Network Tutorial - DH@Ox2023

In this case we will be using Gephi to build and explore a network.

We are going to use a very simplified set of data that might represent an area of interest: communication within and between two communities.

The basic steps we will explore:
1. Examining the data files
2. Data import from text files
3. Data cleaning and transformation
4. Visualization and layout of a network
5. Statistical analysis and refining of visual
6. Export of data, images, and networks.

Raw data:

File 1: letters_edges.csv
File 2: letters_nodes.csv

---

1. Examining the data files

Having a look at the files in a text editor or spreadsheet program (such as Excel) is sometimes a useful first step.

It can help you check to formats and separators in a file are as you might expect.

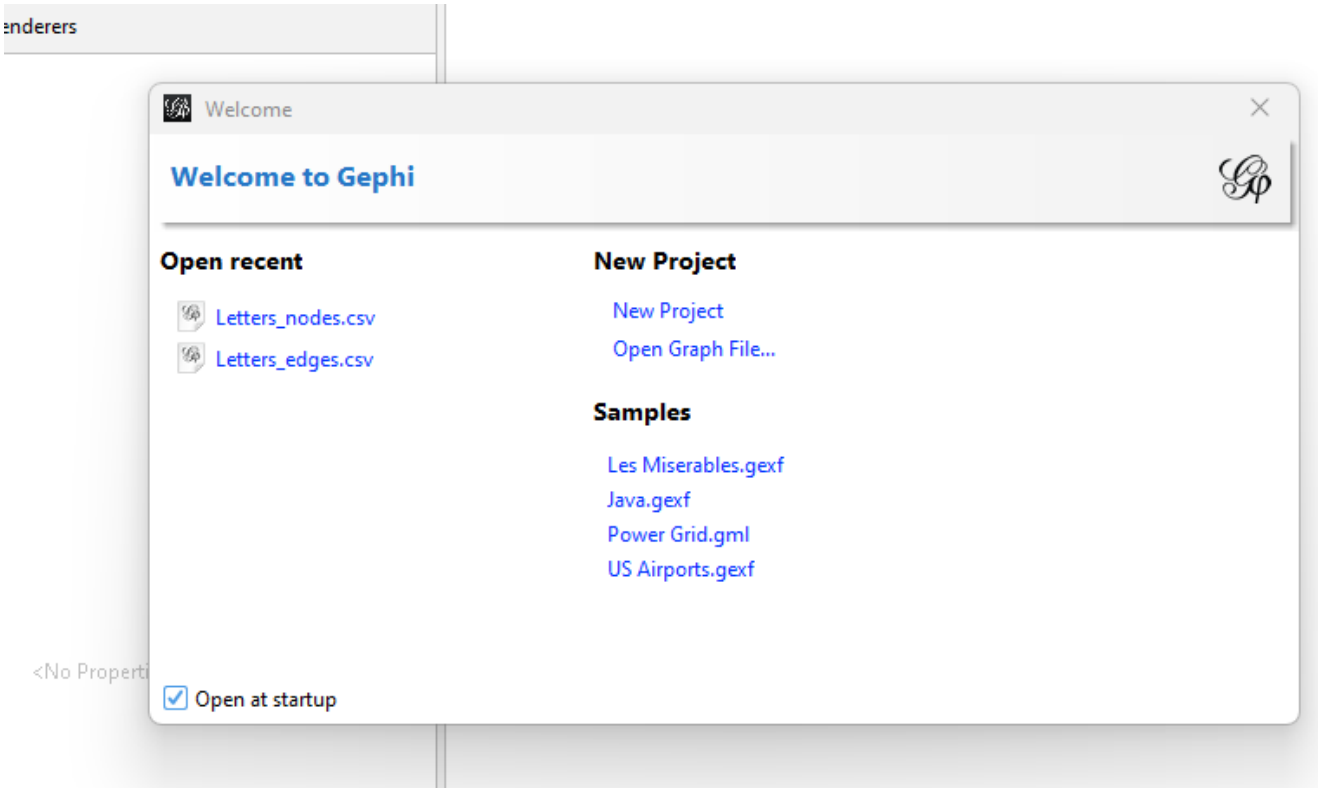
It can also help identify if there are errors preventing loading into other programs.

Don't save and changes as you close these.

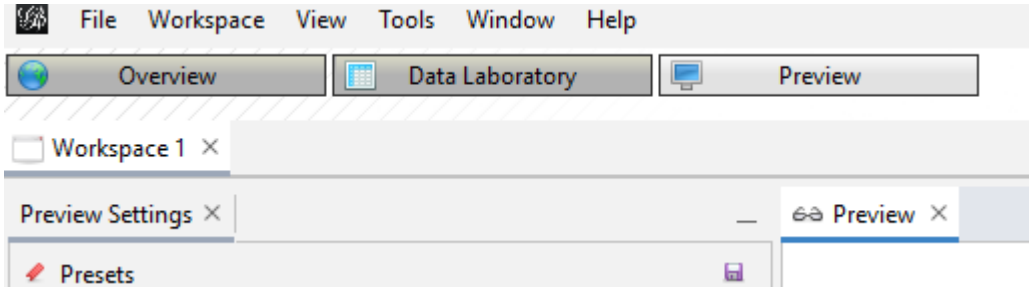
> In this case `letters_edges.csv` shows us a series of individual entries with `source` `target` and `weights` . There is only a single entry for each combination of nodes (no repeat entries) and no reversed entries (node 2 -> node 1 etc.). Do you think the network is **directed** or **undirected?** Do you think it is has **summed edges** or is a **multigraph?**

2. Data import from text files

Start by opening the Gephi Program on your machine, and selecting a new project.

