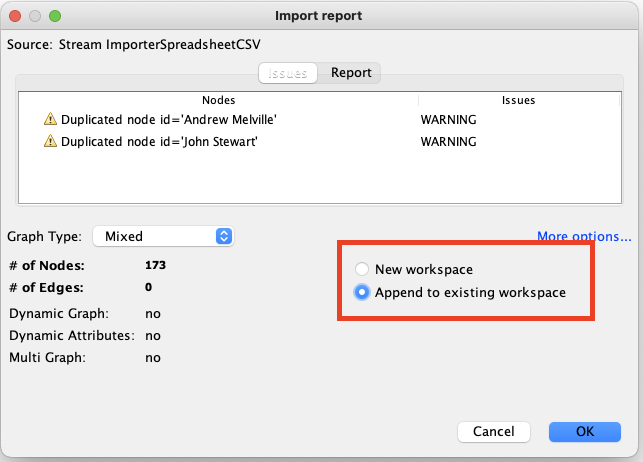


Figure 8. Important step. Choose Append to existing workspace.

## Viewing your graph

You should now see a network graph! You can scroll with your mouse or trackpad to zoom in and out, or use the mouse to click on nodes and drag them around. There are a number of options to change the way it looks.

* To see the labels, look for the dark T at the bottom left of the graph viewing pane.
* To adjust the size of the labels, use the slider to the right of the font selection.
* To recenter the graph, use the magnifying glass icon .



Figure 9. Tools for adjusting your graph

## Styling your graph based on attributes

You can use both attribute and edge data to change the way your graph looks using the **Appearance** window in the top left corner of the screen. (Figure 10)

1. First, select whether you want to change the Nodes or Edges. Click on the word “Nodes” or “Edges” but note that the rest of the menu will not change appearance.
2. Next, decide what part of the network you wish to change. In order, the icons represent color of nodes, size of nodes, color of labels, and size of labels.
3. Your next step is to choose how you’d like to use your data to change the graph. **Unique** will affect your whole graph. Partition allows you to change the appearance of the graph based on categories in your data. Figure 11 shows the results of a choosing the Gender field from your data set. **Ranking** is used when you have a scale or range in your data and you want the color of nodes to reflect that range. For instance, if you choose Degree, the nodes with the highest number of connections will be the darkest and the nodes with the fewest connections will be light.
4. Each option from step 3 involves making adjustments via a dropdown menu or color palette.
5. You must click **Apply** to see the changes

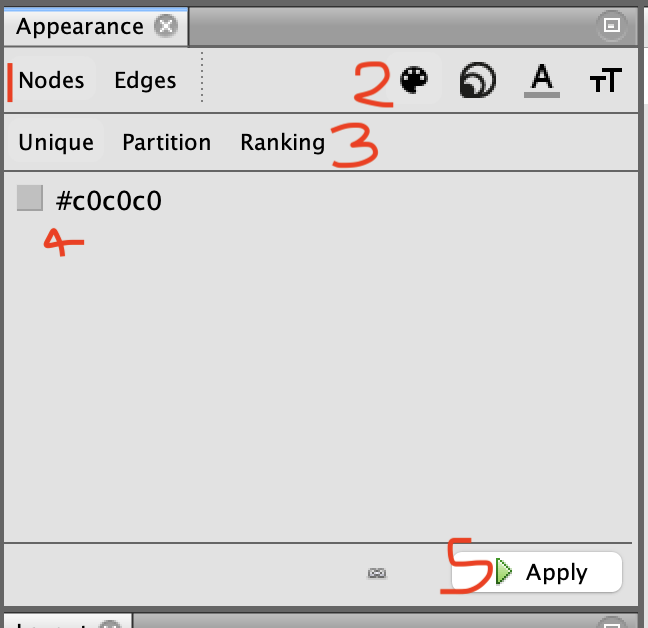


Figure 10. Appearance toolbar

## Layouts

To change the distribution of the nodes in your graph, use the Layout menu options. Each option in the dropdown menu is a different algorithm that causes the nodes and edges to redistribute based on particular criteria. You can use the blue “i” to learn more about each algorithm. You must click **Run** to implement the layout. Note that you can adjust the options for the algorithm as well. In Figure 11, we could increase or decrease the margin by click into the box with 5.0 and replacing it with a higher or lower number.

There are two layouts that improve the readability of your graph: **Label Adjust** creates space around each label so they are easier to read and **Noverlap** creates space around the nodes themselves.

