

Big Data Analytics

Case Study: Social Network Analysis

Joyce Jiyoung Whang

SKKU

Introduction

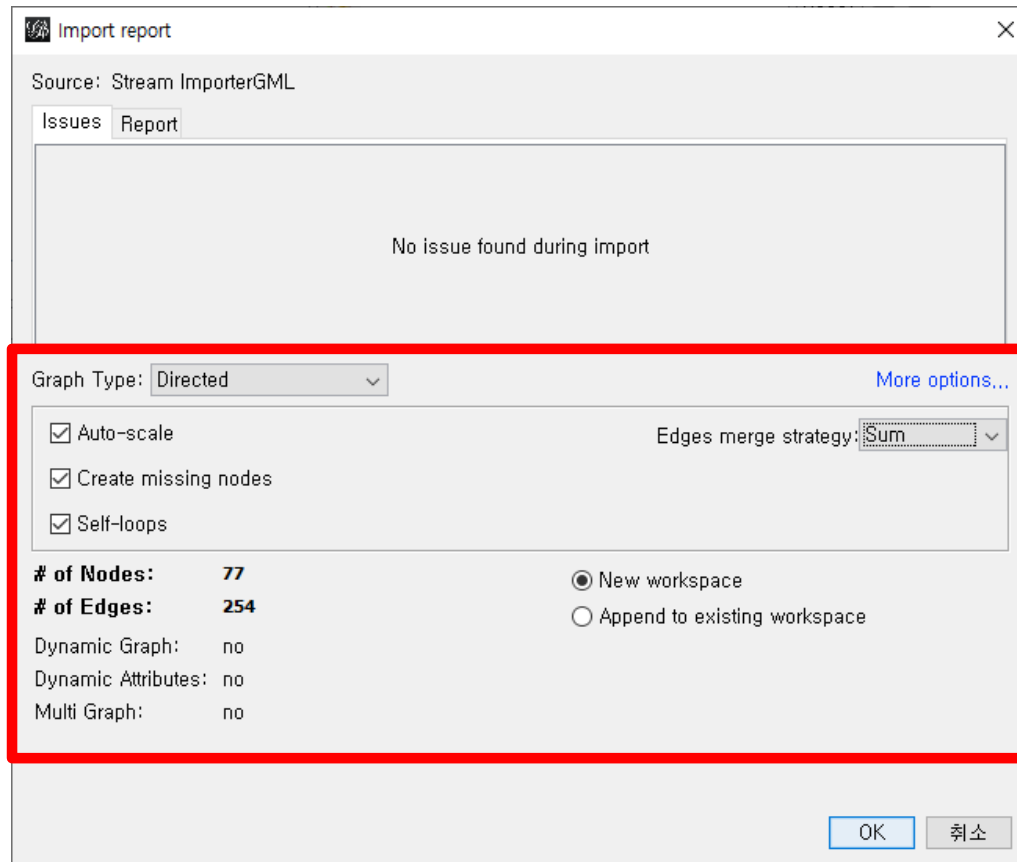
- **Gephi** is an interactive visualization and exploration platform for all kinds of networks.
- OS: Windows, Linux and Mac OS X
- Open-source and free
 - Download: <https://gephi.org/>

Introduction

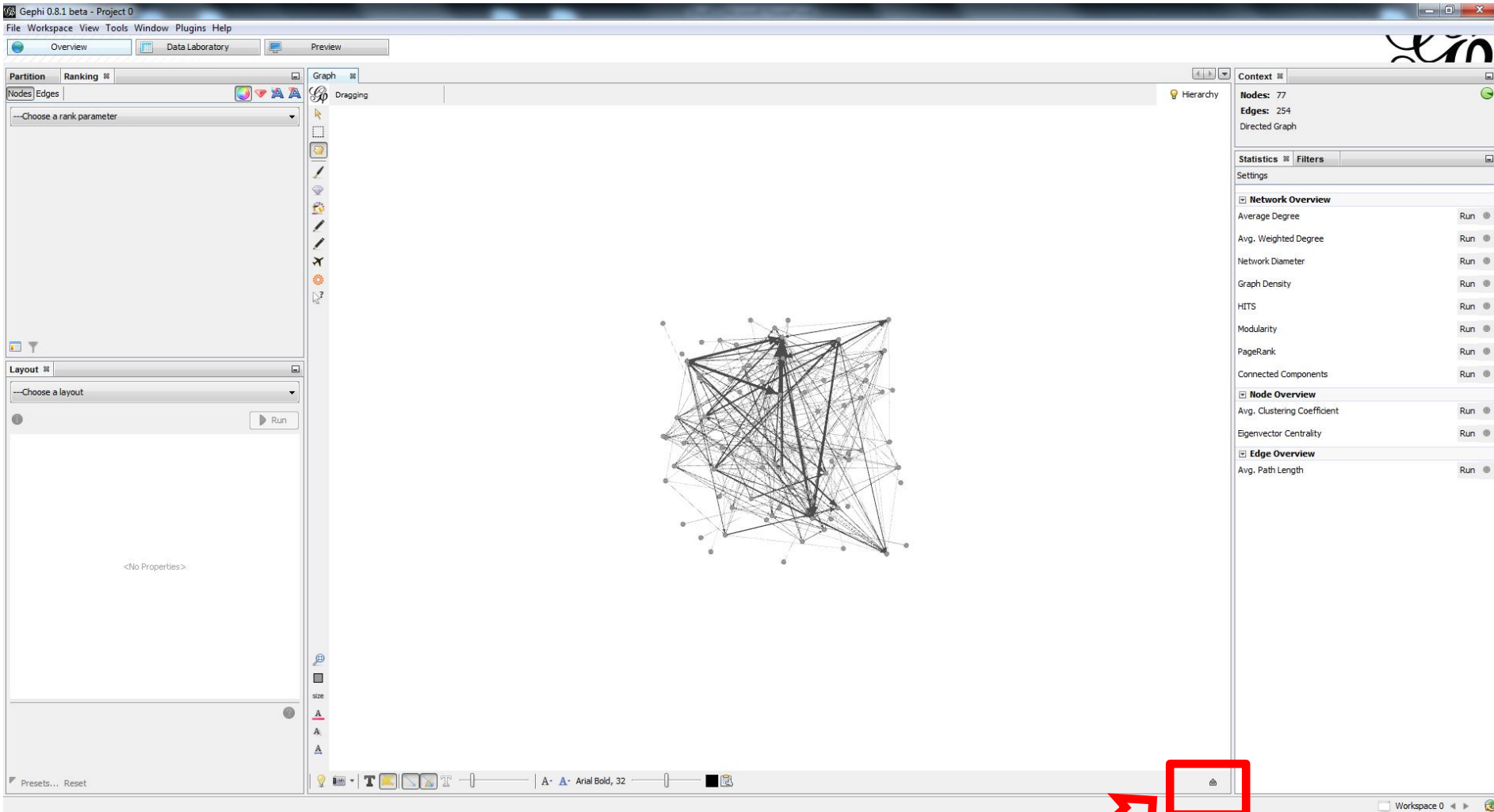
- Supported Data Formats:
 - [GEXF](#), [GDF](#), [GML](#), [GraphML](#), [Pajek NET](#), [GraphViz DOT](#), [CSV](#), [UCINET DL](#), [Tulip TPL](#), [Netdraw VNA](#), [Spreadsheet](#)
- Sample datasets are available at:
<https://github.com/gephi/gephi/wiki/Datasets>
- Gephi Wiki
<https://github.com/gephi/gephi/wiki>

Import Graph

- File > Open > Select “lesmiserables.gml”