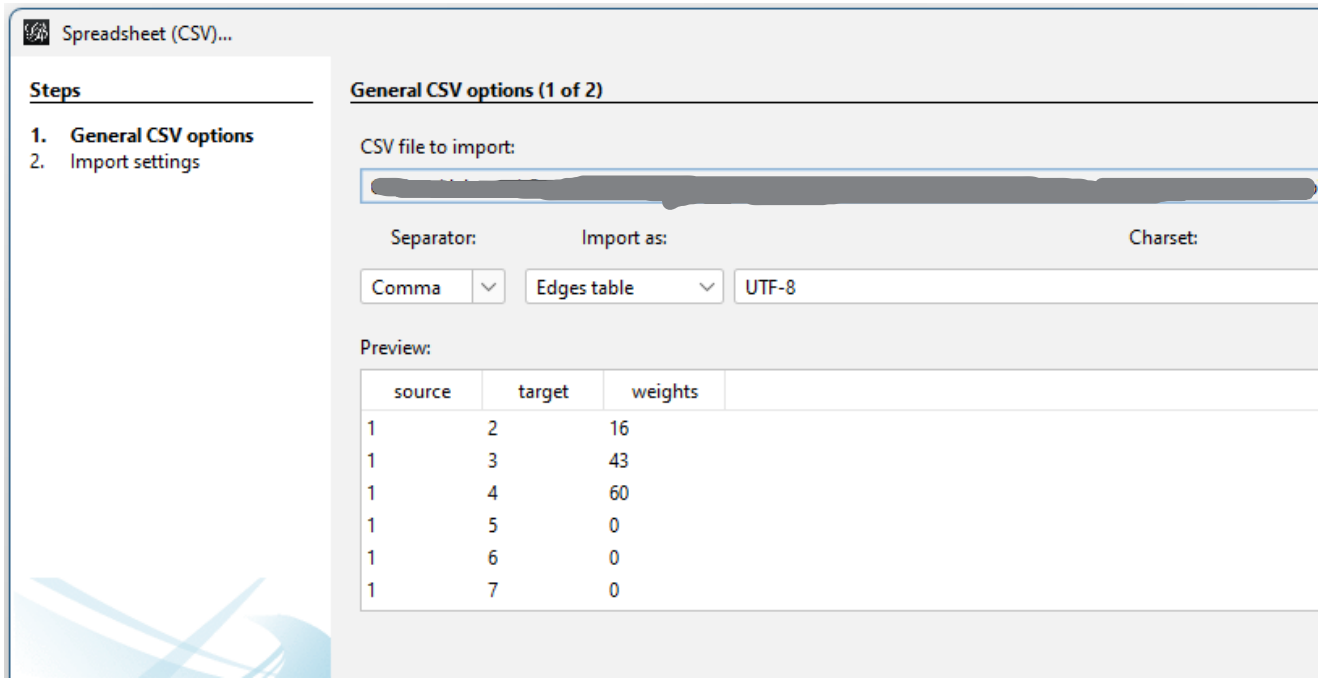


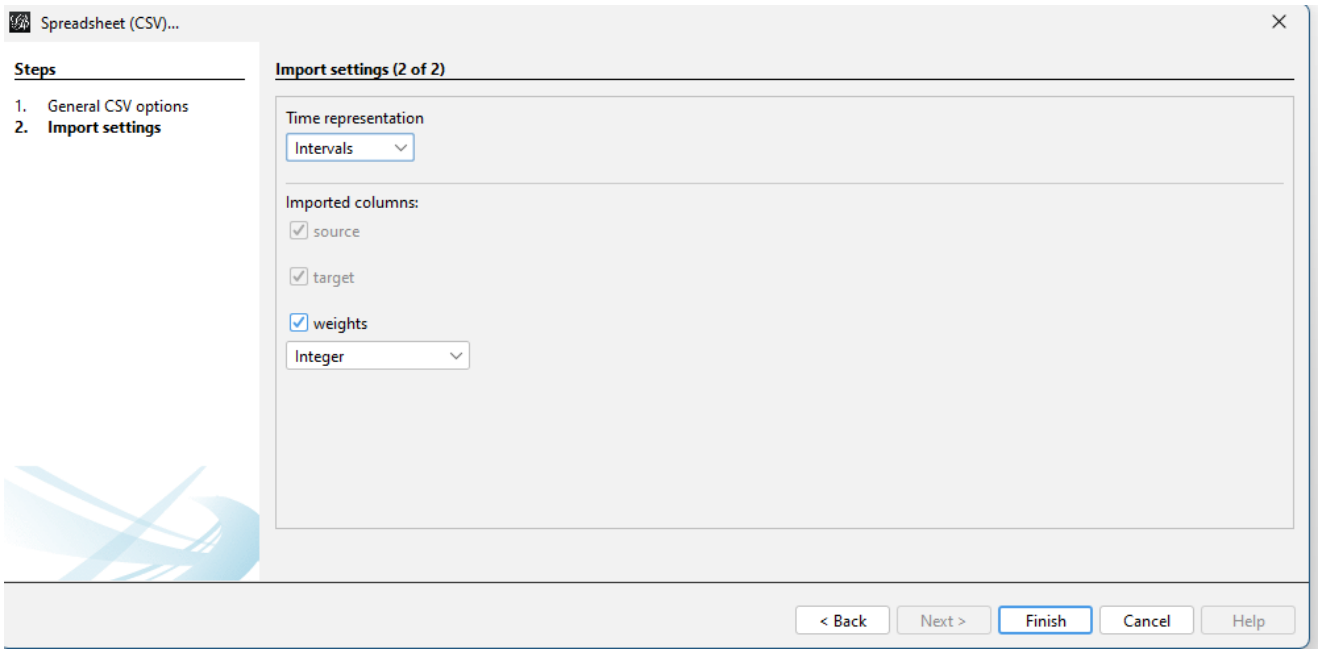
Take a moment to explore the menu options available, including the range of options in the `File` menu for creation and import of a range of networks and file types.

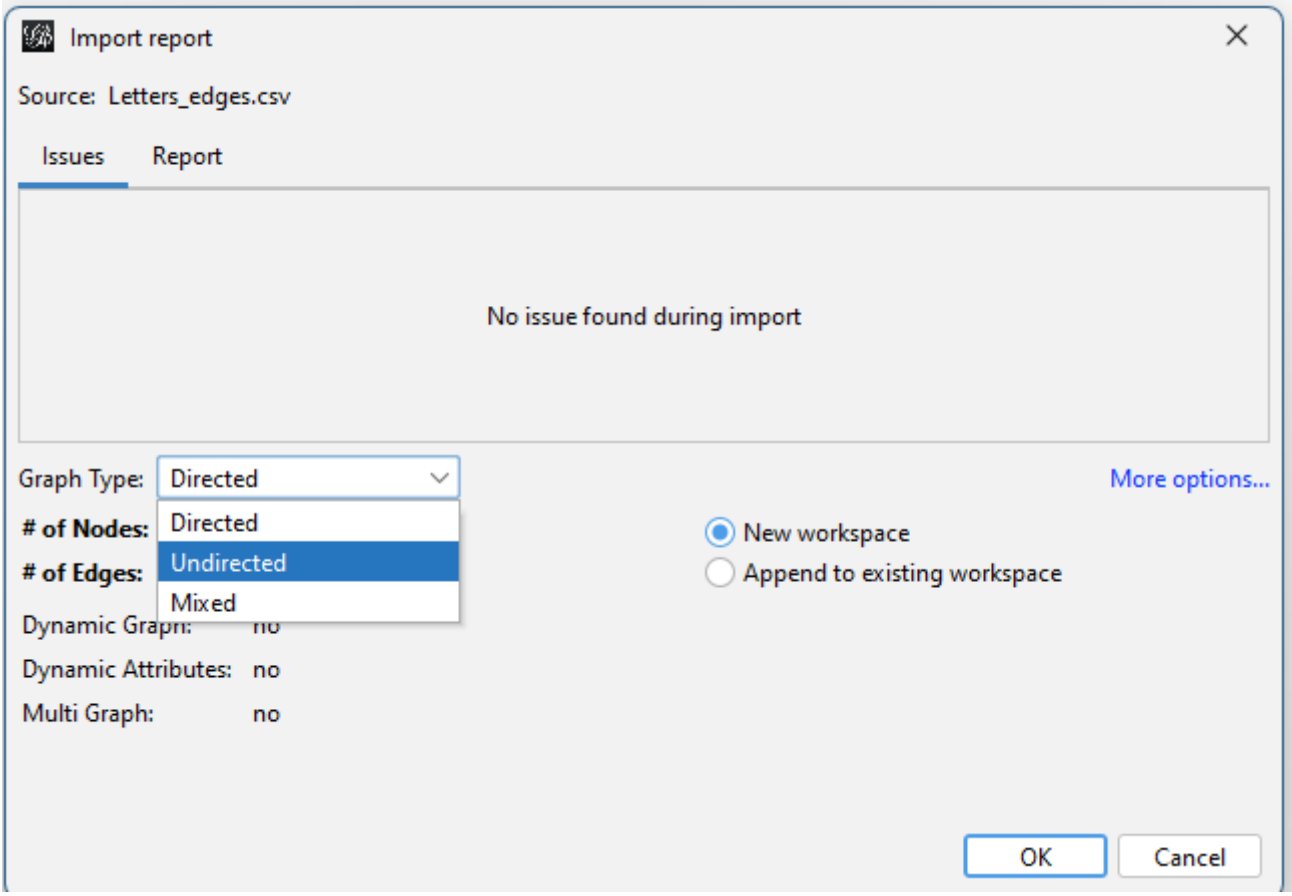


We will do a series of imports, one for our edges and one for or additional node characteristics.

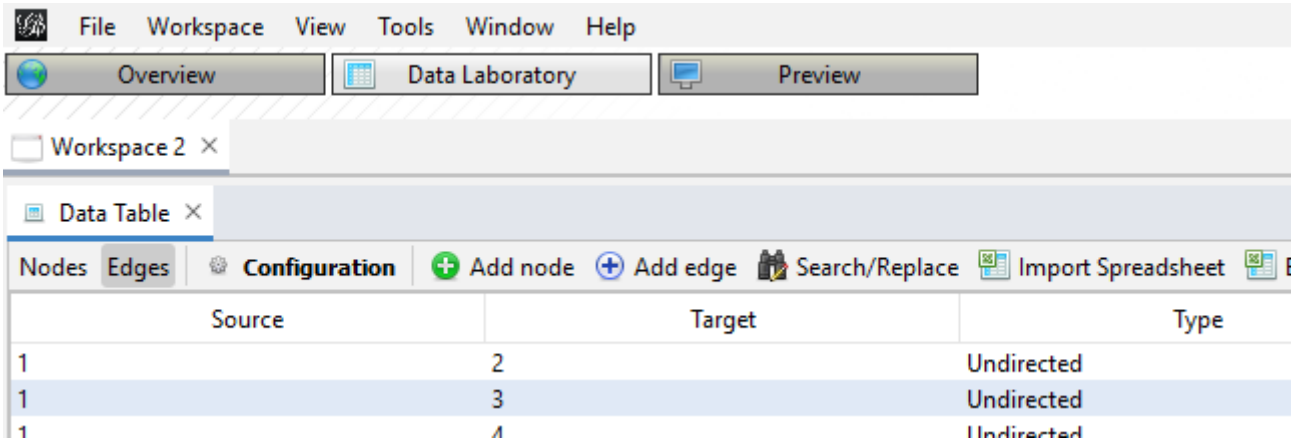
`File > Import Spreadsheet` for letters_edges.csv







In response the questions raised in the previous section, single edge for each possible combination of nodes, with a range of `weights` would suggest the data does not represent a multigraph. The lack of any source node numbers higher than target nodes, plus no repetition of `source` and `target` in reversed order, suggest this is an **undirected** network (edges represent sum total of communication between nodes).

