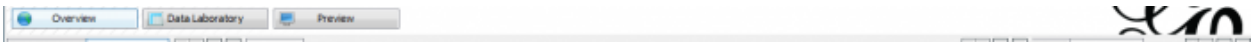
ALE #5
GEPHI HANDOUT

Gephi is a network analysis and visualization software. The user is free to rearrange the environment, move panels, show/hide windows, etc. The Graphical User Interface is set by default for three task families grouped as Overview, Data Laboratory and Preview.

Organization: *Modes are accessible under the main menu:*
- ***Overview****: graph manipulation mode.*
- ***Data Laboratory****: data visualization in tables.*
- ***Preview****: visual tuning before vectorial export.*



Workspaces: *A workspace is a whole environment for exploring one graph. If you want to analyze more graphs simultaneously, you need to create new workspaces. Note that each new data import creates a new workspace if you select the "New workspace" option in the import window.*

1. UPLOADING A NETWORK IN CSV FILE FORMAT

Gephi can import following standard graph file formats: GEXF, GDF, GML, GraphML, Pajek NET, GraphViz DOT, CSV, UCINET DL, Tulip TPL, Netdraw VNA, Spreadsheet

We will mainly work with CSV (Comma Separated Values) file format. CVS files can be created via Excel or Google Sheets by saving your file in CSV format. Gephi supports CSV files that simply represents relationships. Each line must contain at least two elements, separated by a separator (comma, semicolon, pipe or whitespace). By default graphs imported from CSV are directed graphs, but the user can select undirected in the import report dialog.

This manual shows what steps are necessary to import CSV files into Gephi, which will expect that each row of the file is a node or an edge. Note that the import can be done at any moment, the workspace does not need to be empty.

Here we'll learn how to use the import wizard to customize the data to our needs based on the options available.
1. Click on "**File > Import spreadsheet**…"
2. Import the **edges.csv**, (**e.g.,** "**les_miserables_links.csv**") file and the **nodes.csv (e.g.,** "**les_miserables_nodes.csv**") file.

If you want to create your own data, read below:

*You will need to create **two .csv files**: a node table and an edge table. Excel files automatically save by default as **.xlsx** format. In order to get the **.csv** format, save the file as **.csv** (UTF 8 format) when you click "Save as." Name the file with "ANameOfYourChoice_nodes.csv" to specify that this is the node table. In general, here is a bit about the difference between files for nodes and edges:*

- ***NODES:*** *The nodes file tells Gephi all the possible nodes in a network. A node is represented by a circle within the Gephi visualization whereas the edges file tells Gephi how all the nodes are related (or connected). The nodes file should at least have the columns "Id" and "Label." The Id could be a number (e.g., 1, 2, 3, 4, 5) or same as label. Labels can be names, places, or any meaningful representation of the node itself. Your nodes table might look like this:*

A	B	C
Id	Label	Gender
Myriel	Myriel	male
Napoleon	Napoleon	male
MlleBaptistine	MlleBaptistine	female
MmeMagloire	MmeMagloire	female

The node table can also include attributes. Attributes offer a way for you to distinguish between your nodes by categorizing your data by, for example, color, size, gender, or age.

- ***EDGES:*** *The edges table (the second .csv table) tells Gephi how the nodes are connected. It has the columns* ***Source, Target, and Type. Source*** *refers to a node that you've identified and labeled in your nodes.csv file.* ***Target*** *also refers to a node you've listed in your nodes.csv file.* ***Type*** *refers to how the two nodes are connected. If the source drives the relationship (for example, a sender of a letter versus a receiver), the relationship is "Directed." In this example, the sender of the letter is the source and the receiver of the letter is the target. If the relationship goes both ways - for example, the graph visualizes friendships, the graph will be* ***undirected.*** *Here is an example of what your file will look like:*

A	B	C	D
Source	Target	Type	Weight
Napoleon	Myriel	Undirected	1
MlleBaptistine	Myriel	Undirected	8
MmeMagloire	Myriel	Undirected	10
MmeMagloire	MlleBaptistine	Undirected	6

In the edges table, you can also add a column to define the weight for each relationship. ***Weight*** *gives you the option to show the importance of certain relationships by giving them a numerical weight.*

2. PLOTTING UNDIRECTED, DIRECTED, AND WEIGHTED NETWORKS

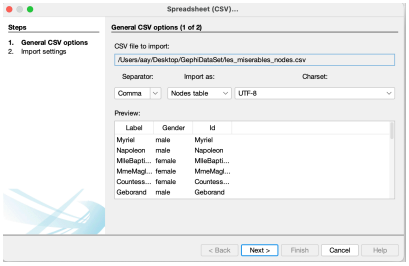
To plot different types of networks, we will mainly work on a CSV file.

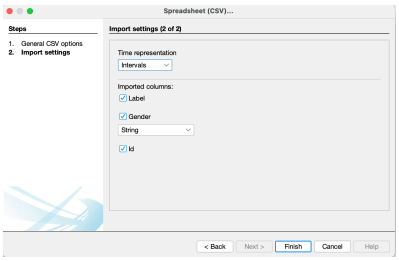
- ***Undirected Network (also Weighted):***

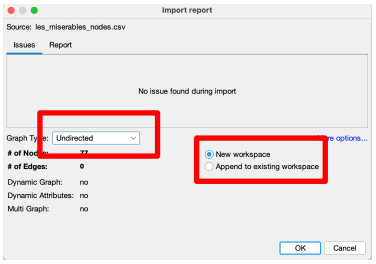
We will focus on analyzing the characters network of Les miserables here. The characters are provided in les_miserables_nodes.csv file. Two characters are connected if they appear in the same chapter. In the les_miserables_links.csv file, Source indicates the starting point of the edge, and Target indicates

*the ending point. You **MUST** use the Id of the nodes for these edges. Since the network is undirected, either node can be source/target, we specified '**Undirected**' for ALL edges.*