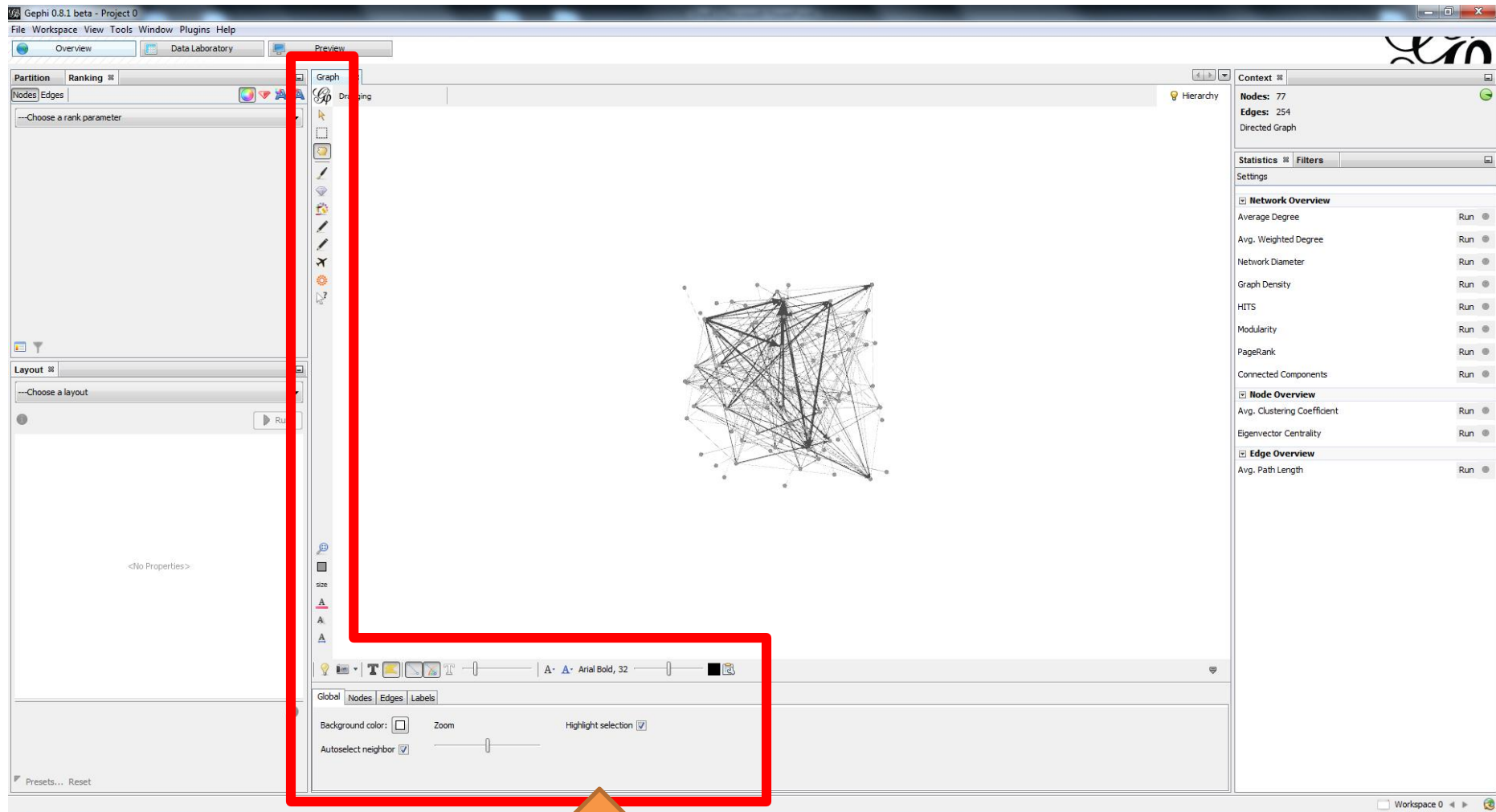


Visualization



Click this icon to see more visualization options.

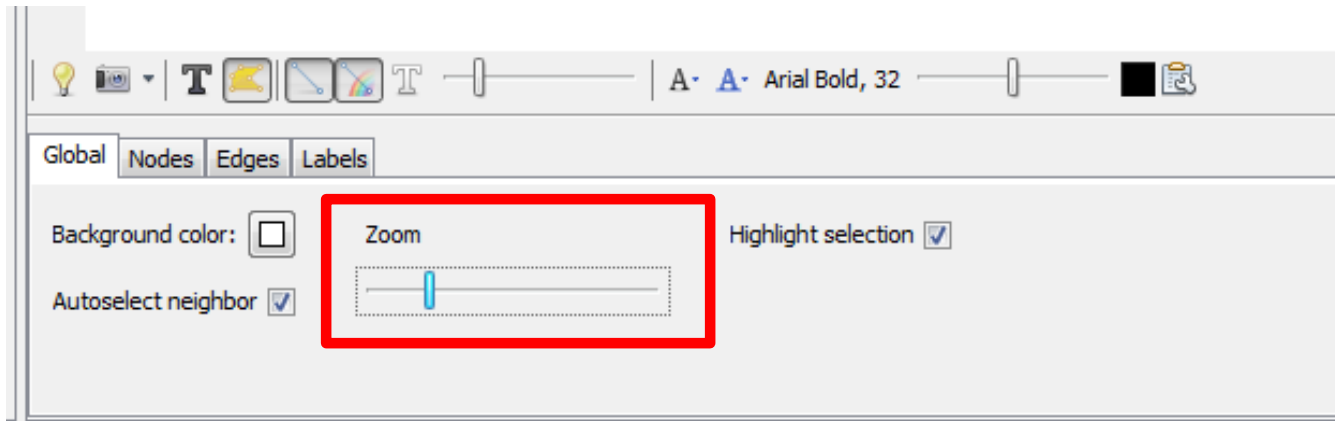
Visualization



Visualization Options

Visualization

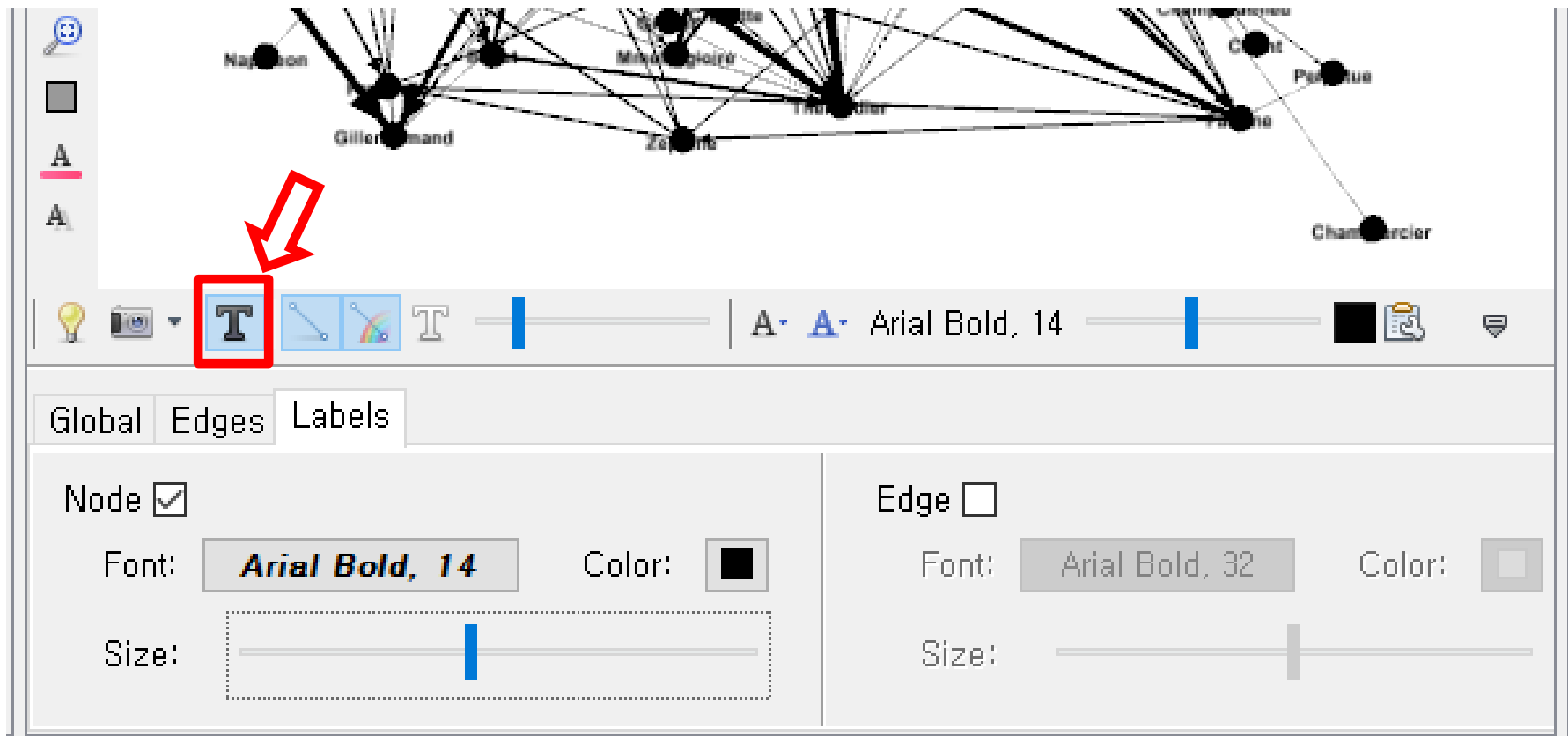
- Zoom In/Out (bottom panel)