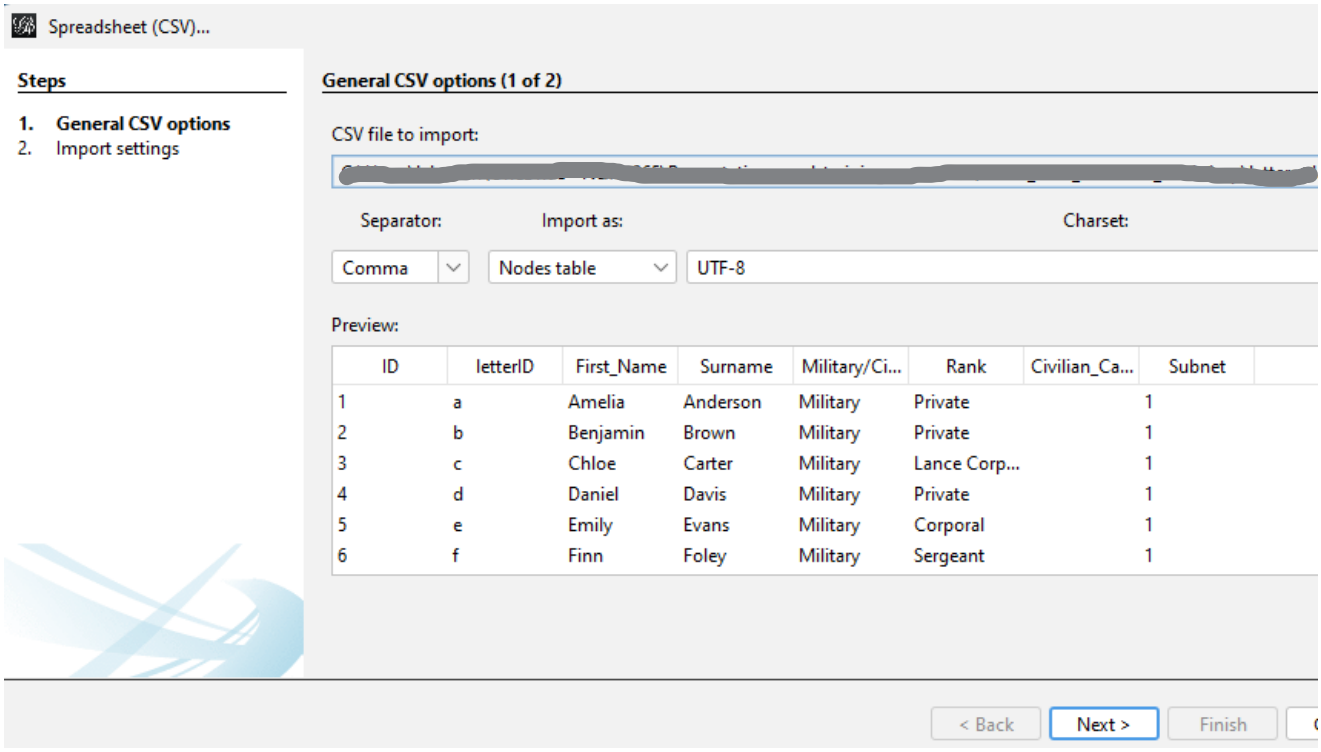


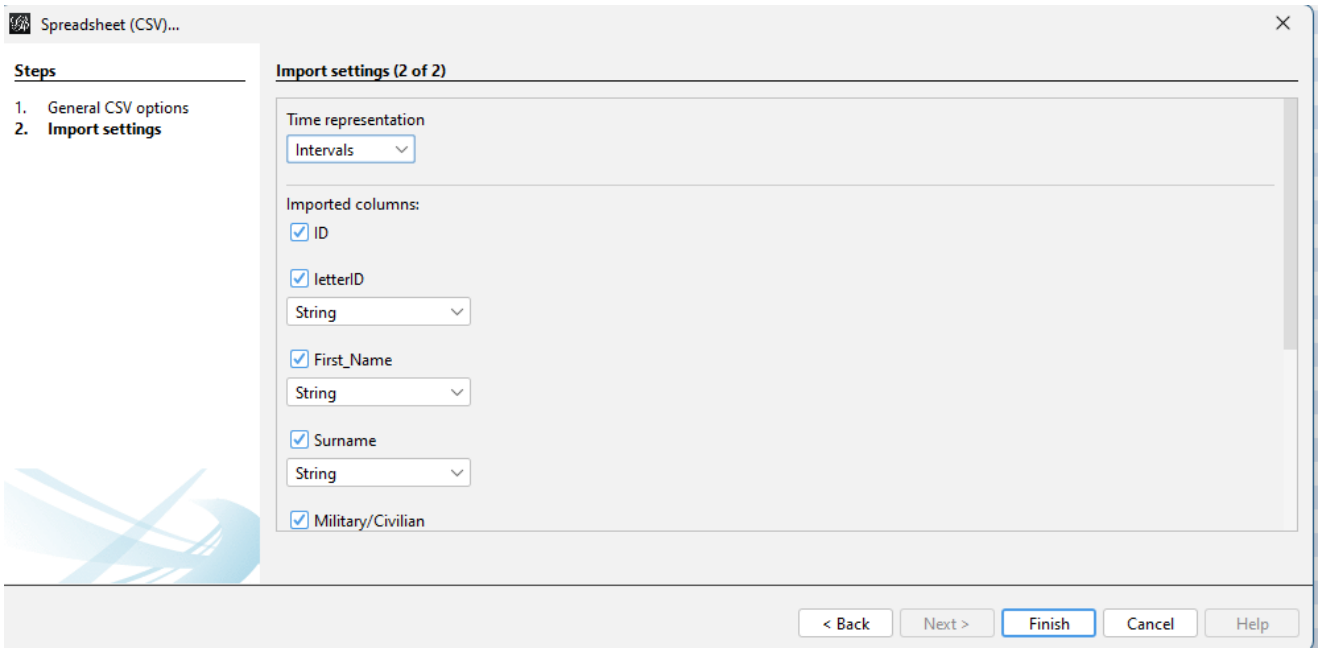
Once the data is loaded, this can be found in the **Data Laboratory** tab of the Gephi interface (under **Edges**).

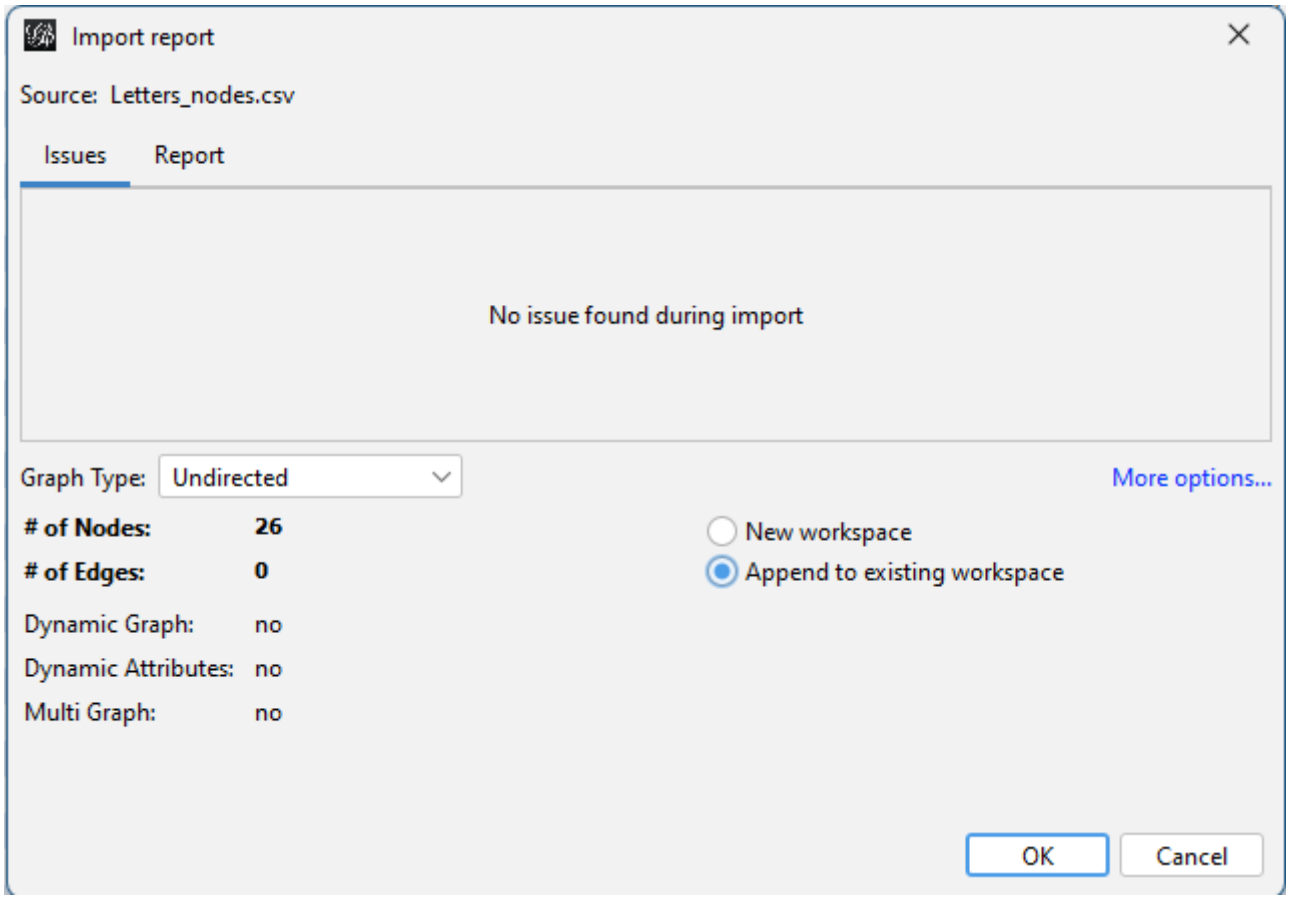
One thing to Note:
- There the same value of 1.0 for every edge `weight` at the moment, plus there is a column called `Weights` imported. How could we fix this.