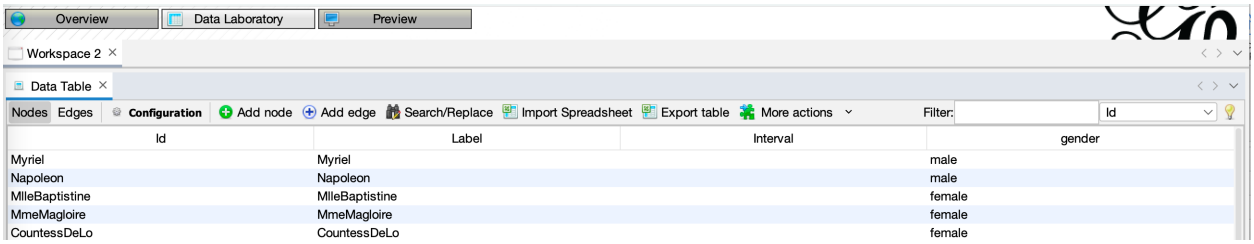
*Click Import Spreadsheet and choose **les_miserables_nodes.csv.** Click **Next,** then click **Finish**. Make sure to choose **New Workspace**, and choose the graph type to be **Undirected** since **les_miserables_nodes.csv** file is for an undirected network.*





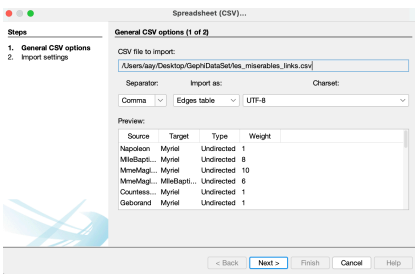


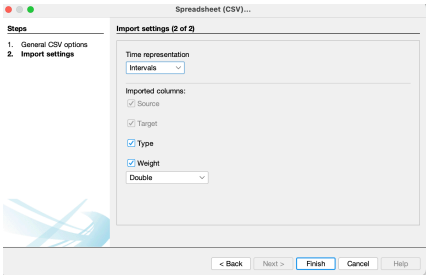
*The **Data Laboratory's Nodes** tab should look like this:*

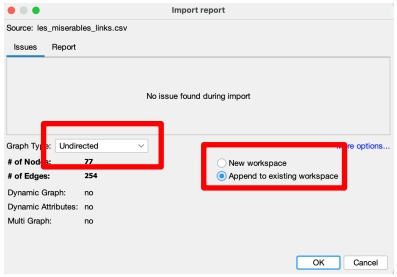


Id	Label	Interval	gender
Myriel	Myriel		male
Napoleon	Napoleon		male
MlleBaptistine	MlleBaptistine		female
MmeMagloire	MmeMagloire		female
CountessDeLo	CountessDeLo		female

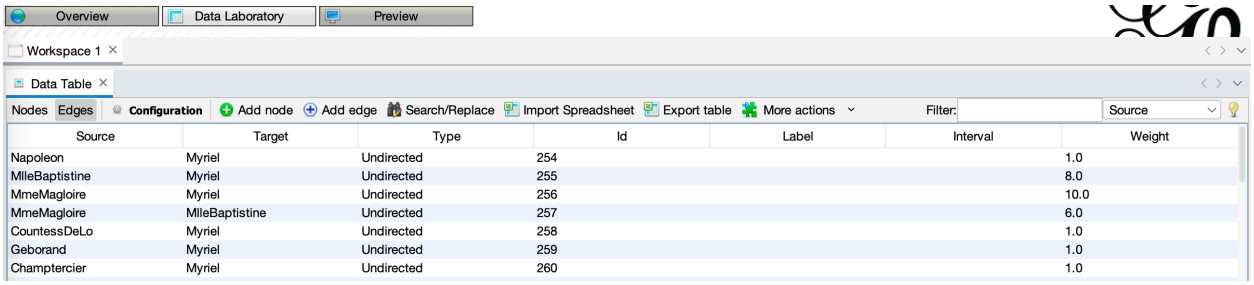
*Import the edges spreadsheet (les_miserables_links.csv). It's important that you choose '**Append to existing workspace'.***





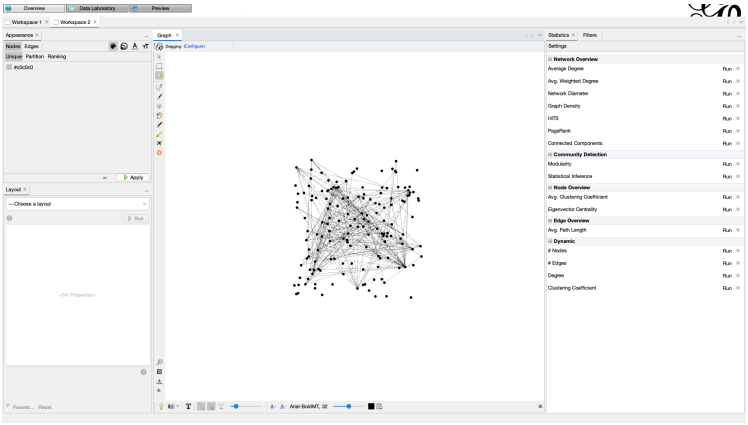


*The **Data Laboratory's Edges** tab should look like this:*

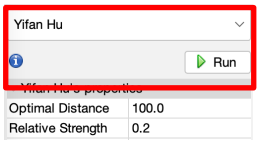


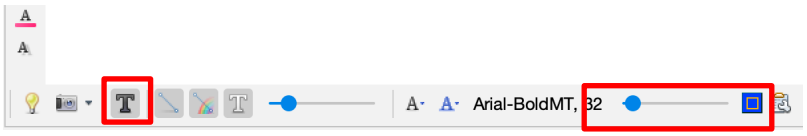
Source	Target	Type	Id	Label	Interval	Weight
Napoleon	Myriel	Undirected	254			1.0
MlleBaptistine	Myriel	Undirected	255			8.0
MmeMagloire	Myriel	Undirected	256			10.0
MmeMagloire	MlleBaptistine	Undirected	257			6.0
CountessDeLo	Myriel	Undirected	258			1.0
Geborand	Myriel	Undirected	259			1.0
Champtercier	Myriel	Undirected	260			1.0

*Click **Overview** to see the network. It looks something like this:*

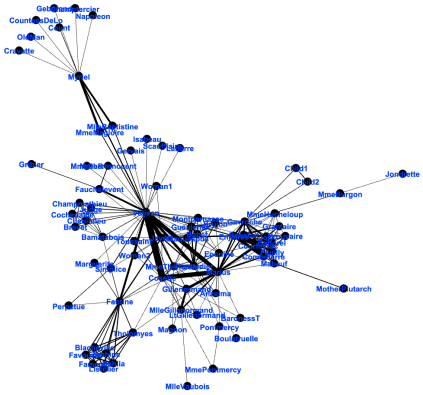


You can use a network layout like Yifan Hu (on the bottom left corner, make sure to RUN). You can also show the labels of the nodes, change the font size and font color for labels using the buttons at the bottom of the window.