- Move the screen
 - Mouse right click + Drag

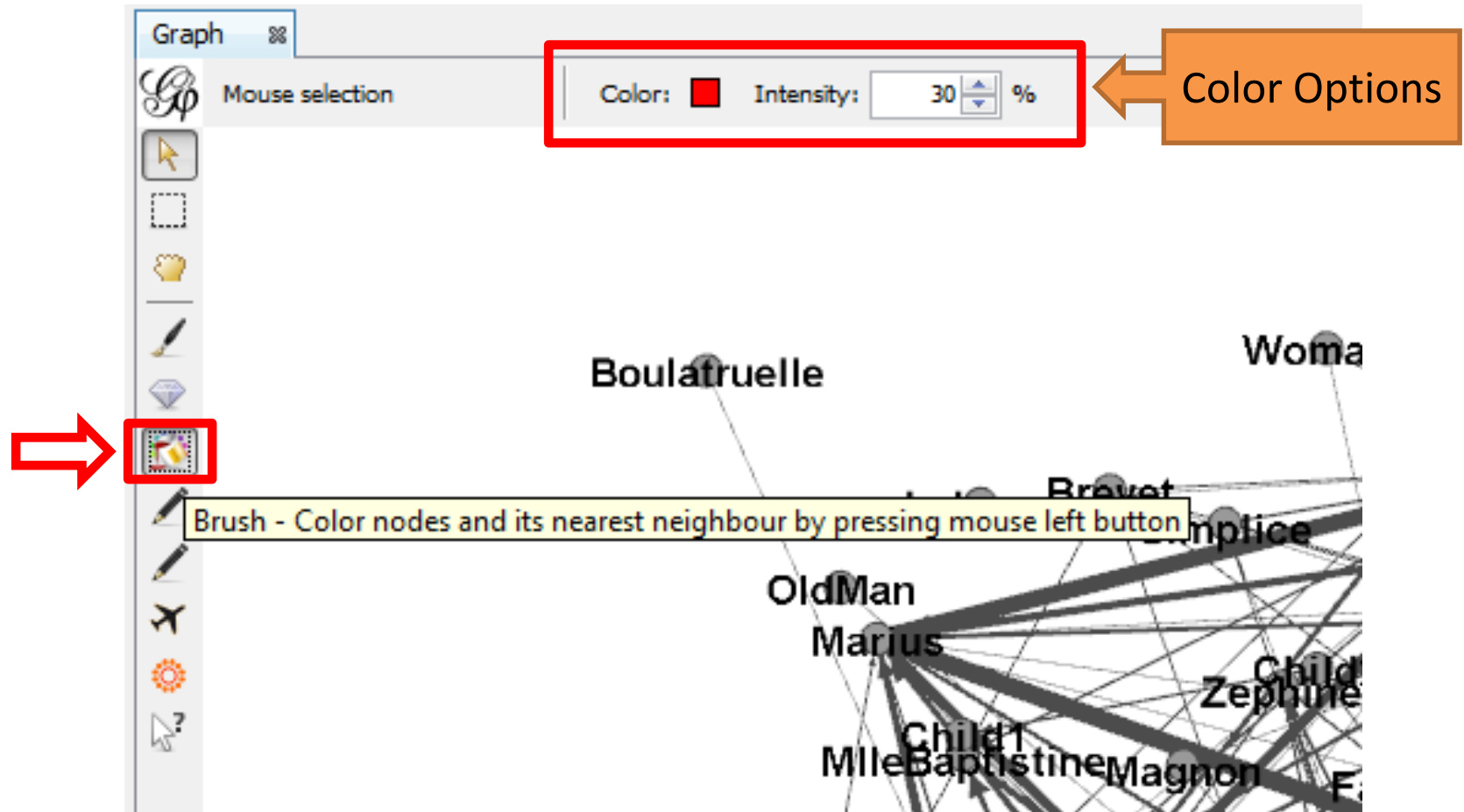
Show Node Label

- At the bottom panel, click 



Color a particular node and its neighbors

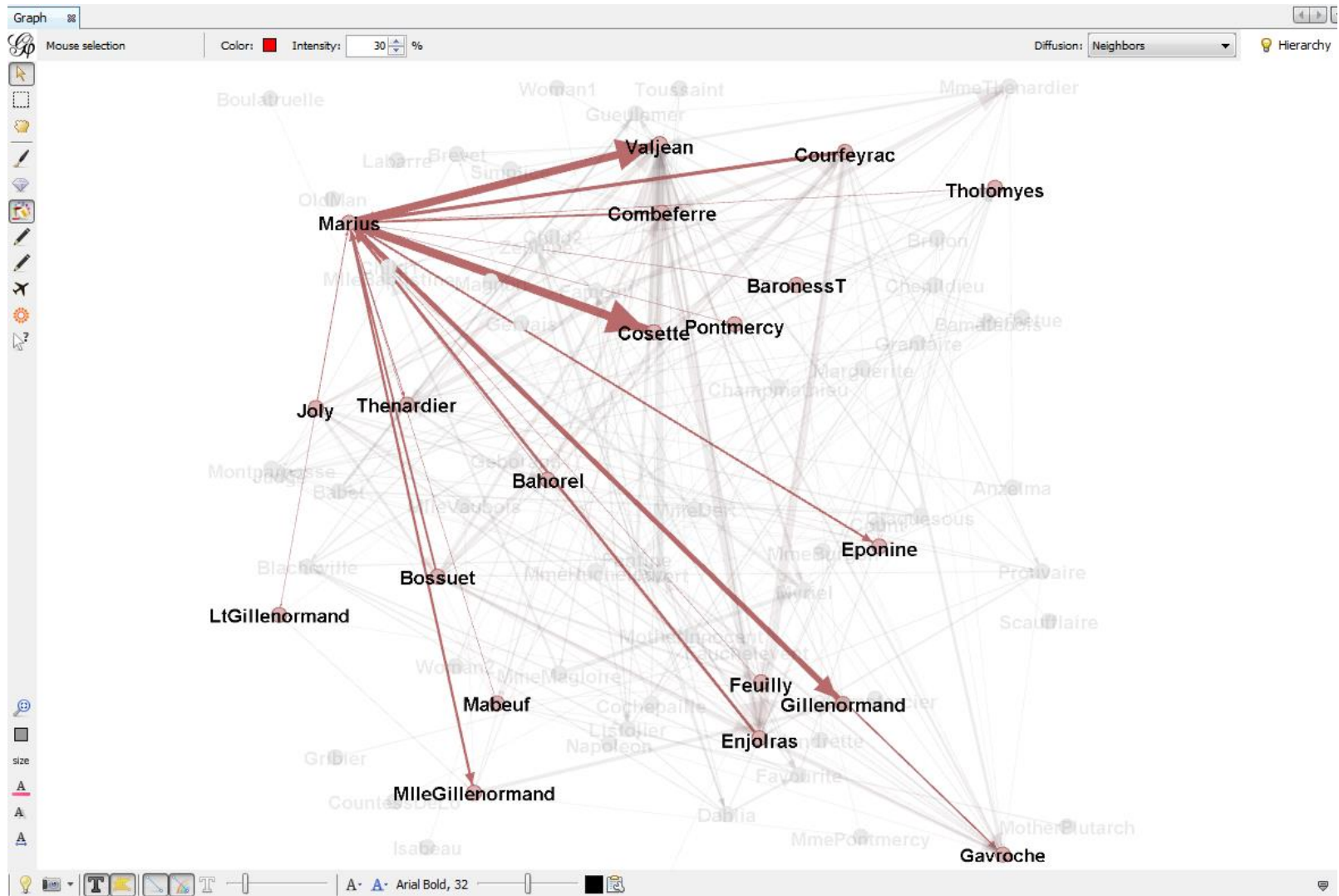
- On the left panel, click



The screenshot shows the Graph tool interface. On the left is a vertical toolbar with various icons. A red arrow points to the 'Brush' icon, which is highlighted with a red box. A tooltip for the Brush icon reads: "Brush - Color nodes and its nearest neighbour by pressing mouse left button". Above the graph, there is a 'Color Options' panel with a red border. It contains a 'Color' dropdown set to red, an 'Intensity' slider set to 30%, and a '%' sign. An orange arrow points from the 'Color Options' label to this panel. The main graph area displays a network of nodes and edges. Nodes are labeled: Boulatruelle, Woma, Brevet, Complice, OldMan, Marius, Child1, MlleBaptistine, Magnon, Zephine, and F. The graph shows a dense cluster of connections between these nodes.

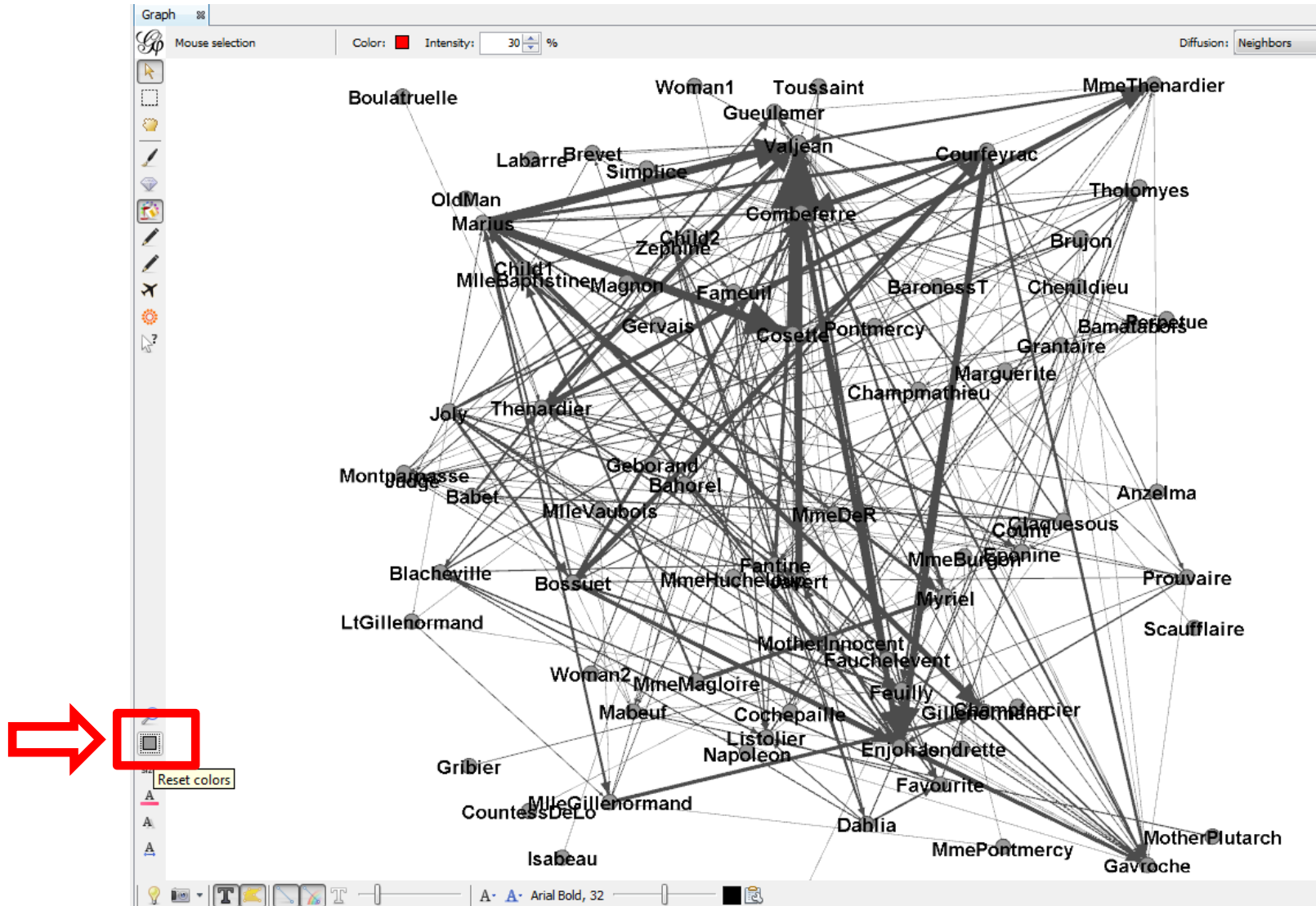
Color a particular node and its neighbors

- Click the target node.