Now we have our nodes and edges, we can add further information about the nodes they connect.

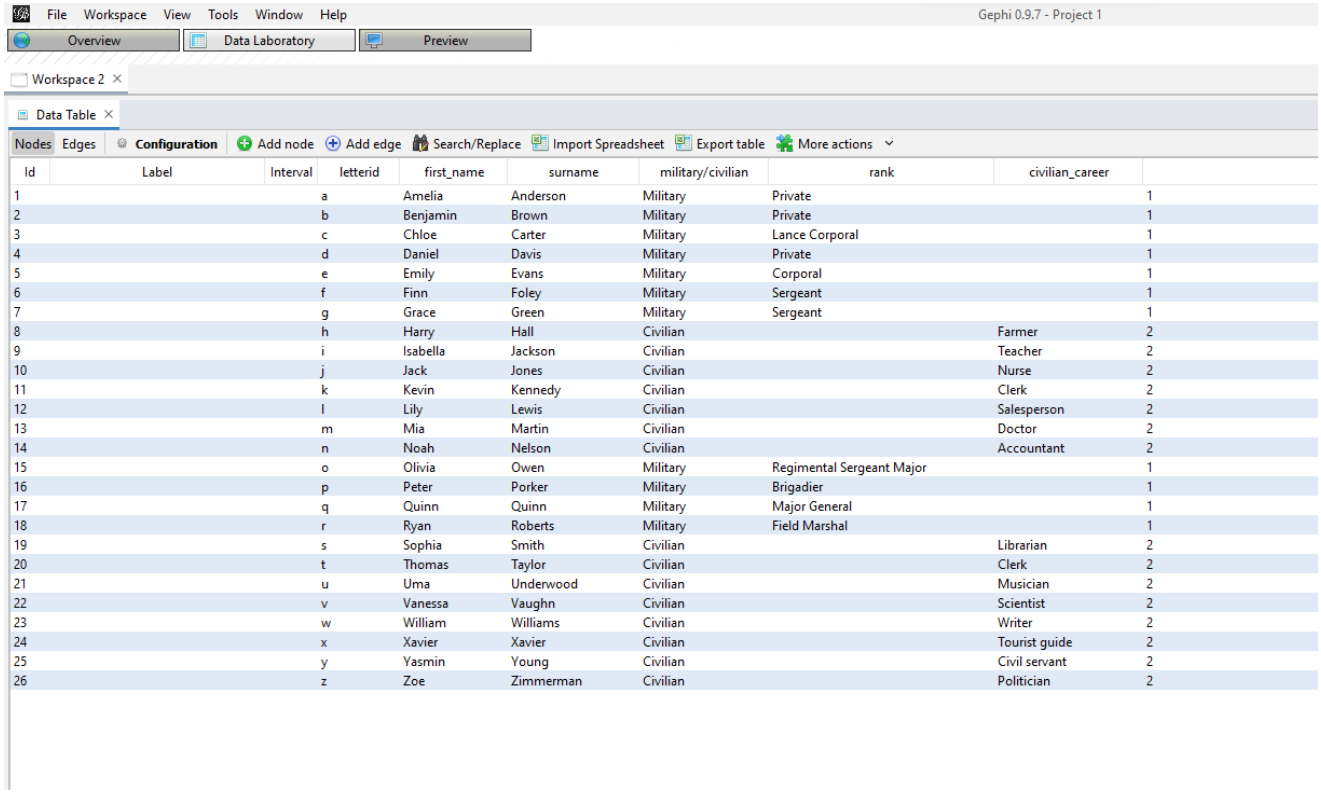
`File > Import Spreadsheet` for letters_nodes.csv







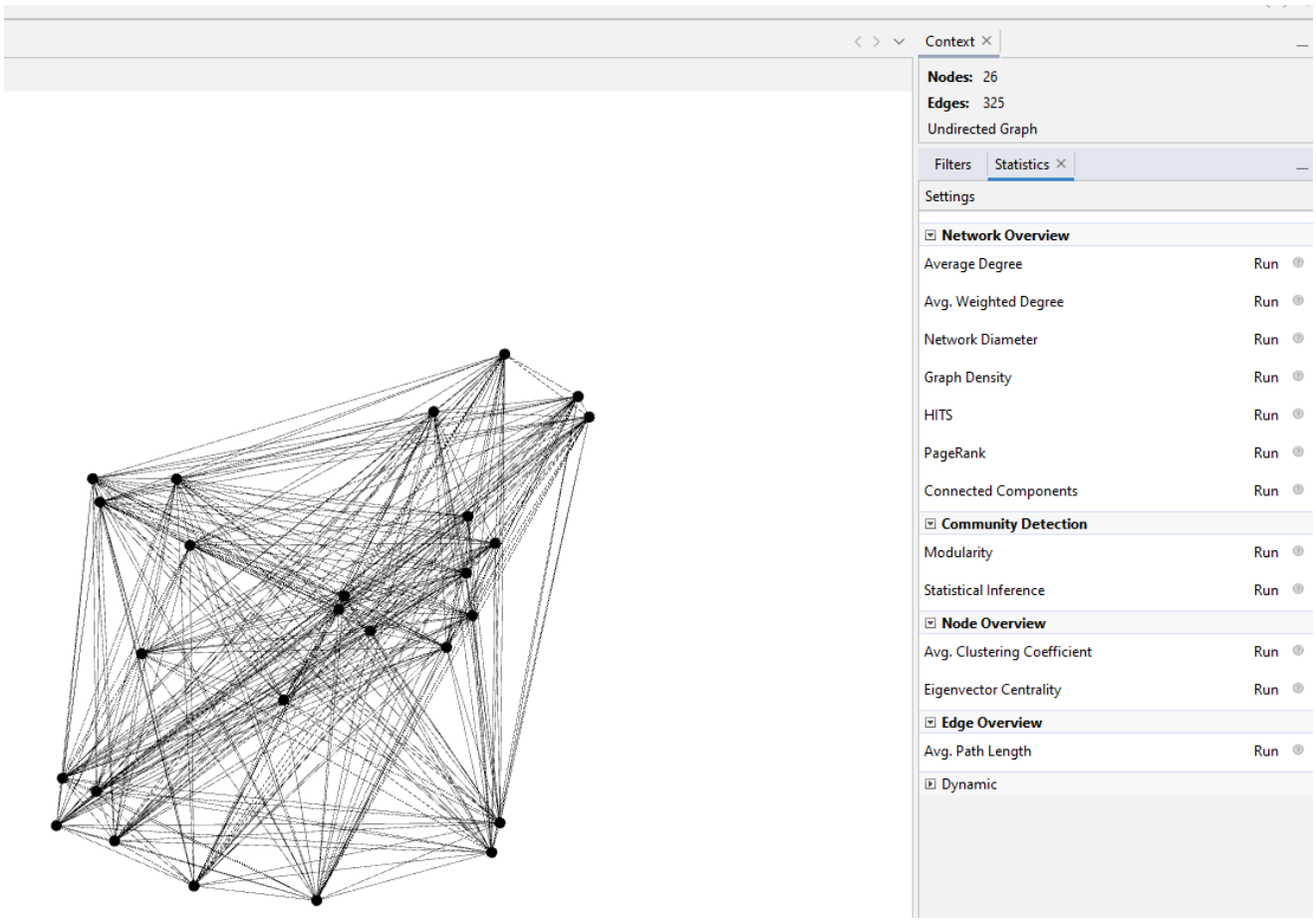
A couple of things to notice again:
- There are no entries in the `Label` column
- Names are provided as two different columns