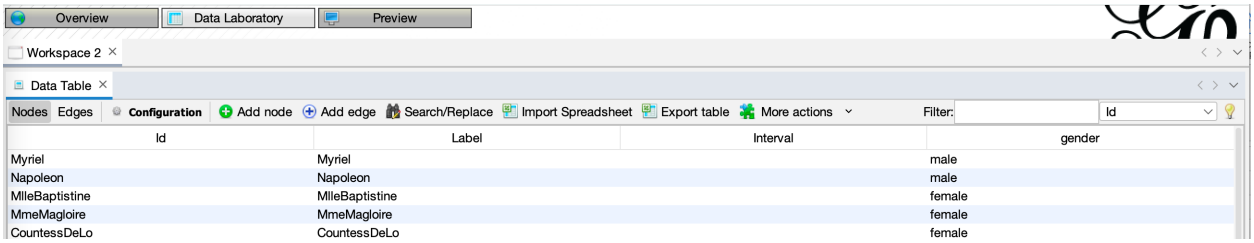
You can get a network plot similar to this.



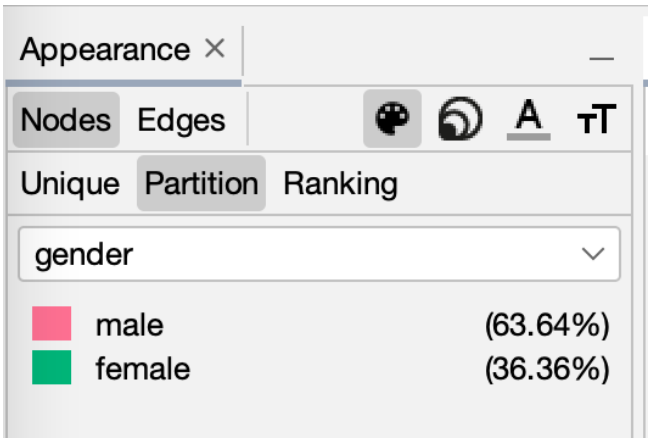
- ***Directed Networks:** The only thing we need to change in the data is to modify **Type** to **Directed**. Following similar steps to shown above, we can obtain a plot of a directed network.*

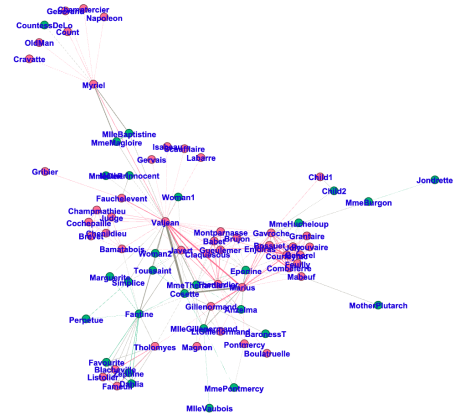
- ***Weighted Networks:** The edges.csv file is the only file where the directed network differs. We'll have a column called "Weight" in the csv file with numerical values. Following similar steps to shown above, we can obtain a plot of a weighted network. Please note that the files I used above is already weighted.*

❖ ***Coloring by Node Attributes:*** *A network consists in entities and their relations. This is what we just visualized. Yet, the properties of these entities remain invisible. For instance: the characters in the novel "Les Misérables" are male or female. Are males more likely to be connected to males, or females? Just looking at the network in Gephi, we can't tell.*

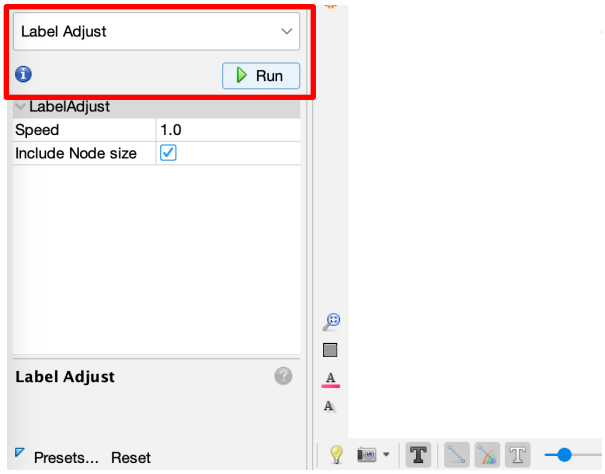


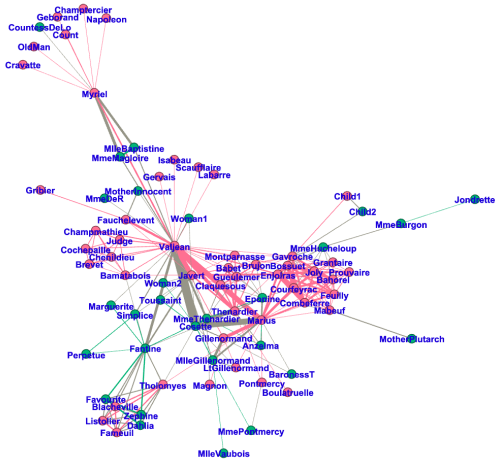
We see there is a Gender attribute for each character. We will color the nodes based on their gender. To do that, we select ***Partition****, then select* ***Gender*** *in the drop-down menu in the* ***Appearance*** *panel. Network after coloring characters according to their gender is given below.*

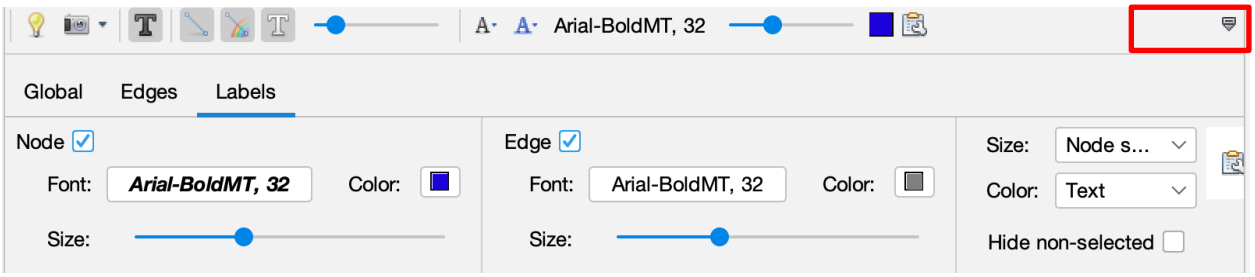




❖ **Prevent the labels from overlapping and change the edge weights:** *In the layout panel, choose "****Label Adjust****" or "****Noverlap****": these layouts will move the nodes just so that the Labels stop overlapping. Don't forget to click on "****Run****" to apply these layouts.*