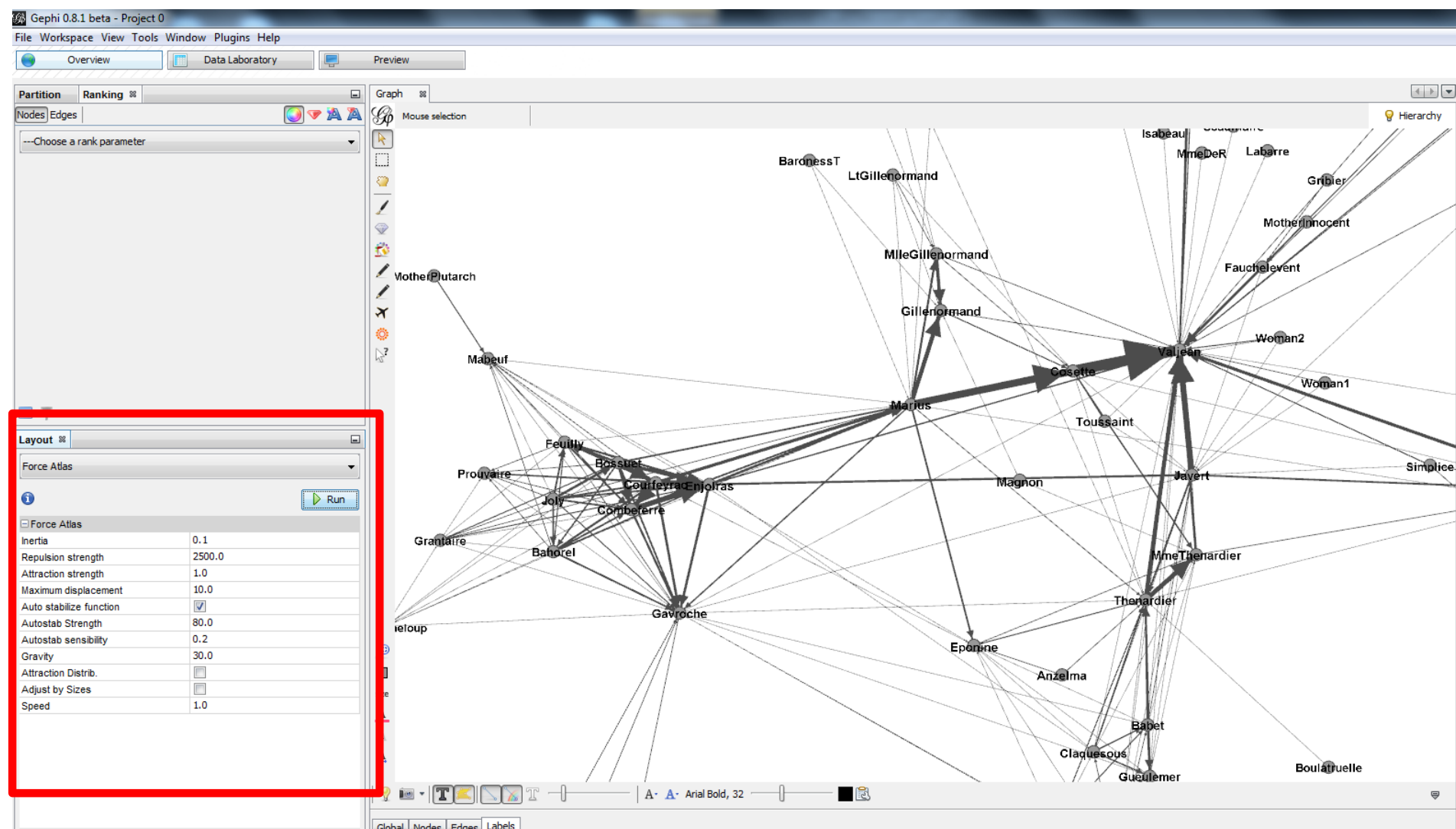


Reset the color

- On the left panel, click 

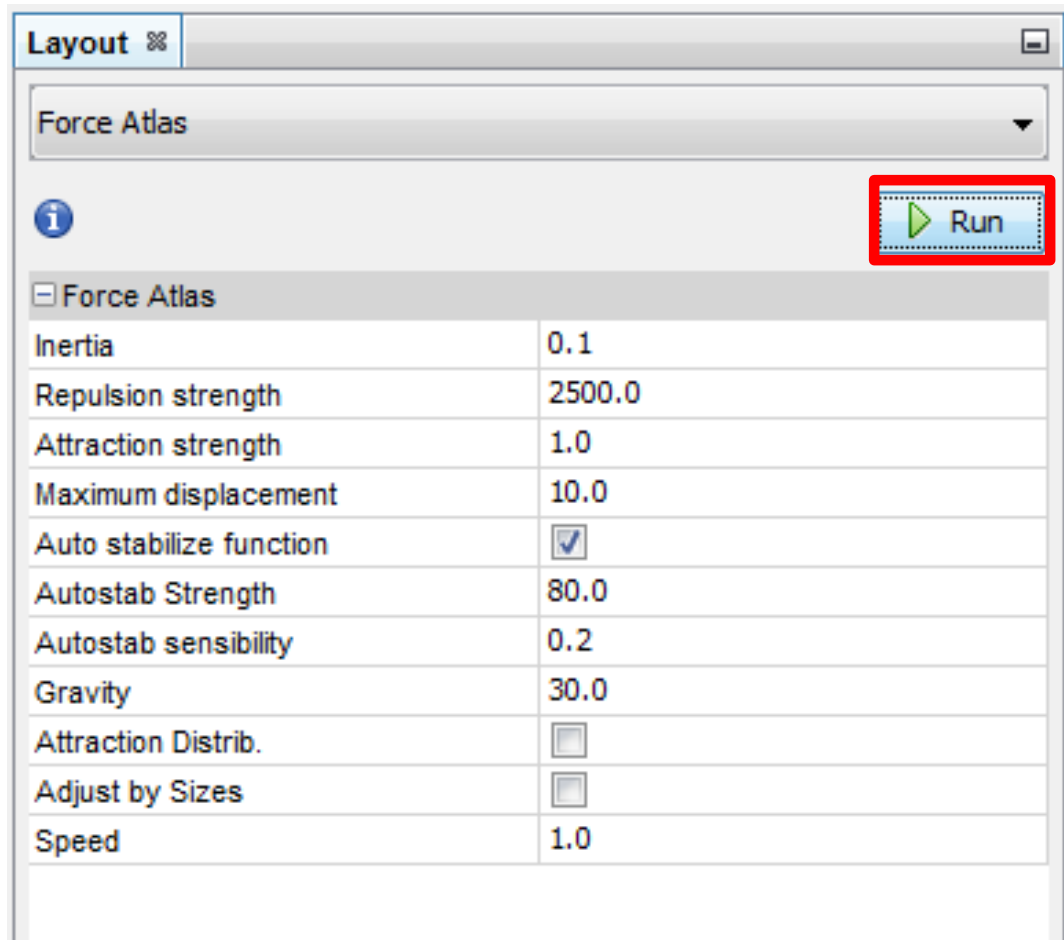


Graph Layout



Graph Layout

- Change the parameters as follows:



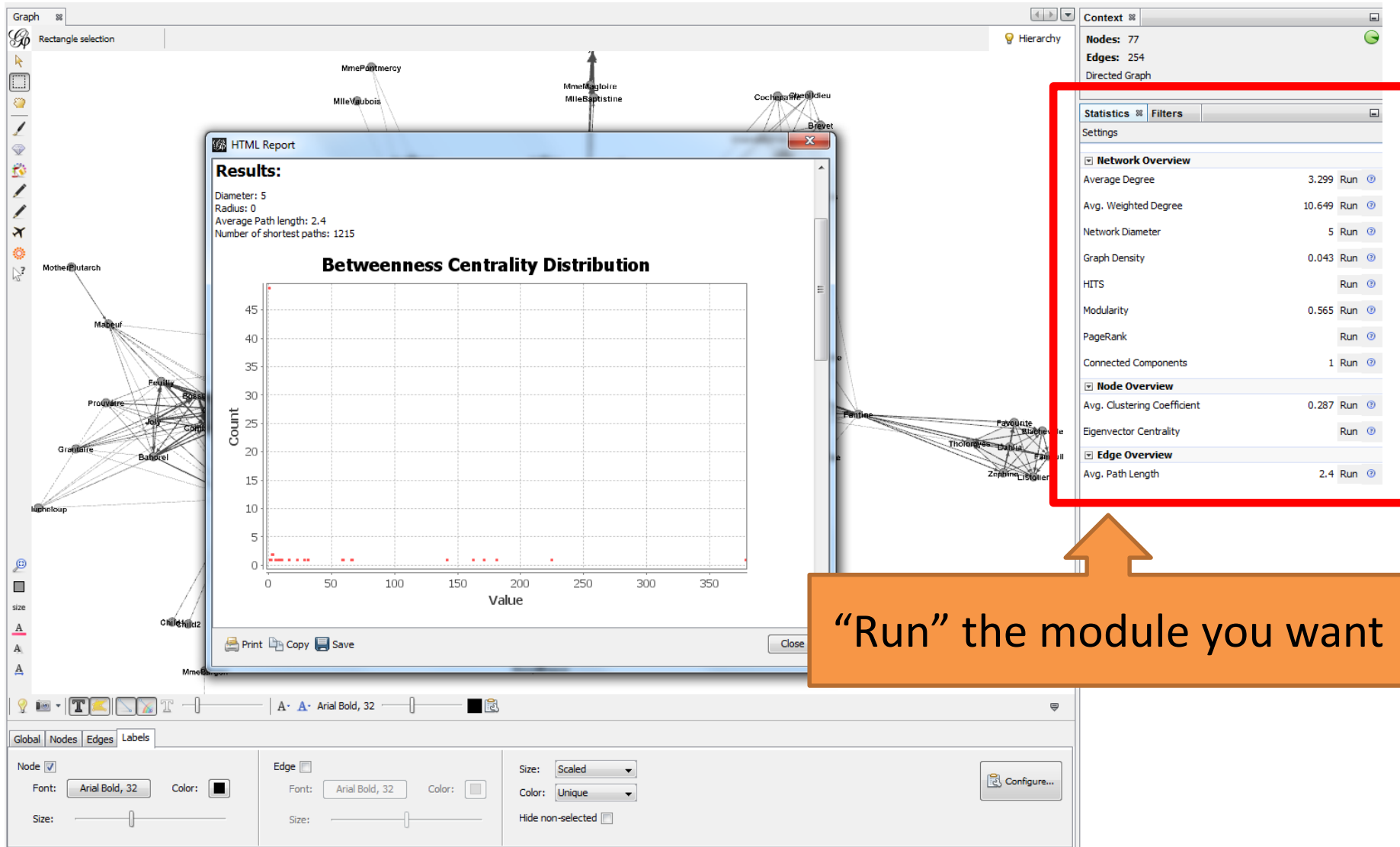
The screenshot shows a window titled 'Layout' with a dropdown menu set to 'Force Atlas'. Below the menu is an information icon (i) and a 'Run' button, which is highlighted with a red dashed border. The 'Run' button features a green play icon and the text 'Run'. Below these elements is a table of parameters for the 'Force Atlas' layout.

Force Atlas	
Inertia	0.1
Repulsion strength	2500.0
Attraction strength	1.0
Maximum displacement	10.0
Auto stabilize function	<input checked="" type="checkbox"/>
Autostab Strength	80.0
Autostab sensibility	0.2
Gravity	30.0
Attraction Distrib.	<input type="checkbox"/>
Adjust by Sizes	<input type="checkbox"/>
Speed	1.0

Network Statistics

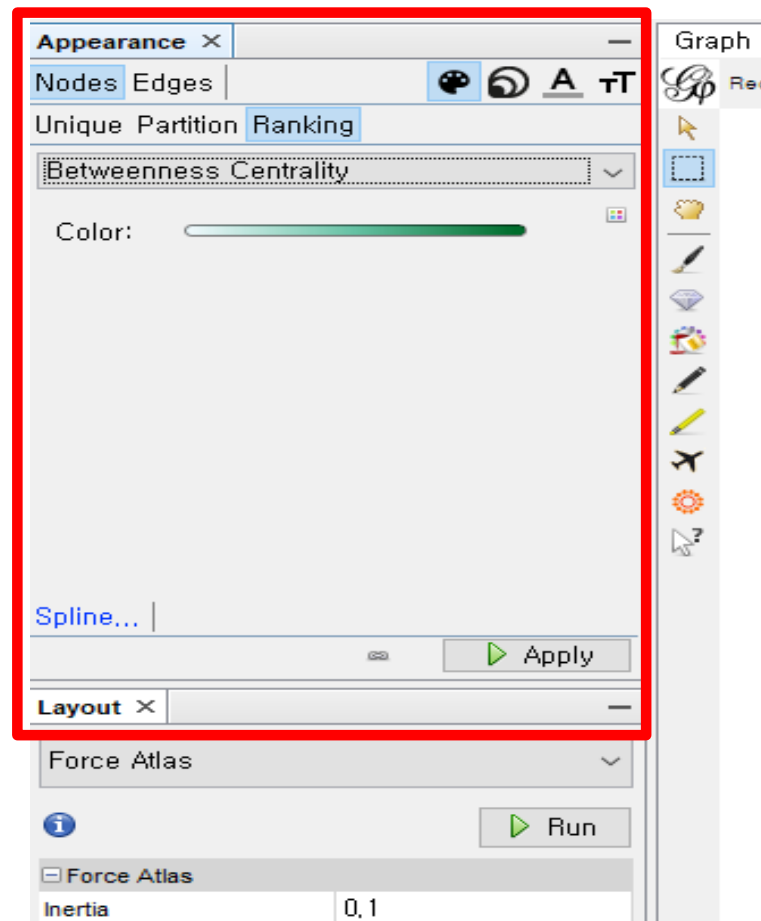
- We can calculate the following statistics:
 - Average Degree
 - Network Diameter
 - Graph Density
 - Hubs and Authorities
 - Modularity
 - PageRank
 - Clustering Coefficient
 - Eigenvector Centrality

Network Statistics



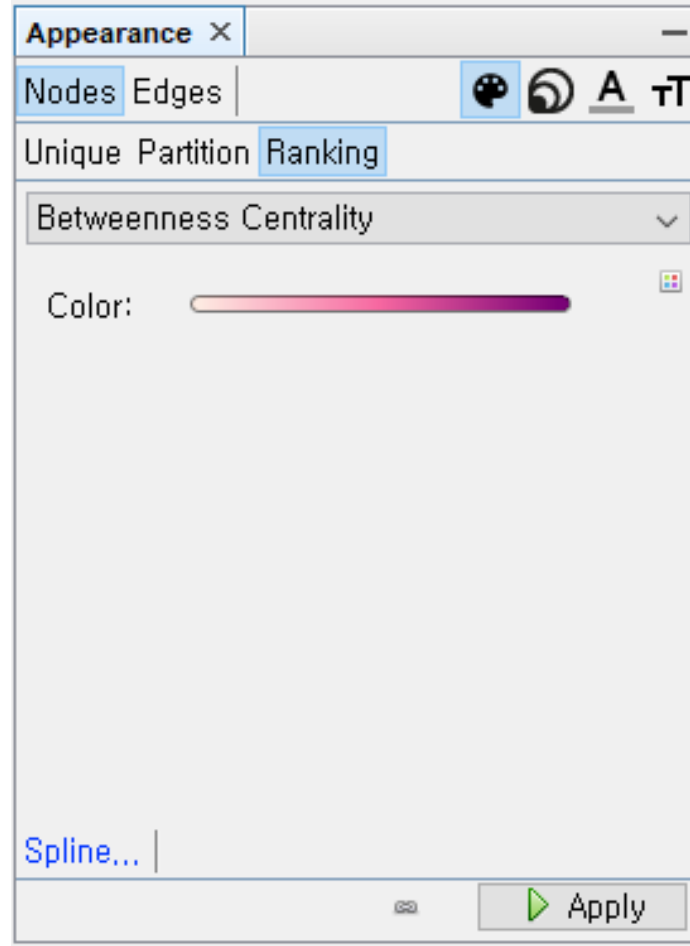
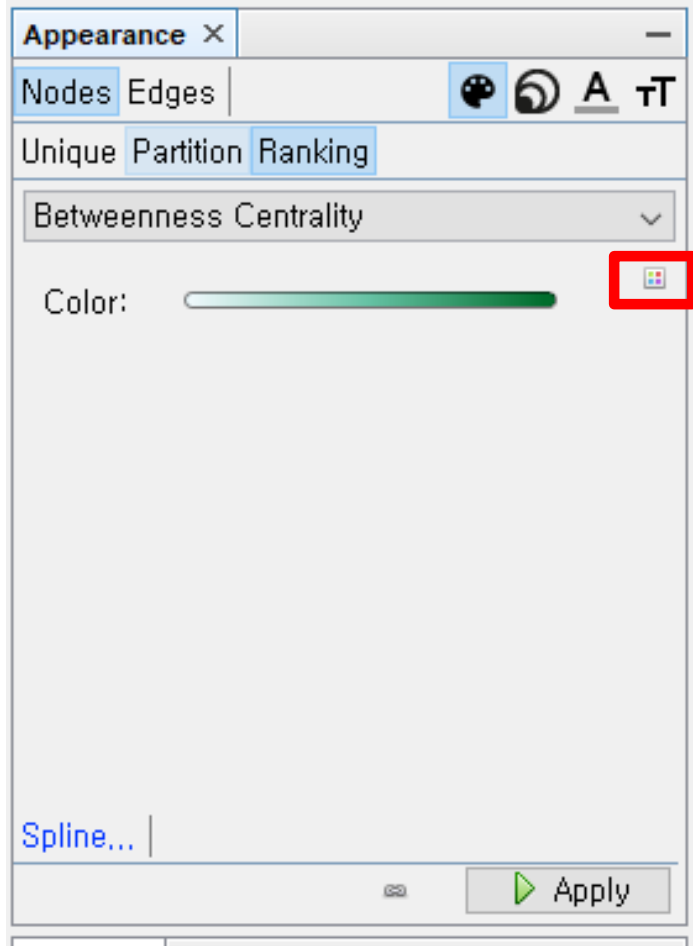
Change Node Color

- We can change node color according to its statistics (e.g. degree, Betweenness centrality, ...)