Id	Label	Interval	letterid	first_name	surname	military/civilian	rank	civilian_career	
1			a	Amelia	Anderson	Military	Private		1
2			b	Benjamin	Brown	Military	Private		1
3			c	Chloe	Carter	Military	Lance Corporal		1
4			d	Daniel	Davis	Military	Private		1
5			e	Emily	Evans	Military	Corporal		1
6			f	Finn	Foley	Military	Sergeant		1
7			g	Grace	Green	Military	Sergeant		1
8			h	Harry	Hall	Civilian		Farmer	2
9			i	Isabella	Jackson	Civilian		Teacher	2
10			j	Jack	Jones	Civilian		Nurse	2
11			k	Kevin	Kennedy	Civilian		Clerk	2
12			l	Lily	Lewis	Civilian		Salesperson	2
13			m	Mia	Martin	Civilian		Doctor	2
14			n	Noah	Nelson	Civilian		Accountant	2
15			o	Olivia	Owen	Military	Regimental Sergeant Major		1
16			p	Peter	Porker	Military	Brigadier		1
17			q	Quinn	Quinn	Military	Major General		1
18			r	Ryan	Roberts	Military	Field Marshal		1
19			s	Sophia	Smith	Civilian		Librarian	2
20			t	Thomas	Taylor	Civilian		Clerk	2
21			u	Uma	Underwood	Civilian		Musician	2
22			v	Vanessa	Vaughn	Civilian		Scientist	2
23			w	William	Williams	Civilian		Writer	2
24			x	Xavier	Xavier	Civilian		Tourist guide	2
25			y	Yasmin	Young	Civilian		Civil servant	2
26			z	Zoe	Zimmerman	Civilian		Politician	2

3. Data cleaning and transformation

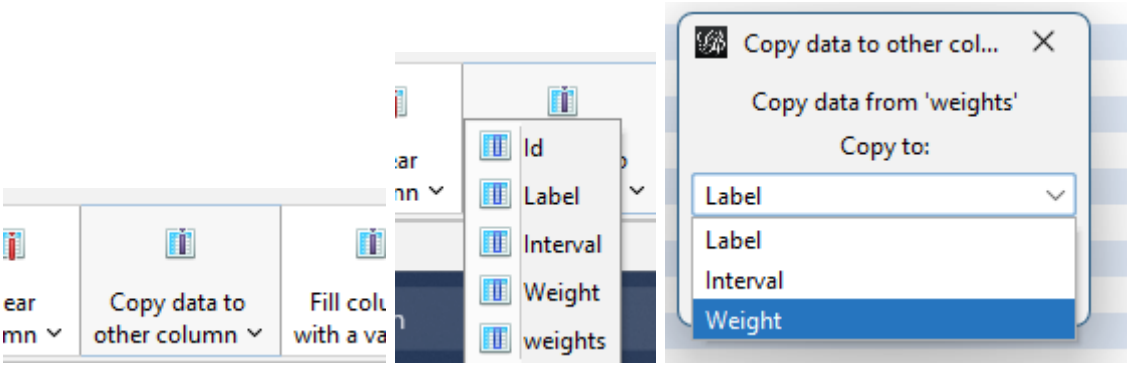
Looking at the list of edges, or at the **Overview** tab, what can you see?



We will carry out 3 bits of data cleaning and transformation:

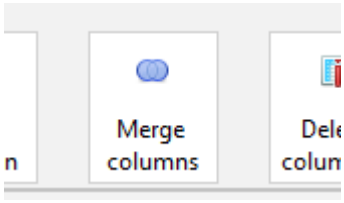
1. Correct edge `Weights`
2. Combine name columns for the nodes to a `full_name`
3. Add this column as a `Label`

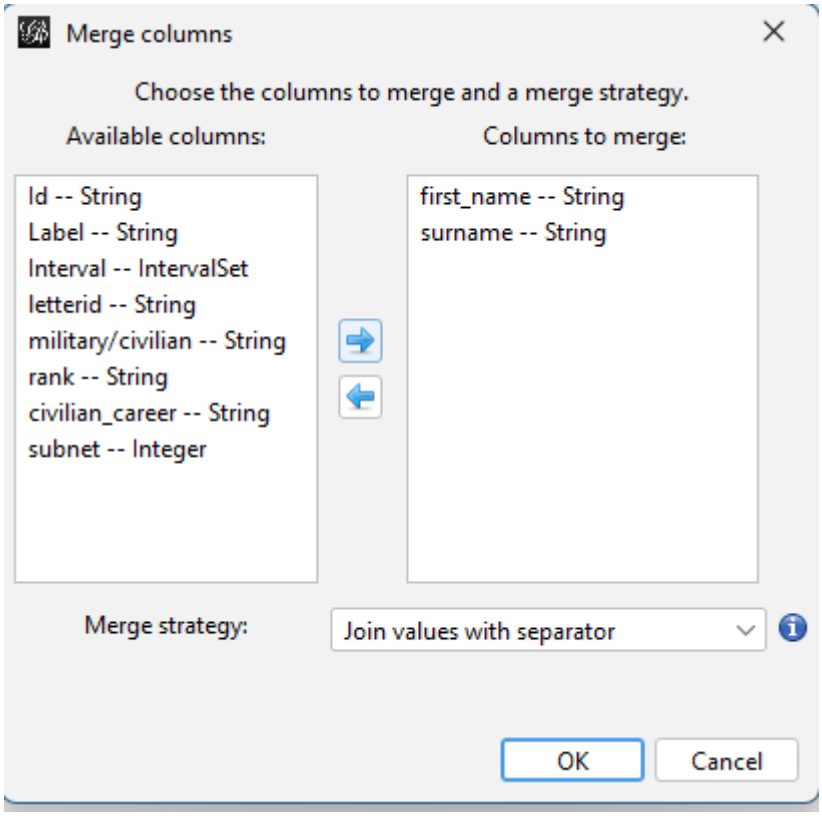
1. Currently the imported data had all possible combinations of nodes in a list, with edge weight of zero indicating no interaction. Due to naming of the column, this was not automatically recognized by Gephi on import. Therefore we will need to correct the edge weights, then remove edges that don't exist. First we use the option to copy data from one column to another

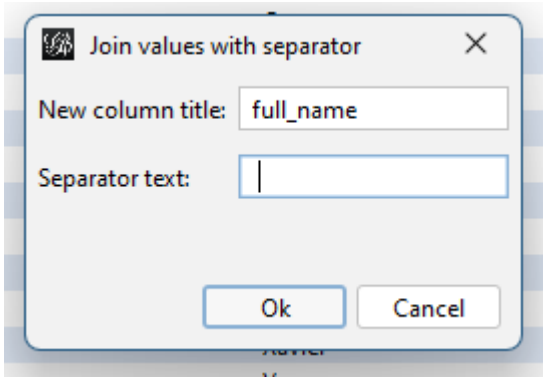


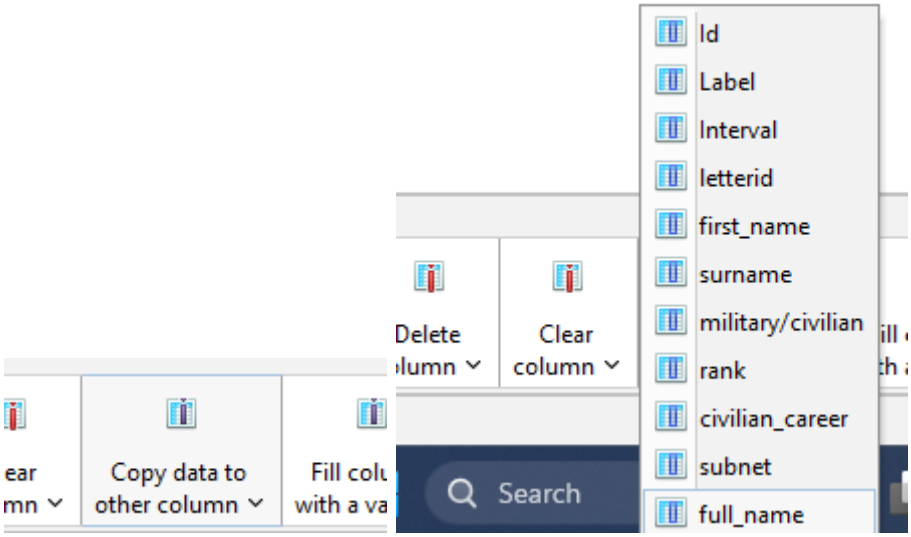
   Then we can sort columns by `weight` and delete all rows with `weight` = 0 or, we can filter in the **Overview** tab.