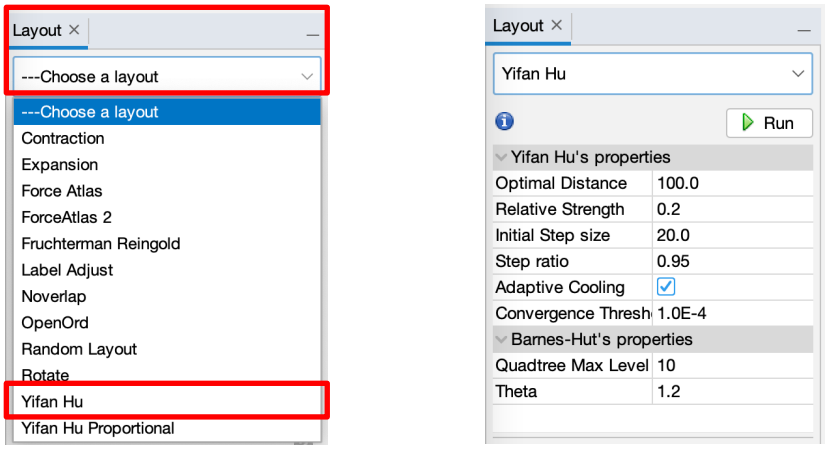
❖ ***Changing the size of the labels.*** *Open the bottom panel of Gephi by clicking on tiny arrow head (1). Then select "nodes" (2), then move the slider (3). You can adjust the thickness of links using the options here also. After all modifications, we have a nicer bipartite network:*

3. NETWORK LAYOUT OPTIONS

Gephi adjusts the nodes and edges in the network by the layout feature. It prioritizes different properties of the network. ***The general steps to modify the Network layout are:***

- ❖ *Choose a layout from the drop-down list (e.g., ForceAtlas 2)*
- ❖ *Adjust parameters for the layout algorithm*
- ❖ *Click the "Run" button*
- ❖ *Continue to refine the layout until you are happy with the results*

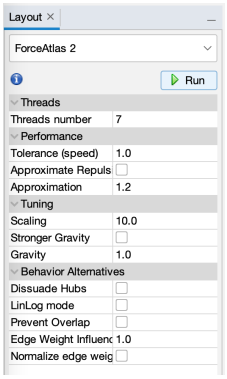
Layout Options:

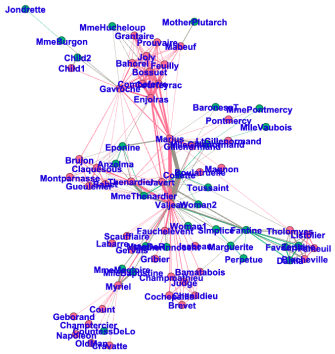


Graphs are usually laid out with "Force-based" algorithms. They follow a simple principle: linked nodes attract each other and non-linked nodes are pushed apart. Layout algorithms set the graph shape, it is the most essential operation. When you choose a lay out, you can see the layout properties below. You can leave default values for these options most of the time. Click run to launch the algorithm. You will see the positions of nodes changing in real time. PLEASE BE AWARE THAT

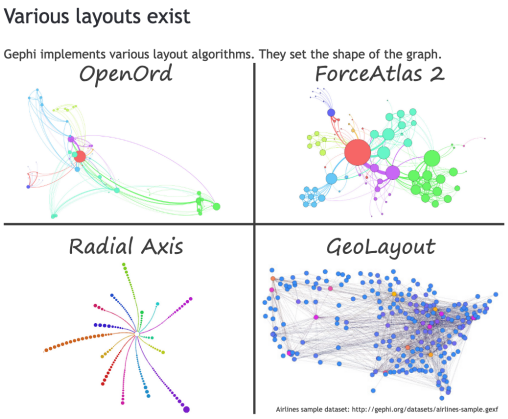
LAYOUT SELECTION IS AN ART, AND YOU SHOULD PLAY WITH DIFFERENT LAYOUTS UNTIL YOU FIND ONE THAT SHOWS A PLOT REFLECTING YOUR OBSERVATIONS ABOUT THE NETWORK.

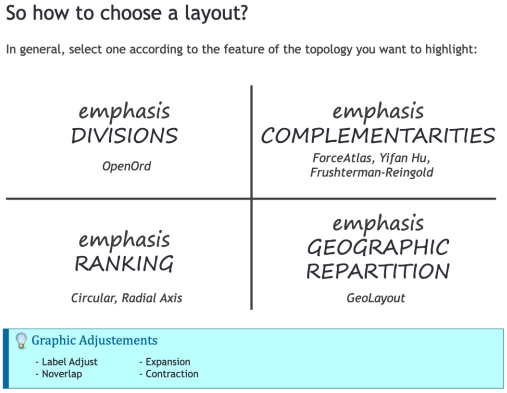
Force Atlas 2 *is a layout which bring together nodes which are connected, and spreads apart unconnected nodes. As an effect we can easily detect communities of nodes. Let's examine ForceAtlas 2 on this Les miserables network. You can change scale of the graph. You can choose prevent overlap option to avoid nodes being on top of each other. Make sure to RUN the layout.*





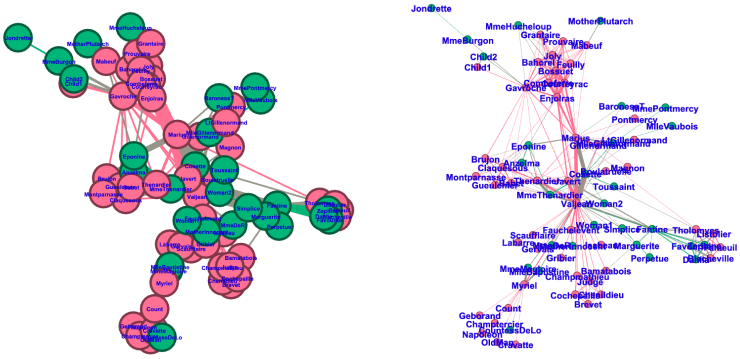
Some Important Layout Types:





ForceAtlas: *Home-brew layout of Gephi, it is made to spatialize Small-World / Scale-free networks. It is focused on quality (meaning "being useful to explore real data") to allow a rigorous interpretation of the graph (e.g. in Social Network Analysis) with the fewest biases possible, and a good readability even if it is slow, O(N^2). It can be used for 1 to 10,000 nodes.*

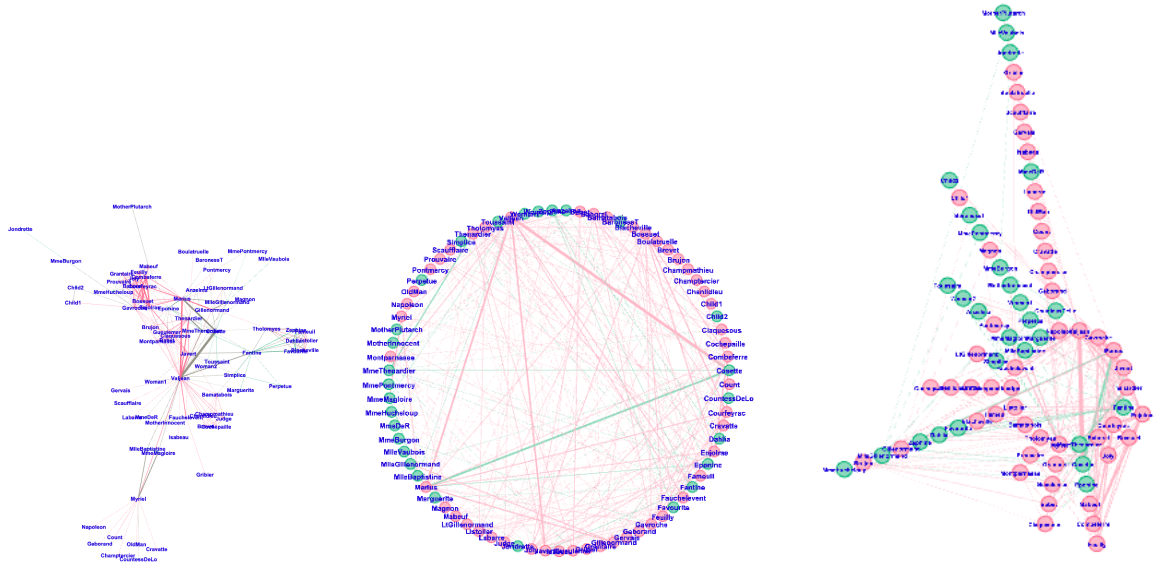
ForceAtlas 2: *Improved version of the Force Atlas to handle large networks while keeping a very good quality. This is O(N*log(N)) algorithm. It can be used for 1 to 1,000,000 nodes. Activate "Approximate Repulsion" on large graphs only. You can play with the "Tolerance" option. Linear attraction & logarithmic repulsion (lin-lin by default), makes clusters tighter. Increase Scaling to make the graph sparser. Change edge weight from 0 (no influence) to 1 (normal).*



Yifan Hu Multilevel Layout: *It is a very fast algorithm with a good quality on large graphs. Its complexity is O(N*log(N)). It can be used for 100 to 100,000 nodes. It does not use edge weight. It combines a force-directed model with a graph coarsening technique (multilevel algorithm) to reduce the complexity. It stops automatically. Step ratio (e.g., 0.99) is used to update the step size. Increase it for a better quality (vs speed). Optimal distance (e.g., 300) is the natural length of the springs. Increase it to place nodes farther apart. Smaller theta (e.g., 1) values mean more accuracy*

Circular Layout: *It draws nodes in a circle ordered by ID, a metric (degree, betweenness centrality...) or by an attribute. Use it to show a distribution of nodes with their links. Its complexity is O(N), and can be used for 1 to 1,000,000 nodes.*

Radical Axis Layout: *It groups nodes and draws the groups in axes (or spars) radiating outwards from a central circle. Groups are generated using a metric (degree, betweenness centrality...) or an attribute. Use it to study homophily by showing distributions of nodes inside groups with their links. Its complexity is O(N), and can be used for 1 to 1,000,000 nodes.*

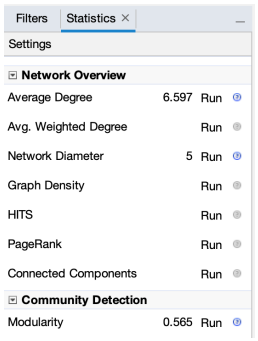


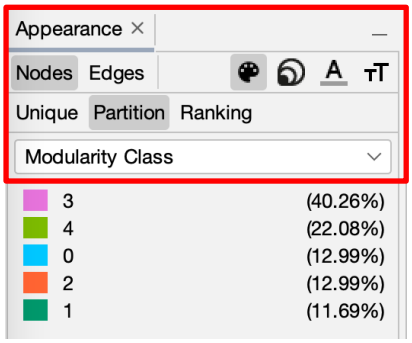
Community Structure: *We now want to show the community structure in this network: does it divide naturally into groups of nodes with dense connections within groups and sparser connections between groups?*

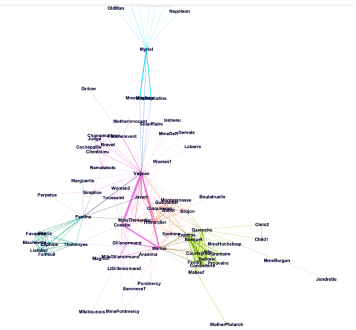
*In the **Statistics panel** (on the right), click on the **"Modularity"**. The community detection algorithm created a "Modularity Class" value for each node. The partition module can use this new data to colorize communities.*

Locate the Partition module on the left panel. Click on the "Refresh" button to populate the partition list. Select "Modularity Class" in the partition list.

You can see multiple communities were found, could be different for you. A random color has been set for each community identifier. Click on Apply to colorize nodes.