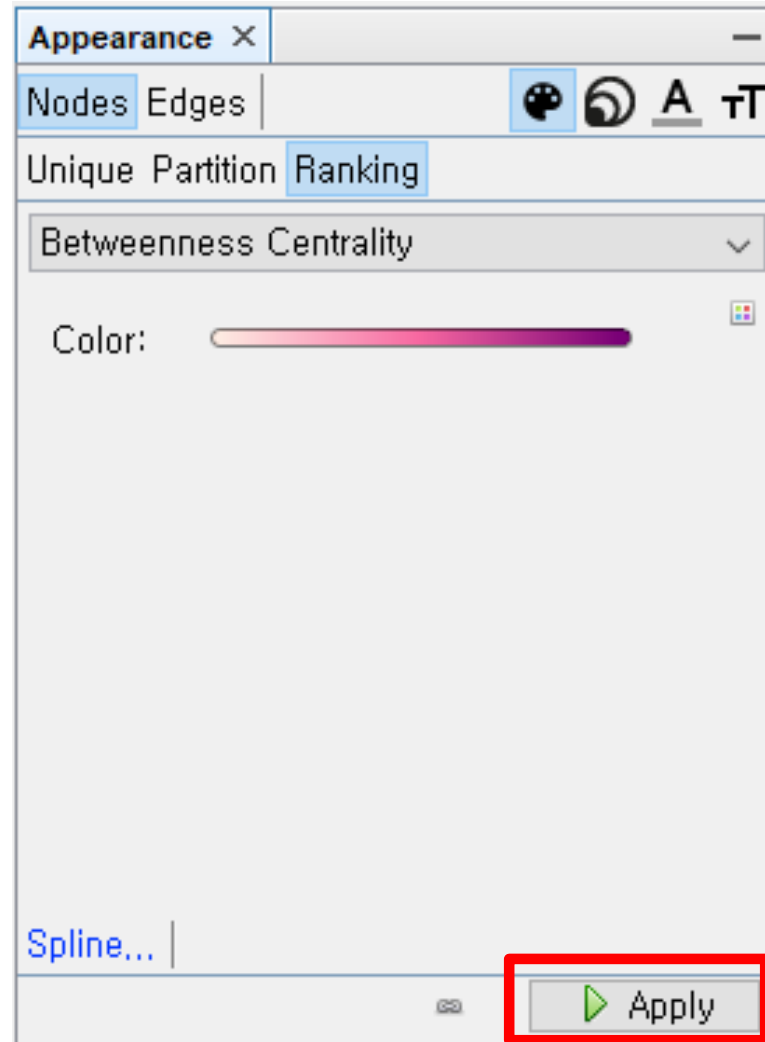
Change Node Color

- Right Icon click > Default > Select color

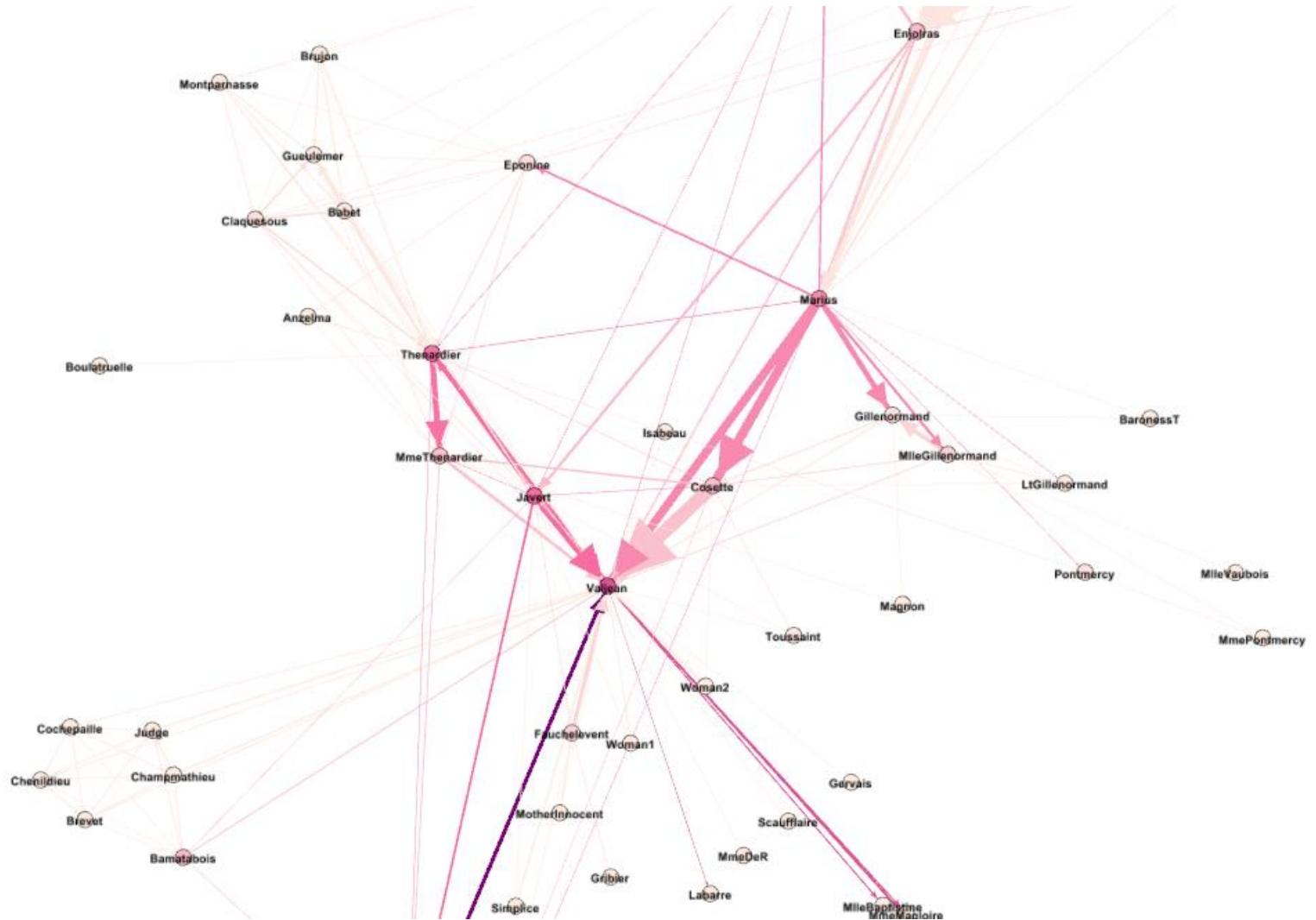


Change Node Color

- Click “apply”