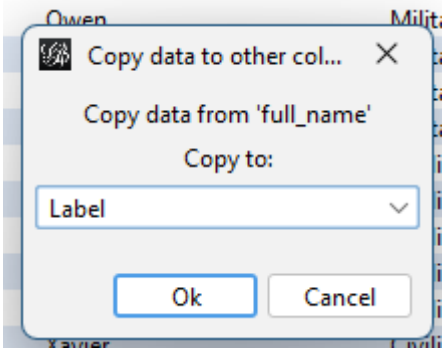
2. To construct our ideal `Label` for each node we will combine (merge) to columns into a new one. This time in the nodes section of the **Data Laboratory**





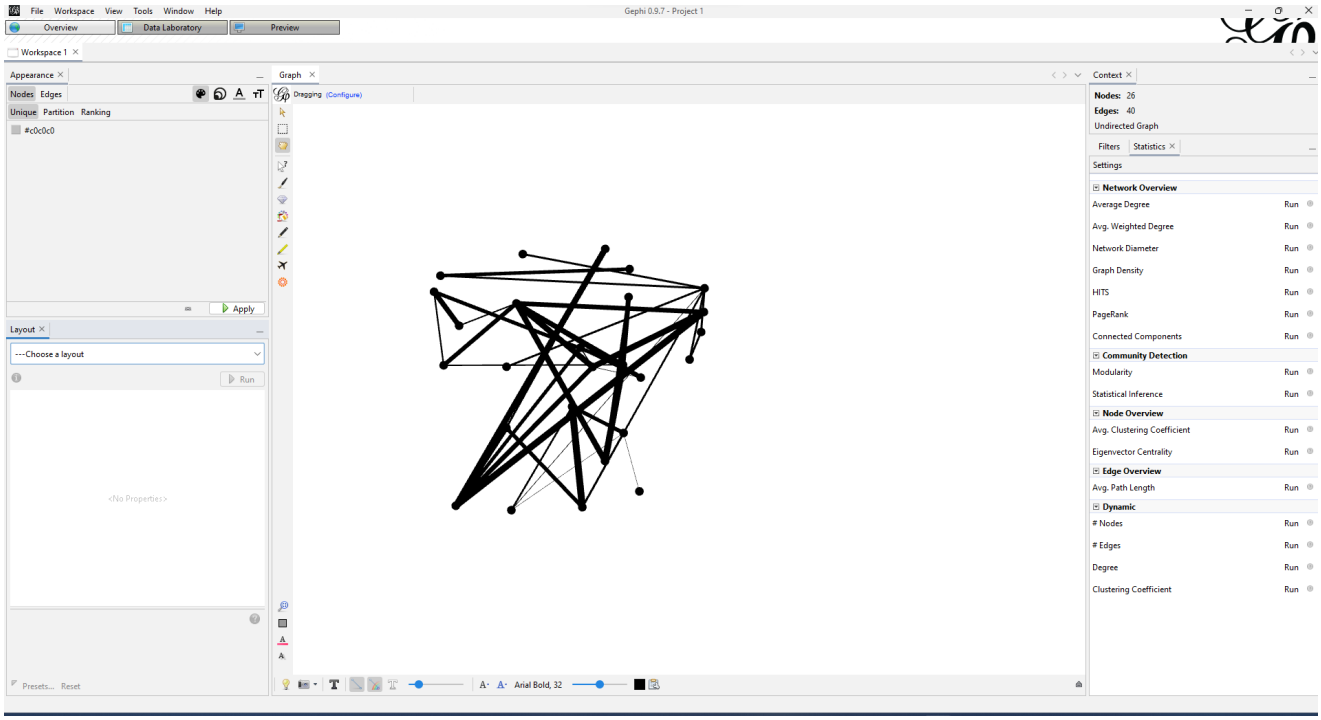


3. again 

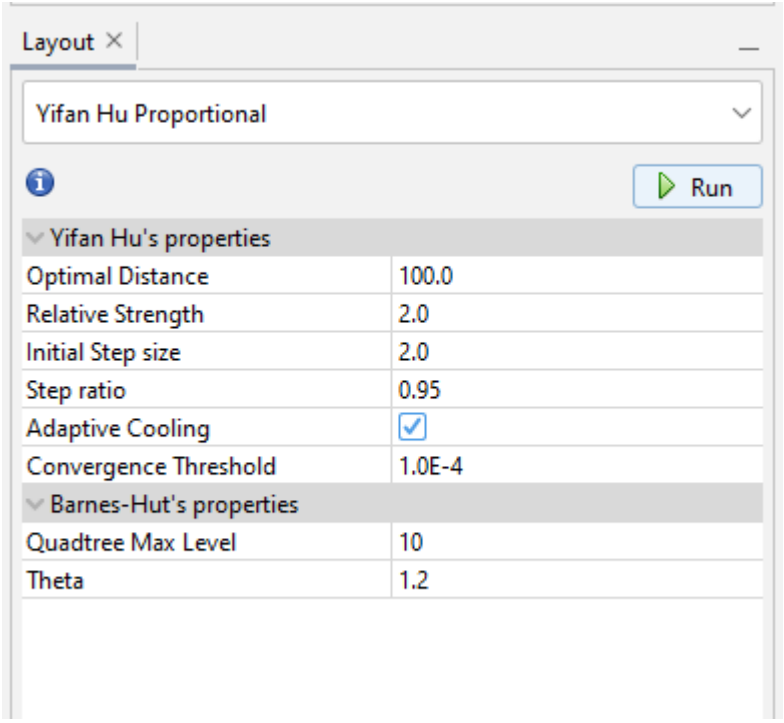


4. Visualization and layout of a network

Have a play around with the layout of the network.

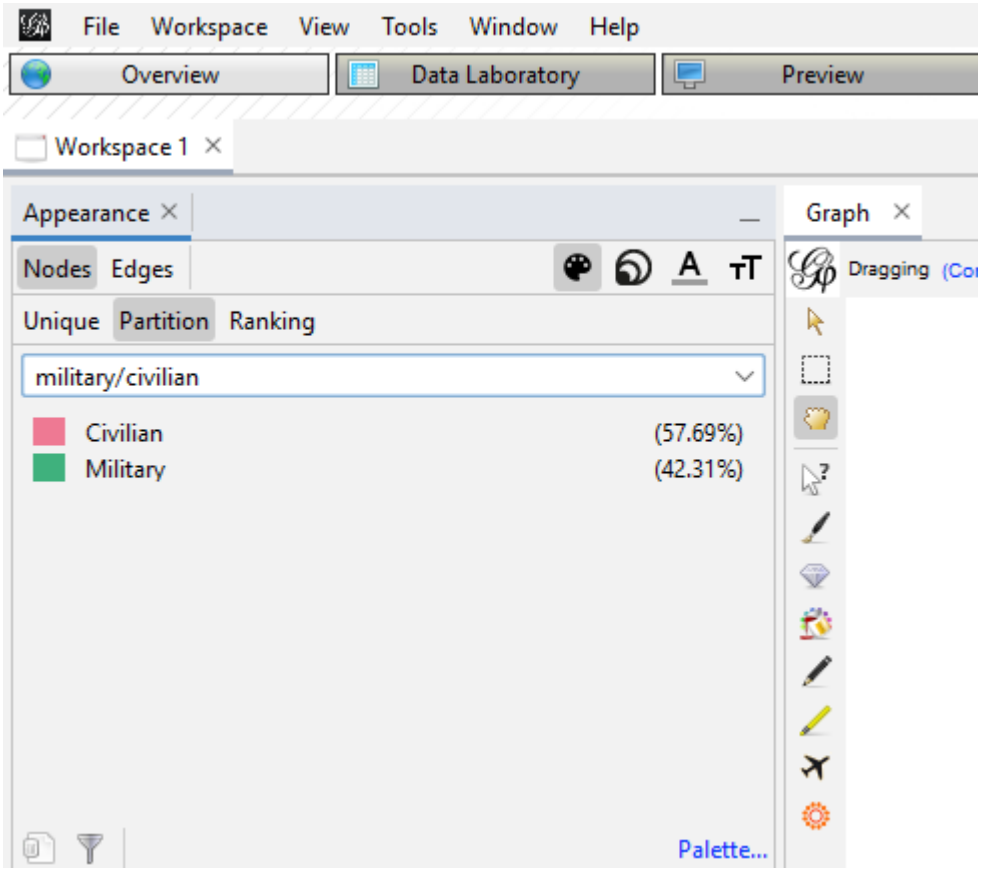


if you get stuck try ...