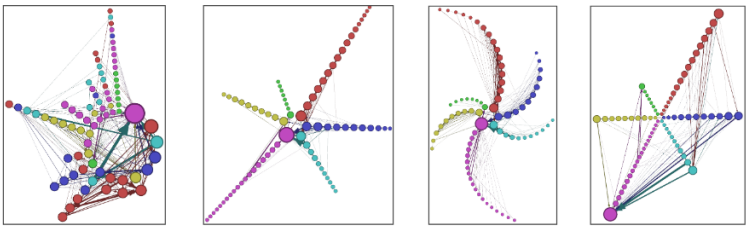




Run Radial Axis Layout

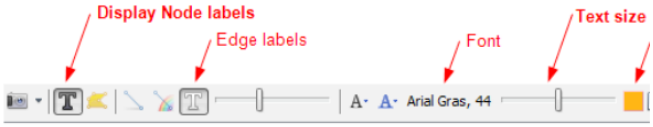
Run the layout by applying the following settings step by step:

- Group nodes by = "Degree" — Run — Homophily by degree?
- Group nodes by = "Modularity Class"
Order nodes by = "Degree" — Run — Distribution of nodes by degree inside each community.
- Draw spar/axis as spiral = checked — Run — Better show links inside communities
- Draw spar/axis as spiral = unchecked
Ascending order = checked — Run — Better show links between communities



Label Adjust Layout: *It works on text size to repulse nodes and therefore makes every label readable. It only runs on the visible nodes in the Visualization panel.*

- Locate the Visualization settings.
- Click on T to activate text display.
- Increase the text size to the maximum.

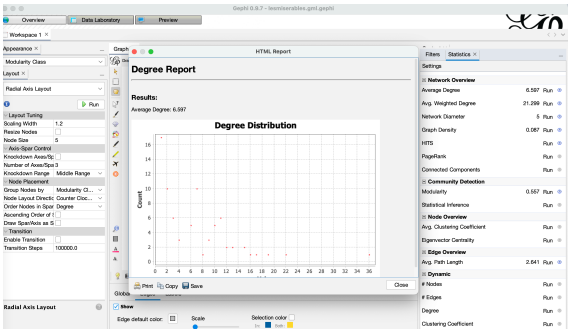


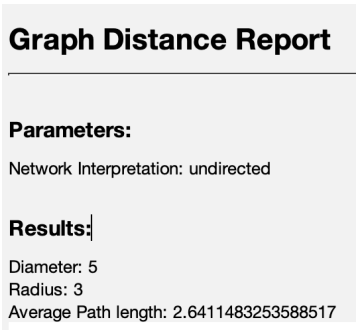
4. NETWORK METRICS

The general steps for obtaining network metrics:

- ❖ *Navigate to the right **'Statistics'** Panel*
- ❖ *Select the desired network metric*
- ❖ *Click **Run** on the desired measure*
- ❖ *View the Report and save*

For instance, we can calculate the degree centrality of the network. Click Run next to the Average Degree tab. This will open a report window with parameters. Similarly, we can calculate the average path length.





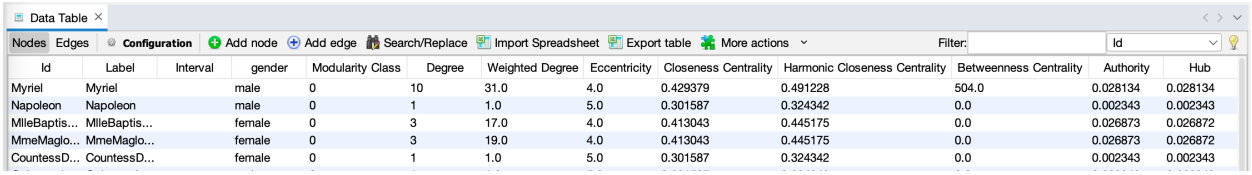
5. COLORING AND SIZING NODES BY ATTRIBUTES OR NETWORK METRICS

A. Node Sizes:

*It would be interesting to resize the nodes according to their centrality: the more central a node, the bigger. This would allow for a very quick visual appreciation of which nodes are the most central. This method can be applied to resize the nodes using different metrics. The general steps to resize **nodes according to a numerical variable are:***

- ❖ *Click on "**Nodes" and "Ranking" tabs.***
- ❖ *Select a variable (e.g., Degree) from the drop down*
- ❖ *Choose a minimum and maximum size as a range for the size of the nodes*
- ❖ *Click the "Apply" button*

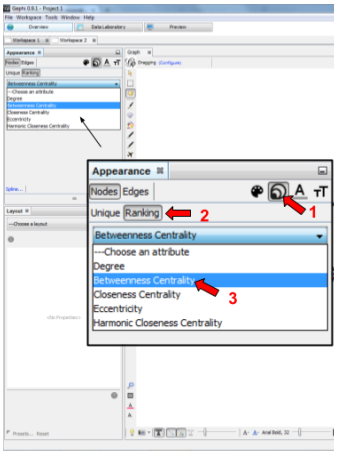
First, let's switch to the data laboratory to see how Gephi stored the metrics of each node:

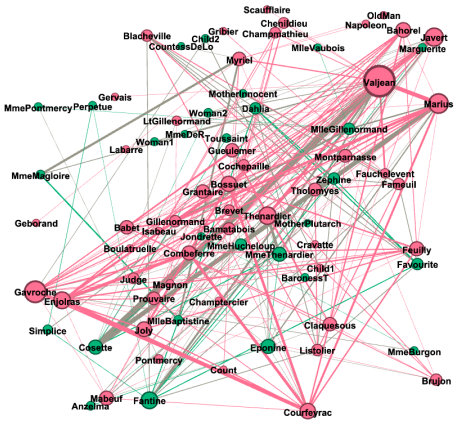


Id	Label	Interval	gender	Modularity Class	Degree	Weighted Degree	Eccentricity	Closeness Centrality	Harmonic Closeness Centrality	Betweenness Centrality	Authority	Hub
Myriel	Myriel		male	0	10	31.0	4.0	0.429379	0.491228	504.0	0.028134	0.028134
Napoleon	Napoleon		male	0	1	1.0	5.0	0.301587	0.324342	0.0	0.002343	0.002343
MlleBaptis...	MlleBaptis...		female	0	3	17.0	4.0	0.413043	0.445175	0.0	0.026873	0.026872
MmeMaglo...	MmeMaglo...		female	0	3	19.0	4.0	0.413043	0.445175	0.0	0.026873	0.026872
CountessD...	CountessD...		female	0	1	1.0	5.0	0.301587	0.324342	0.0	0.002343	0.002343