Change Node Color



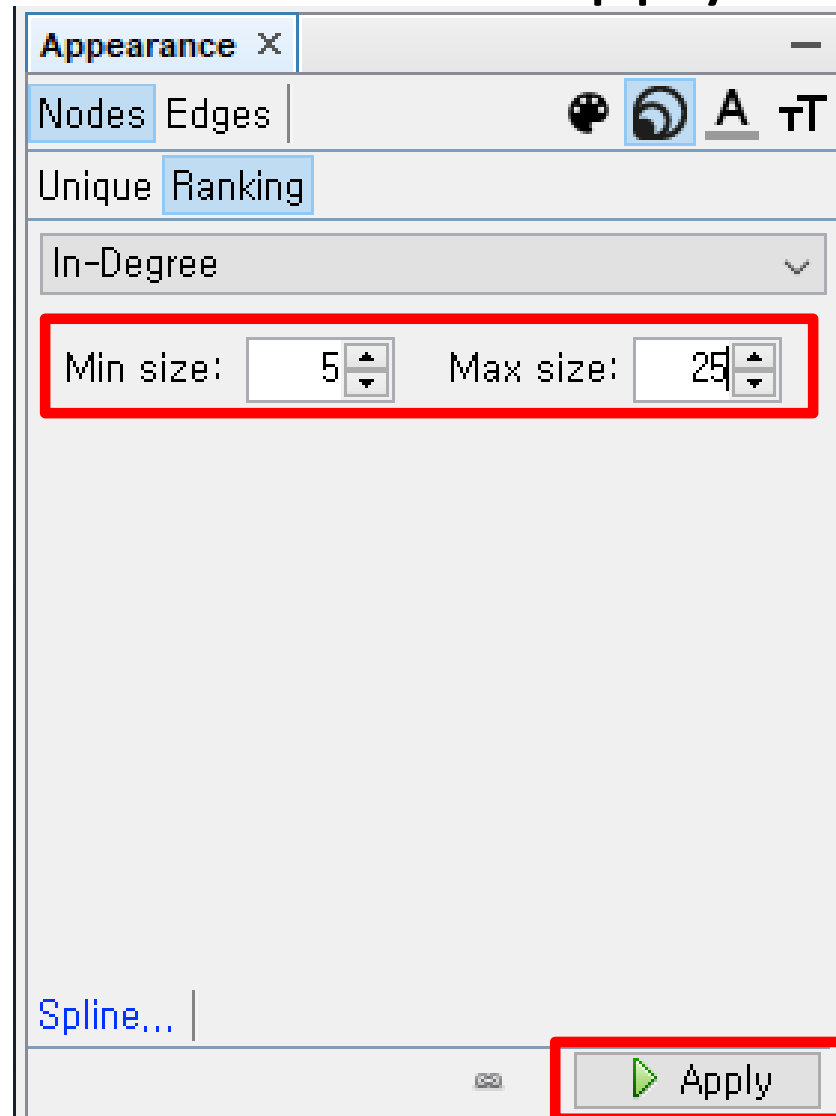
Change Node Size

- Click  > Select statistics

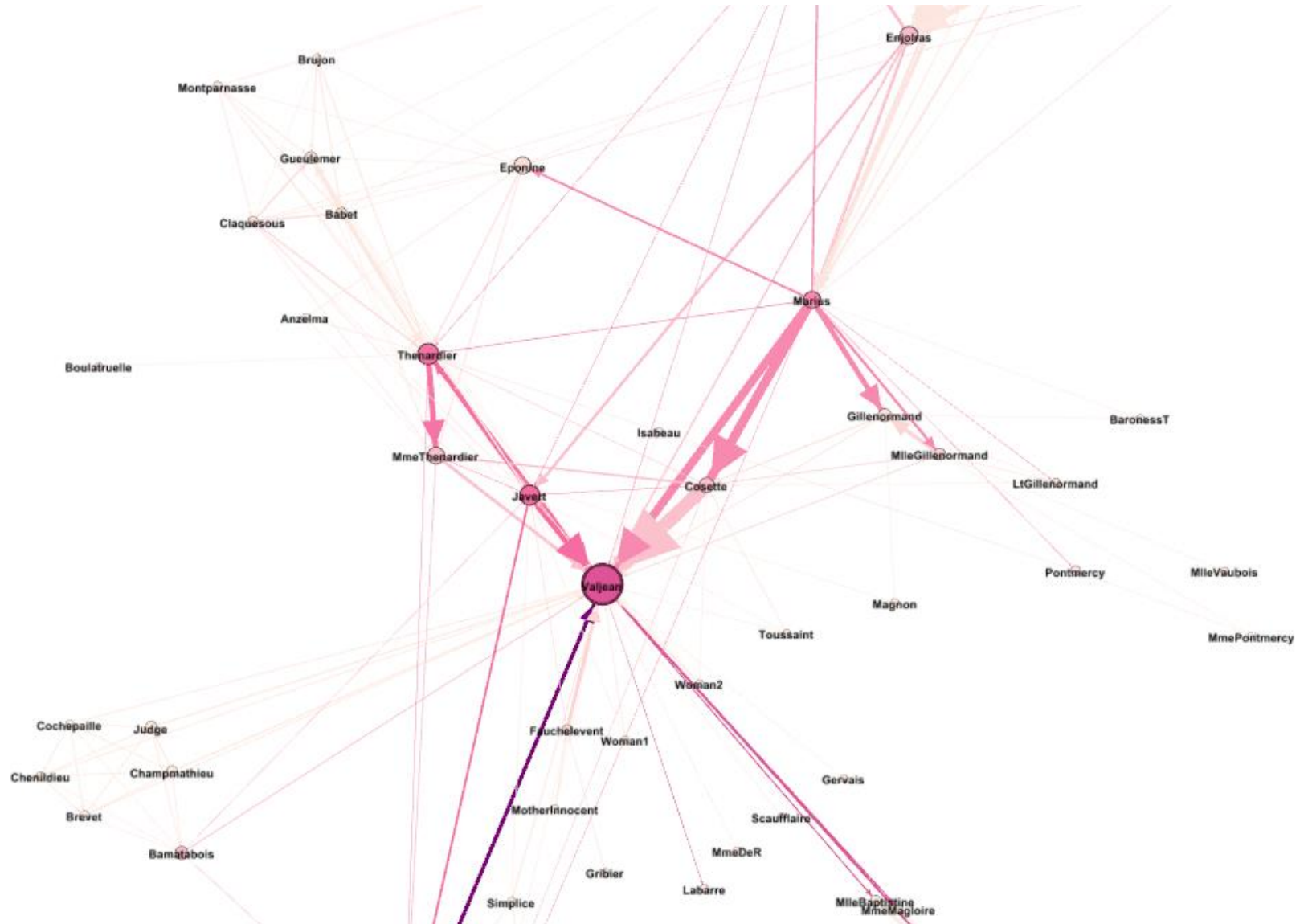


Change Node Size

- Set Min/Max size > Click “apply”

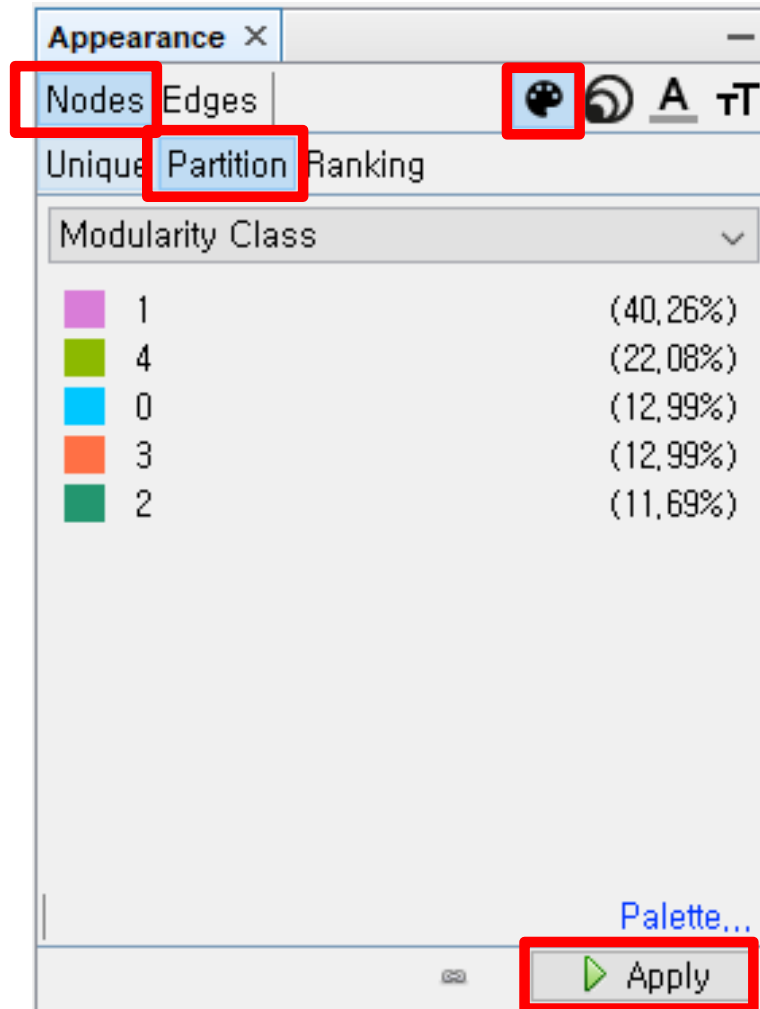


Change Node Size

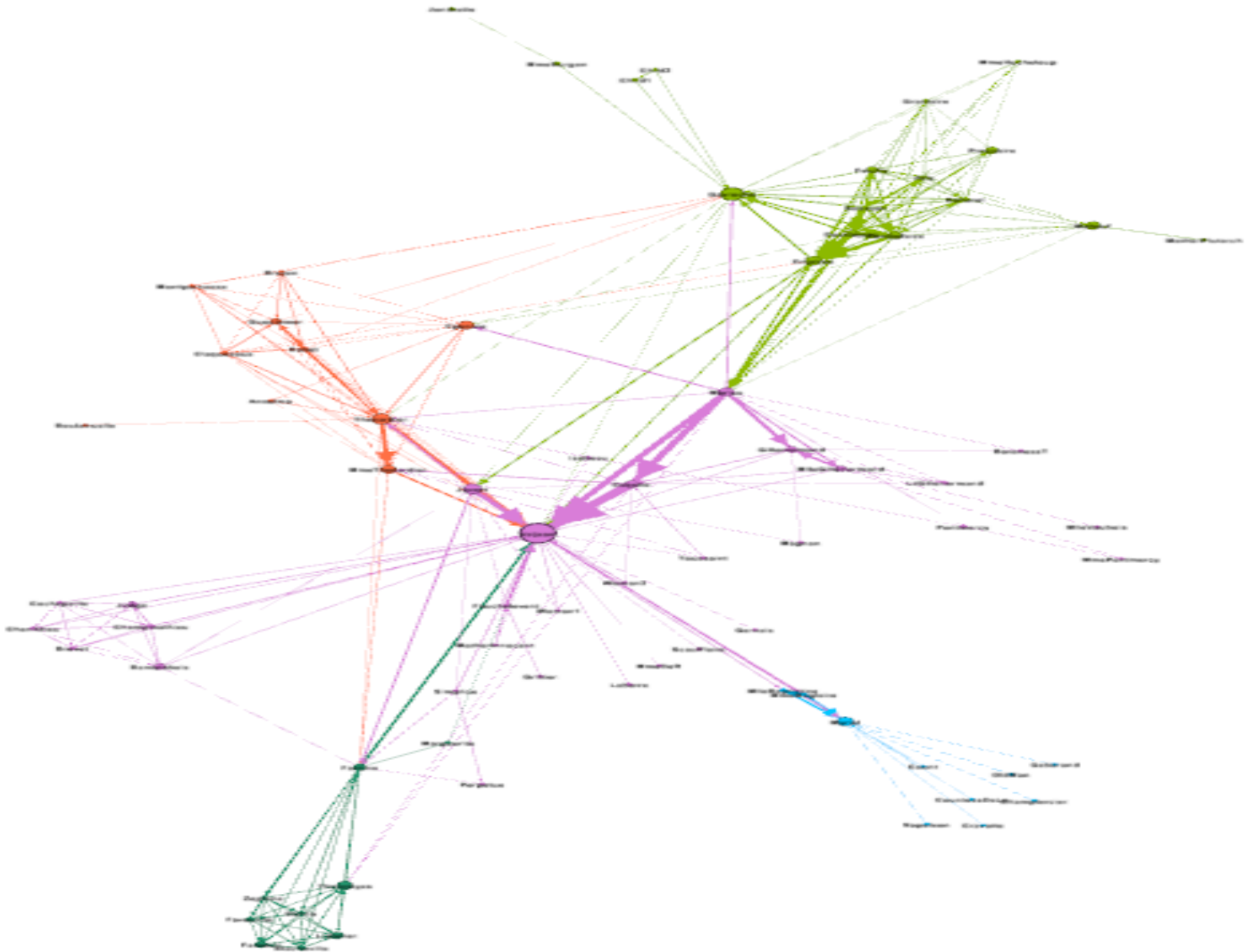


Partition the Graph

- Click Nodes > Click  > Click “Partition” tab > Select “Modularity Class” > Click “Apply”