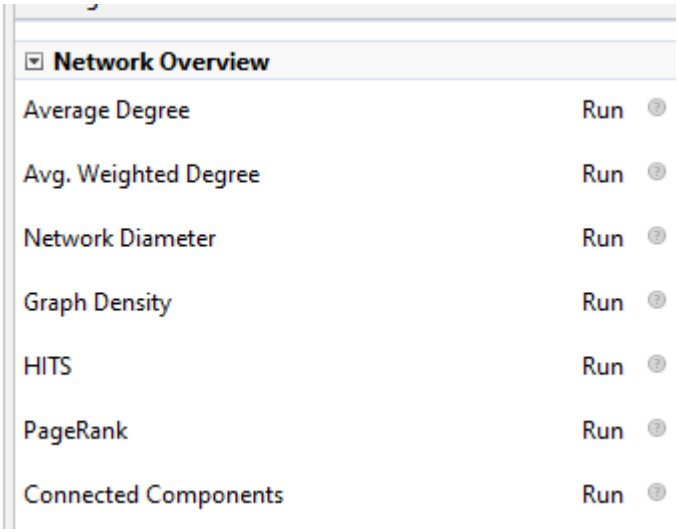


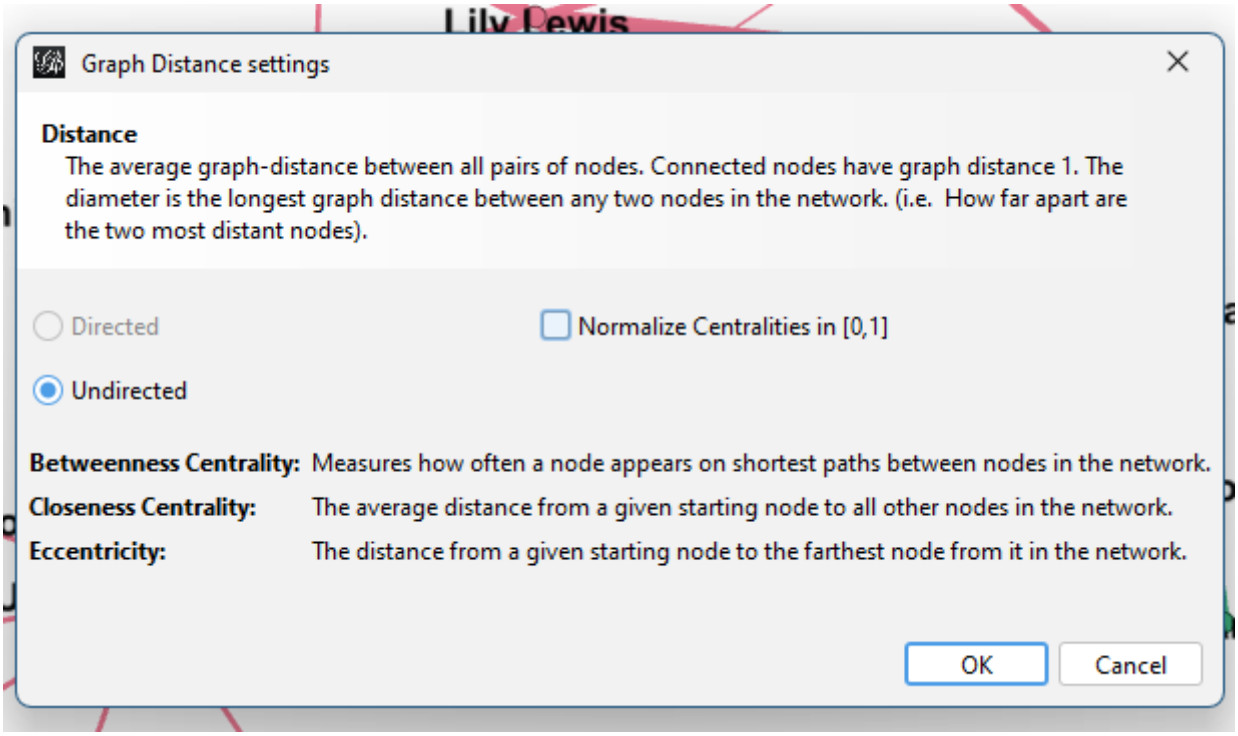
5. Statistical analysis and refining of visual details

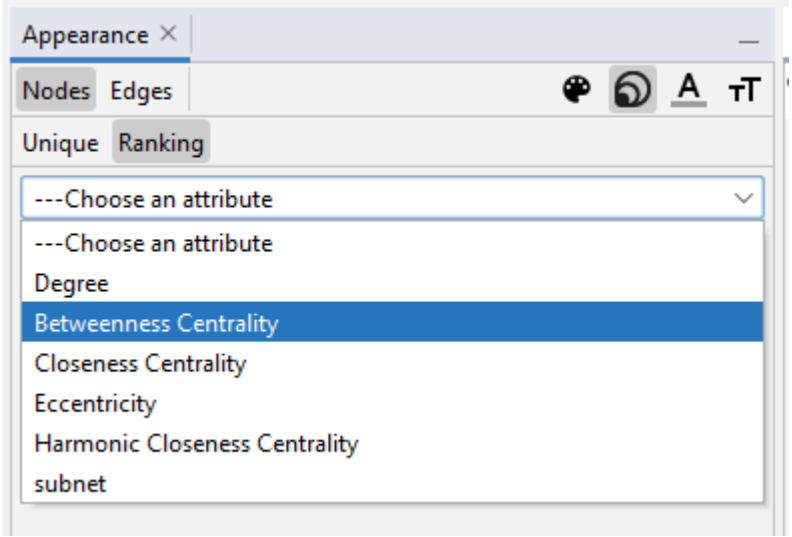
We can now look at the options provided in the Appearance box of the **Overview** tab

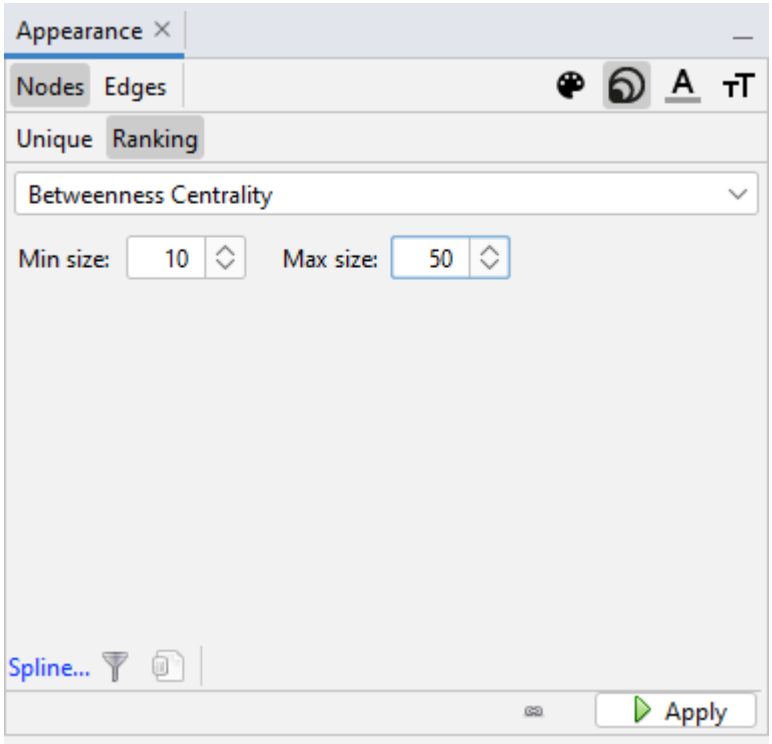


We could also take a look at the analysis of the network to add accurate numbers to what the basic visualization shows.