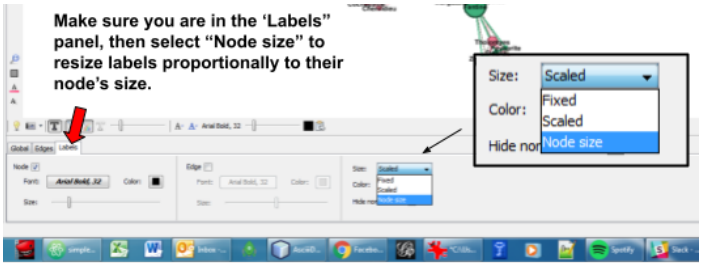
Different centrality measures visible in the data laboratory

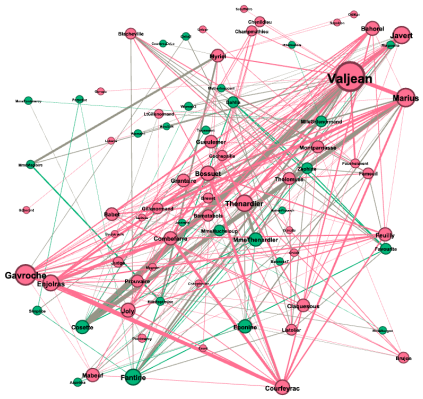
*To resize the nodes according to the value of their degree centrality, we navigate to the **Nodes** tab and chose the **sizing** icon (1). We then use the **Appearance** panel and **Ranking** tab (2). The plot shows betweenness centrality, but please choose Degree Centrality.*





Resizing labels to reflect their node's size





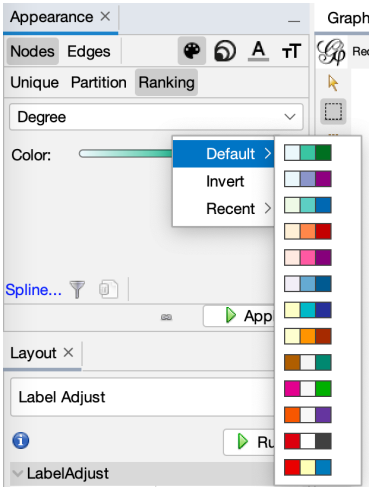
B. Node Color:

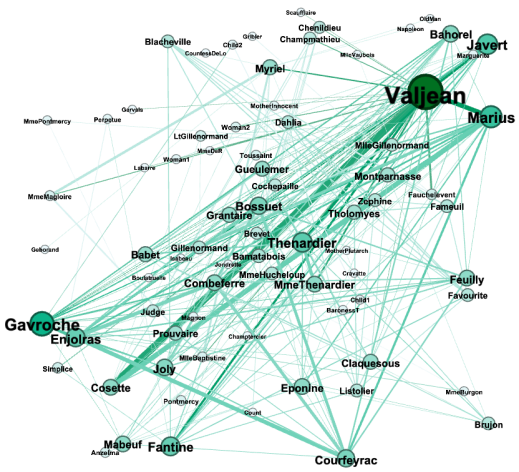
*This method can be applied to recolor the nodes using different metrics. The general steps to recolor **nodes according to a numerical variable are:***

- *Select a "partition" (categorical) node variable from your data. In our example we used Degree.*
- *Click on "**Nodes**"and "**Ranking**" tabs*

- *Choose the metric from the "**Choose an attribute**" drop down menu*
- *Click "**Apply**"*

*The only difference is navigating to the **Color** icon next to the Nodes tab. Lets color the nodes based on its Degree metric.*



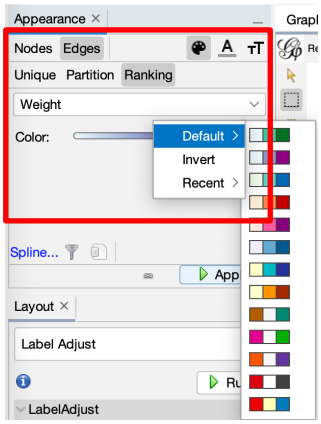


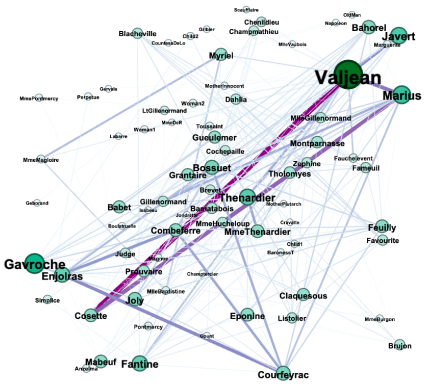
6. COLORING AND SIZING LINKS BASED ON NETWORK INFORMATION

We can also recolor and resize links based on numeric values (or metrics). In this example, the only attribute we can play with is the weight of each edge. Other examples, of course, can have more attributes for you to visualize. The methodology, however, remains the same.

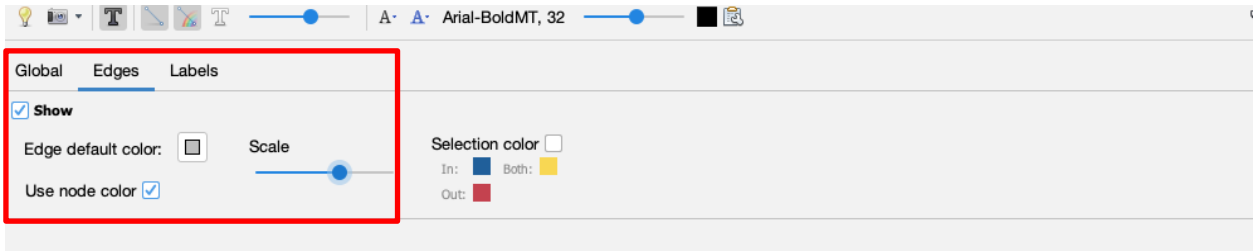
- *Navigate to the **Edges** tab, and choose either the **Color** icon.*
- *Click on "**Ranking**"*
- *Select the attribute from the drop-down menu*
- *Configure the minimum and maximum color range*
- *Click "**Apply**"*

For instance, to color the edges based on Weight:





To adjust the thickness of the edge, you can use the tool bar in the bottom panel.