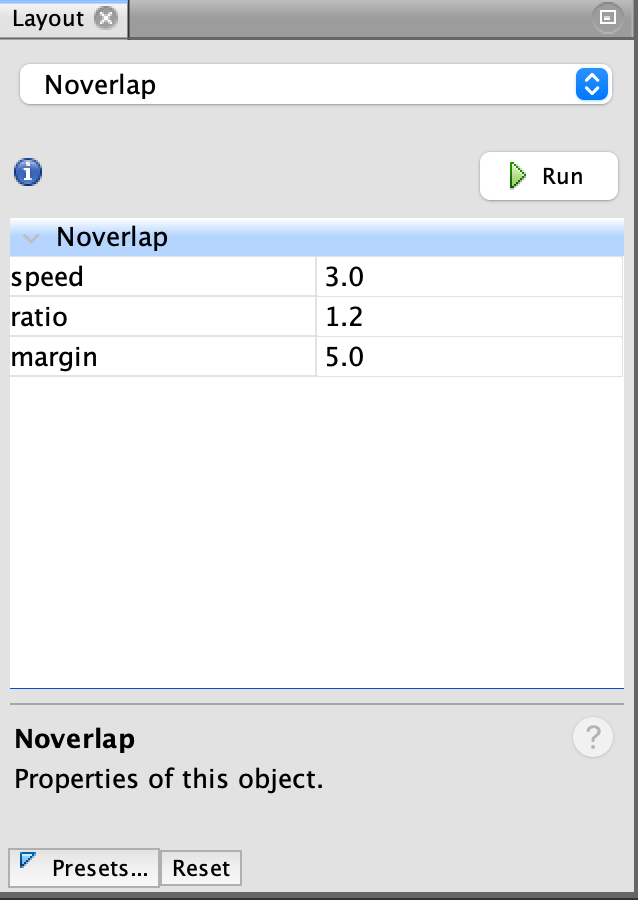


Figure 11. Layout window with the Noverlap algorithm.

## Statistics

Gephi provides a number of ways to calculate statistics about your graph. The Statistics menu is found on the right side of the application. The general process is to press **Run** to calculate. Usually, a window pops up to show you a data visualization of the results. If you want to style your network based on statistics like Degree or Eigenvector, you should run that option before you try to use the Appearance menu. Here are a few of the common statistics:

* Degree – measure the number of connection that each node has.
* Density – measures how close a network is to complete. A complete graph that has all possible edges and density is equal to 1.
* Eigenvector – a measure of a node’s importance
* Network Diameter – a way to measure betweenness.

Once you’ve run statistics, the results will be added to the Node Attribute Table and can exported out of Gephi for use in other programs (a graph in Excel for instance).