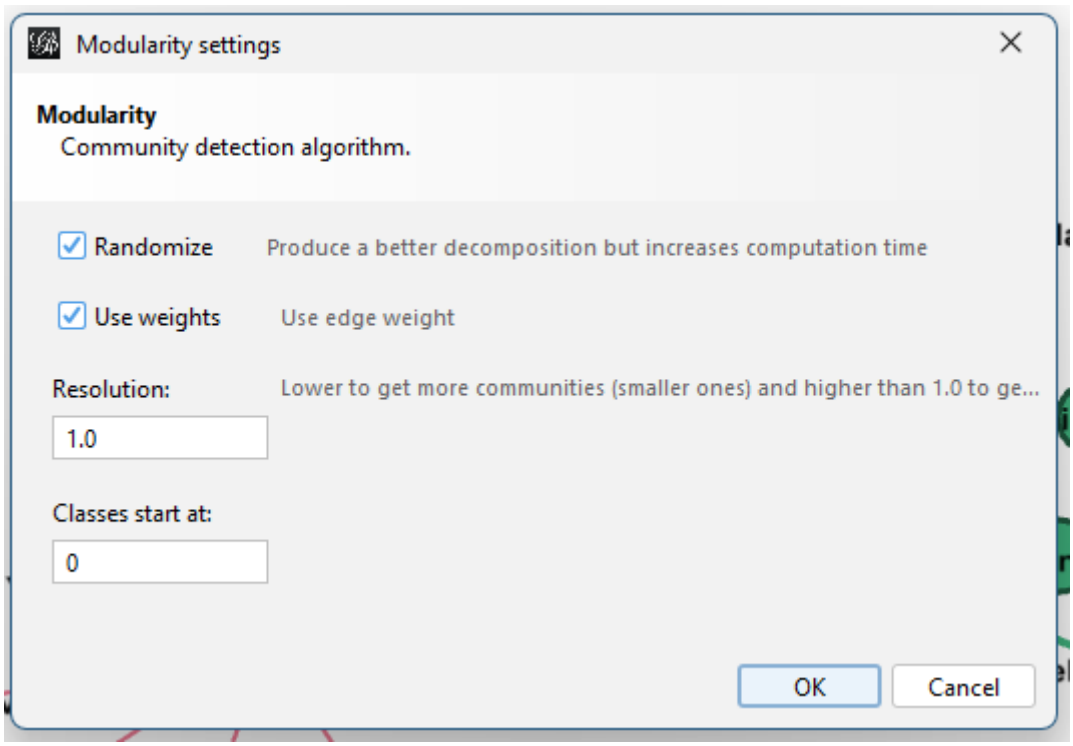






Set Min/Max to 10 and 50pt and hit *Apply*

We can also examine the clusters or sub-groups that might exist in our overall network



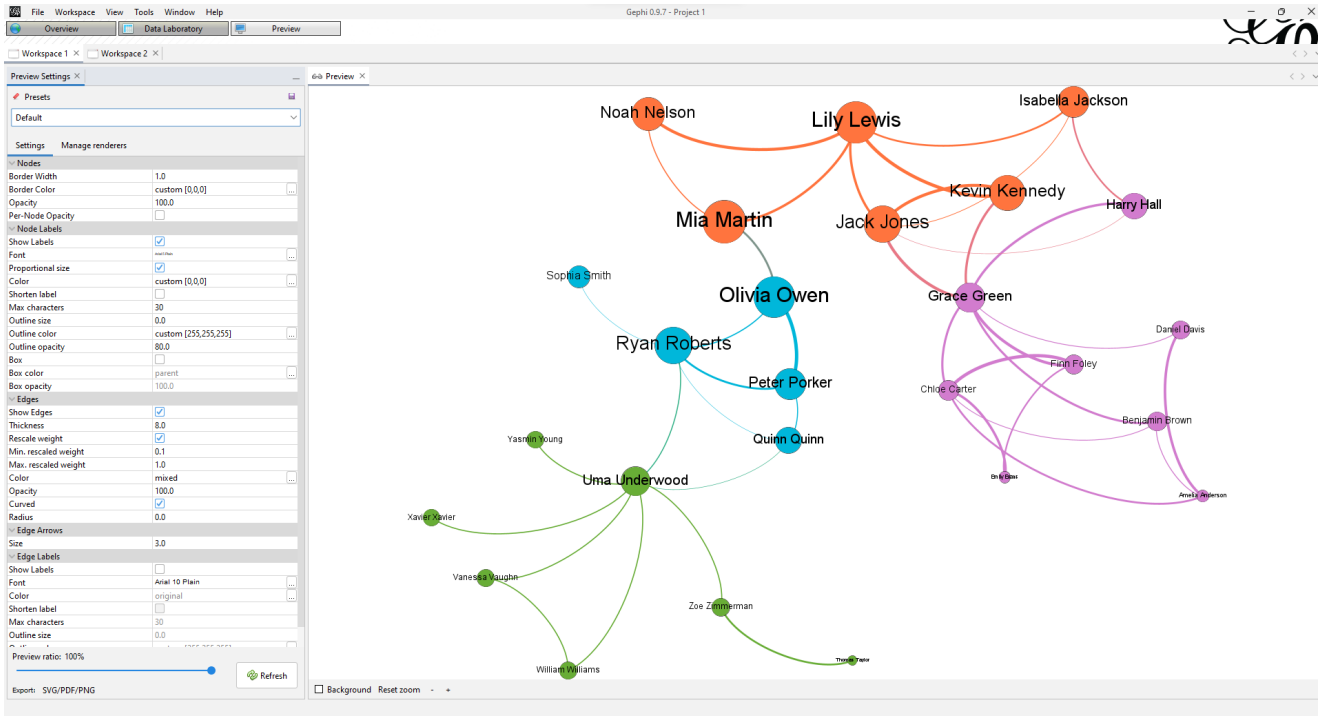
These will also now appear as options for colouring nodes in the **Appearance** box of the **Overview** Tab

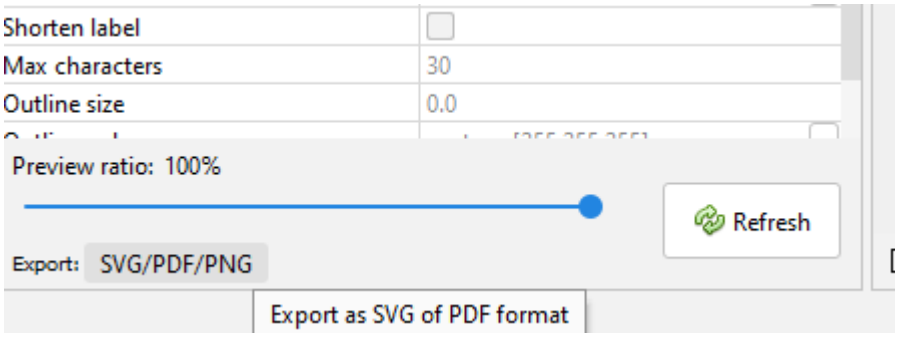
6. Export of data, images, and networks.

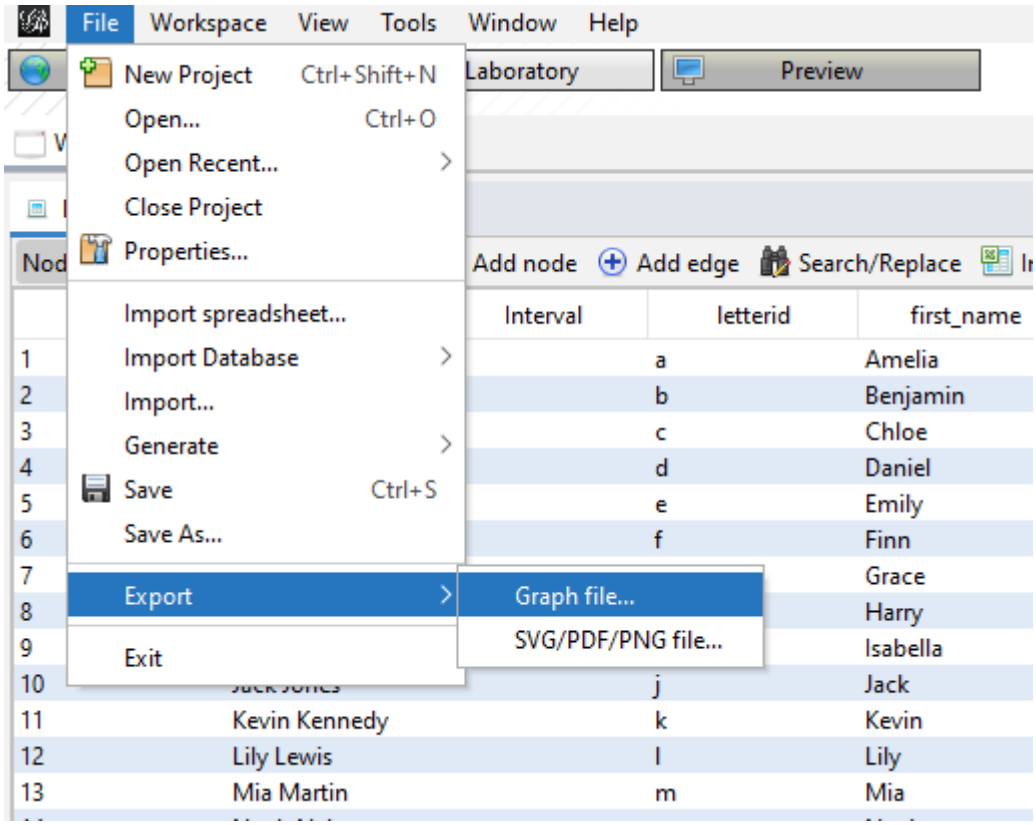
We have a number of options for export of data, images, and the full network file.

Preview tab has a wide range of options for styling you final network(s)

You may find it easier to move between this and the **Overview** tab







Plus a host of Plugins to investigate ...