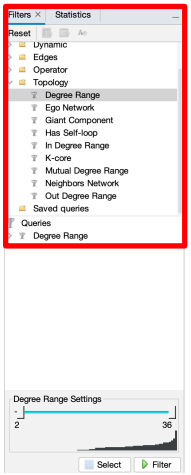


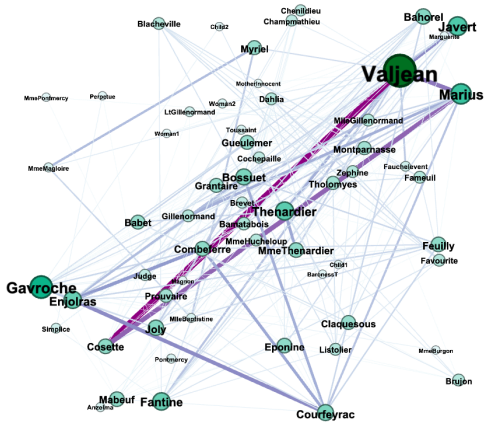
Slide the scale to make the edge thicker.

7. FILTERING NETWORKS TO SELECT NODES AND LINKS

You can create filters that can hide nodes and edges on the network. We will create a filter to remove nodes with only one link, as these are insignificant characters.
The general steps of the filtering operation are:

- *Click the "**Filters**" tab on the right*
- *Expand the folder with respective filtering criteria (Eg. Topology)*
- *Double-click the criteria desired (E.g., Degree Range)*
- *Drag the criteria down to the "**Queries**" below.*
- *Click the "**Filter**" button*





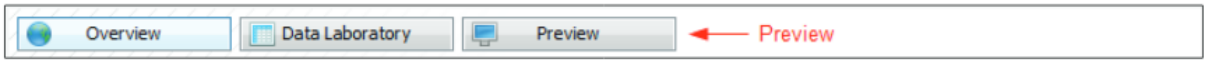
The network after filtering all nodes of degree 1 are hidden.

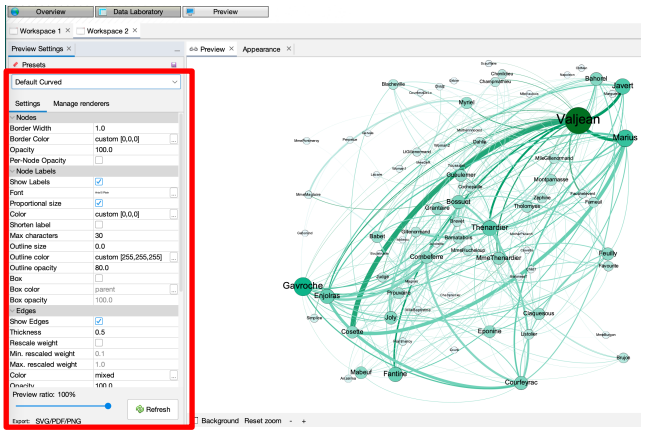
8. SAVING ANALYSIS AND IMAGES:

*If you'd like to open your file again in **Gephi**, you can save it as a Gephi file.*
*If you want to export only the visualization to a **SVG** or **PDF** file, go to the **Preview** to:*

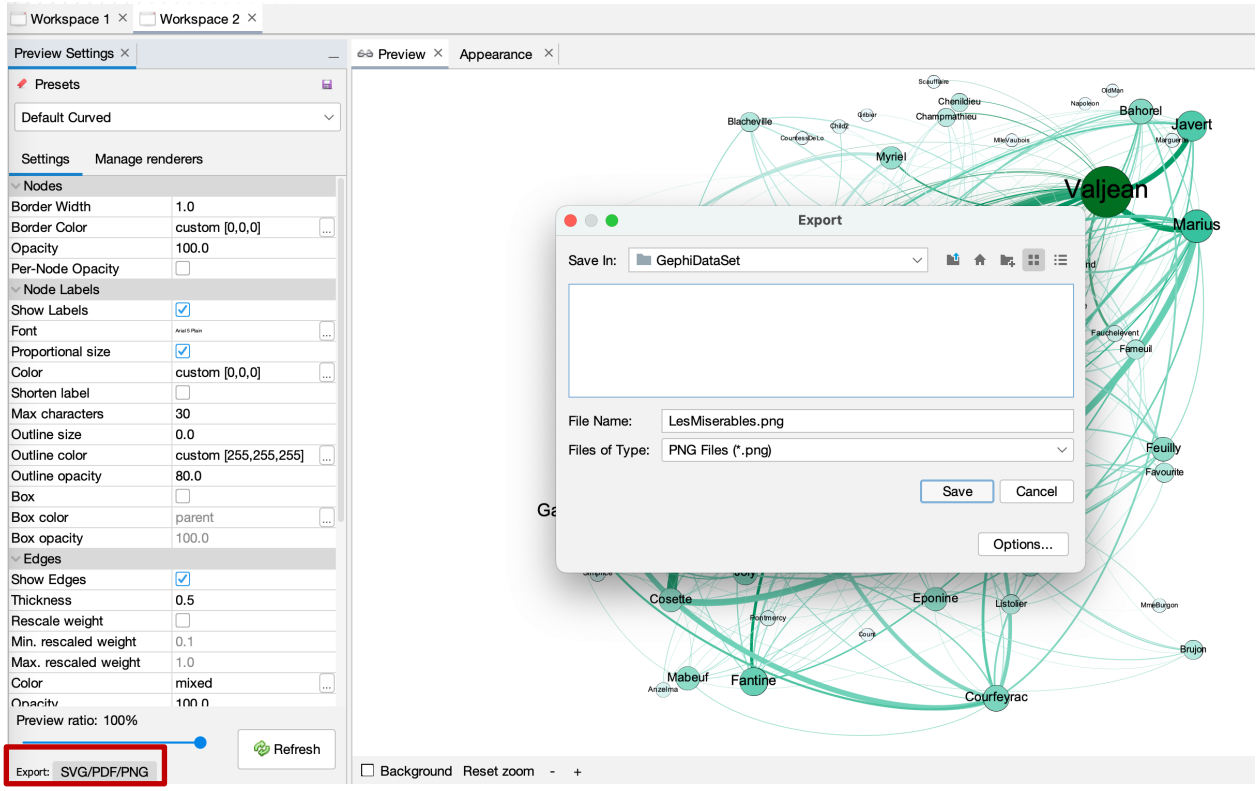
- *See exactly how the graph will look like*
- *Put the last touches*

*Let's select **Preview** tab in the banner. Click **Refresh** to see the preview after making your changes. For example, choose Show Labels. Change edge thickness to 0.5, and node font size to 5.*





To export, just click on the button and select the file format you prefer



Sources:

https://gephi.org/tutorials/gephi-tutorial-quick_start.pdf
https://gephi.org/tutorials/gephi-tutorial-visualization.pdf