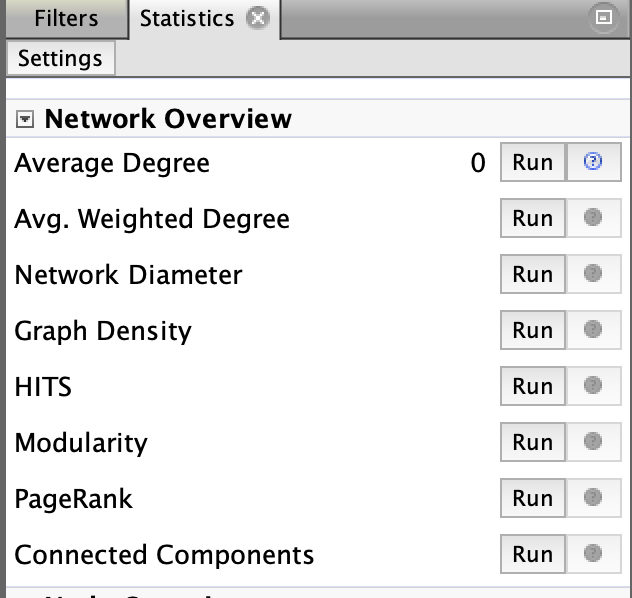


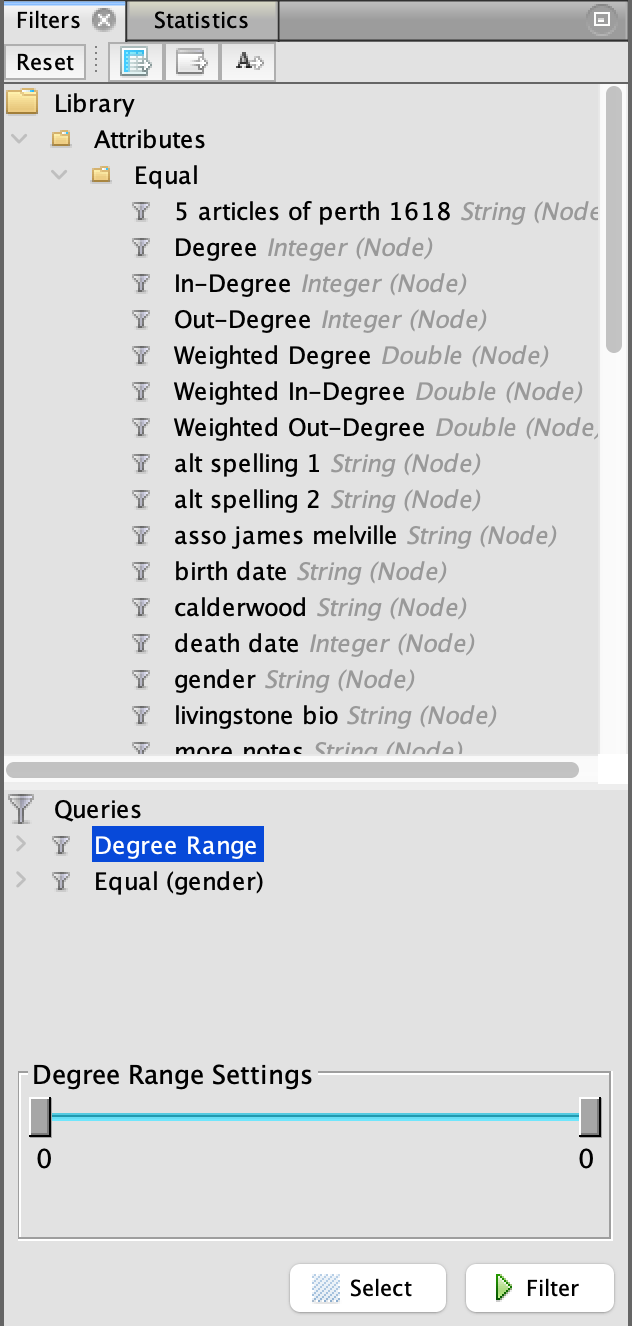
Figure 12. Statistics window

## Filters

Filters provide a way to filter or edit out parts of your graph temporarily. Perhaps you want to remove all the men from your network in order to get a clear picture of how women are connected – this is a job for a filter. Filters can get complex, but to do something like filter based on a field in your Node Attribute Table, follow these steps:

1. Click into the **Filters** window (it’s next to Statistics).
2. Double click through the folder structure to get from **Attributes > Equal > gender**.
3. Double clicking on gender creates a query in the box below.
4. Enter the value of the field you’d like to see, in this case, we’ll entire “female” to just select nodes that have the attribute of female in the gender field. Press **OK** to apply.
5. Click the **Filter** button with the green arrow to actually apply the filter.
6. Press **Stop** to turn off the filter.

You can apply multiple filters if necessary, including filters that pull from statistics about your graph. Perhaps you want to see all the men in your graph who are in the clergy and have degree number between 5 and 10. That is possible with filters.

 Figure 13. Filters window with two filters/queries applied.

## Exporting your graph

To export your graph for use in another program or publication, use the **Preview** button at the top of the page. This brings you to a screen where you can make adjustments to how your graph looks, then **Export** it as a SVG, PDF, or PNG file.

You must press Refresh to see the graph. You may want to adjust the following:

* Show labels (check the box to see labels)
* Font
* Curved (affects the edges/lines).

Plan to experiment with the best format for your graph.

* PNG – this is a typical image file (like GIF of JPG) and while it is web-friendly, it will not shrink or expand with ease. It might be pixelated depending on the size of your graph.
* SVG – this is a scalable vector graphics file, meaning it will stay crisp no matter how big or small it gets (like a logo). It can be a little weird to work with, but it is a good option if you’re planning to publish your graph. Note: there is a bug with exporting SVG files that crops some of the margins of the graph. You can use a program like Inkscape or another vector graphics software tool to expand the canvas of the image. All of the graph information is still in the file, it just needs a bigger canvas.
* PDF – creates a document page on which the graph sits.

## Fixing SVG Canvas in Inkscape

To ensure that your graph can be see in the SVG file, you must expand the canvas using SVG editing software. [Inkscape](https://inkscape.org/) is a free tool for this purpose. Visit the website and download the appropriate version for your operating system.

1. Open Inkscape