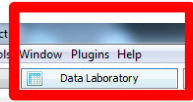


Partition the Graph



Data Editing

- Click “Data Laboratory” tab



Gephi 0.8.1 beta - Project

File Workspace View Tools Window Plugins Help

Overview Data Laboratory Preview

Data Table

Nodes Edges Configuration Add node Add edge Search/Replace Import Spreadsheet Export table More actions Filter: Nodes

Nodes	Id	Label	In-Degree	Out-Degree	Degree	Weighted De...	Weighted In...	Weighted O...	Eccentricity	Closeness Ce...	Betweenness C...	Authority	Hub	Modularity...	PageRank	Componen...	Strongly-C...	Clustering Co...	Eigenvector ...
● Labarre	10.0	Labarre	1	0	1	1.0	1.0	0.0	0	0	0	0.007	0	4	0.027	0	0	0	0.274
● MmeMagloire	3.0	MmeMagloire	1	2	3	19.0	3.0	16.0	1	1	0	0.007	0.008	0	0.027	0	3	0.5	0.274
● MlleBaptistine	2.0	MlleBaptistine	2	1	3	17.0	9.0	8.0	1	1	0	0.01	0.012	0	0.039	0	2	0.5	0.338
● OldMan	9.0	OldMan	0	1	1	1.0	0.0	1.0	1	1	0	0	0.004	0	0.004	0	4	0	0
● Count	8.0	Count	0	1	1	2.0	0.0	2.0	1	1	0	0	0.004	0	0.004	0	5	0	0
● Cravatte	7.0	Cravatte	0	1	1	1.0	0.0	1.0	1	1	0	0	0.004	0	0.004	0	6	0	0
● Champiercier	6.0	Champiercier	0	1	1	1.0	0.0	1.0	1	1	0	0	0.004	0	0.004	0	7	0	0
● MmeBurgon	47.0	MmeBurgon	1	1	2	3.0	2.0	1.0	1	1	22	0.007	0.008	3	0.01	0	9	0	0.131
● Jondrette	46.0	Jondrette	1	0	1	1.0	1.0	0.0	0	0	0	0.007	0	3	0.013	0	8	0	0.035
● Scaufflaire	32.0	Scaufflaire	0	1	1	1.0	0.0	1.0	2	1.8	0	0	0.004	4	0.004	0	11	0	0
● Geborand	5.0	Geborand	0	1	1	1.0	0.0	1.0	1	1	0	0	0.004	0	0.004	0	12	0	0
● Gribier	45.0	Gribier	0	1	1	2.0	0.0	2.0	4	3.211	0	0	0.004	4	0.004	0	33	0	0
● Boulatrue	40.0	Boulatrue	0	1	1	1.0	0.0	1.0	3	2.688	0	0	0.004	2	0.004	0	34	0	0
● Simplic	31.0	Simplic	0	4	4	8.0	0.0	8.0	2	1.789	0	0	0.004	4	0.004	0	36	0.333	0
● Gervais	15.0	Gervais	0	1	1	1.0	0.0	1.0	2	1.8	0	0	0.004	4	0.004	0	37	0	0
● CountessDeLo	4.0	CountessDeLo	0	1	1	1.0	0.0	1.0	1	1	0	0	0.004	0	0.004	0	38	0	0
● MotherInnocent	44.0	MotherInnocent	0	2	2	4.0	0.0	4.0	4	2.947	0	0	0.004	4	0.004	0	39	0.5	0
● Fauchelevent	28.0	Fauchelevent	2	2	4	14.0	5.0	9.0	3	2.333	31	0.01	0.004	4	0.009	0	32	0.167	0.005
● Woman1	33.0	Woman1	0	2	2	3.0	0.0	3.0	3	2.333	0	0	0.004	4	0.004	0	40	0.5	0
● Toussaint	72.0	Toussaint	0	3	3	4.0	0.0	4.0	3	2.222	0	0	0.004	4	0.004	0	41	0.5	0
● Feully	61.0	Feully	5	6	11	38.0	21.0	17.0	4	2.484	1.111	0.02	0.024	3	0.007	0	53	0.464	0.059
● Combeferre	59.0	Combeferre	7	4	11	68.0	40.0	28.0	4	2.586	5.361	0.026	0.032	3	0.01	0	51	0.464	0.119
● Bossuet	64.0	Bossuet	3	10	13	66.0	11.0	55.0	4	2.206	7.837	0.013	0.012	3	0.005	0	56	0.385	0.015
● Prouvaire	60.0	Prouvaire	6	3	9	19.0	12.0	7.0	4	2.833	0	0.023	0.028	3	0.008	0	52	0.5	0.085
● Child2	74.0	Child2	0	2	2	5.0	0.0	5.0	4	3	0	0	0.004	3	0.004	0	58	0.5	0
● Child1	73.0	Child1	1	1	2	5.0	3.0	2.0	4	3.095	0	0.007	0.004	3	0.006	0	57	0.5	0.003
● Anzelma	42.0	Anzelma	0	3	3	5.0	0.0	5.0	3	2.529	0	0	0.004	2	0.004	0	59	0.5	0
● BaronessT	56.0	BaronessT	0	2	2	2.0	0.0	2.0	4	2.815	0	0	0.004	4	0.004	0	60	0.5	0
● Brujon	75.0	Brujon	0	7	7	13.0	0.0	13.0	4	2.515	0	0	0.004	2	0.004	0	65	0.452	0
● Perpetue	30.0	Perpetue	1	1	2	3.0	2.0	1.0	3	2.214	0	0.007	0.004	4	0.005	0	35	0.5	0.003
● Fantine	23.0	Fantine	6	9	15	47.0	12.0	35.0	2	1.308	377.767	0.023	0.024	5	0.044	0	21	0.157	0.358
● MmePontmercy	52.0	MmePontmercy	0	2	2	2.0	0.0	2.0	4	2.95	0	0	0.004	4	0.004	0	66	0	0
● Pontmercy	39.0	Pontmercy	2	1	3	3.0	2.0	1.0	3	2.688	9.5	0.01	0.008	4	0.007	0	43	0.167	0.082
● LtGillenormand	54.0	LtGillenormand	1	3	4	5.0	1.0	4.0	4	2.842	0	0.007	0.008	4	0.005	0	47	0.5	0.079
● Tenchov	14.0	Tenchov	0	1	1	1.0	0.0	1.0	2	1.8	0	0	0.004	4	0.004	0	63	0	0

Add column Merge columns Delete column Clear column Copy data to other column Fill column with a value Duplicate column

Create a boolean column from regex match Create column with list of regex matching groups Negate boolean values

Workspace 0

Export Graph

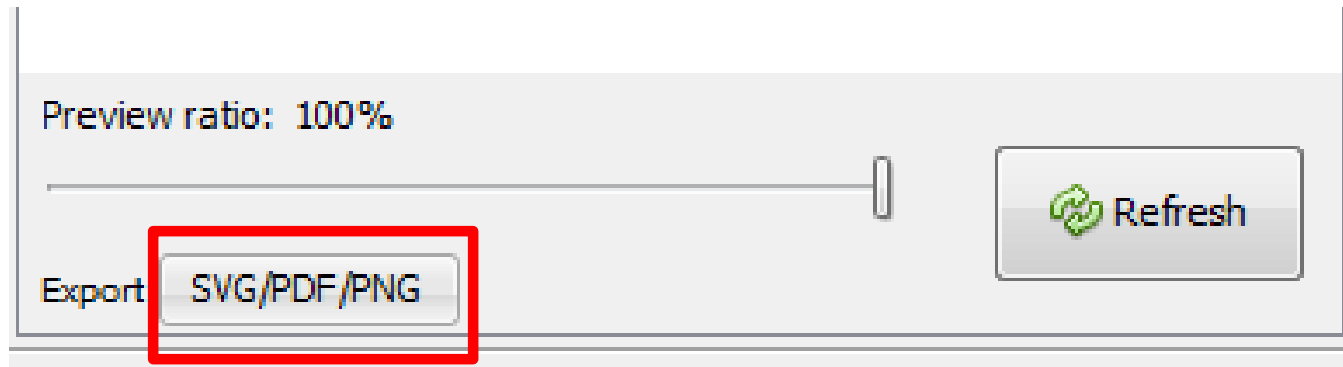
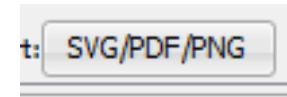
- Click “Preview” tab > Click

Refresh

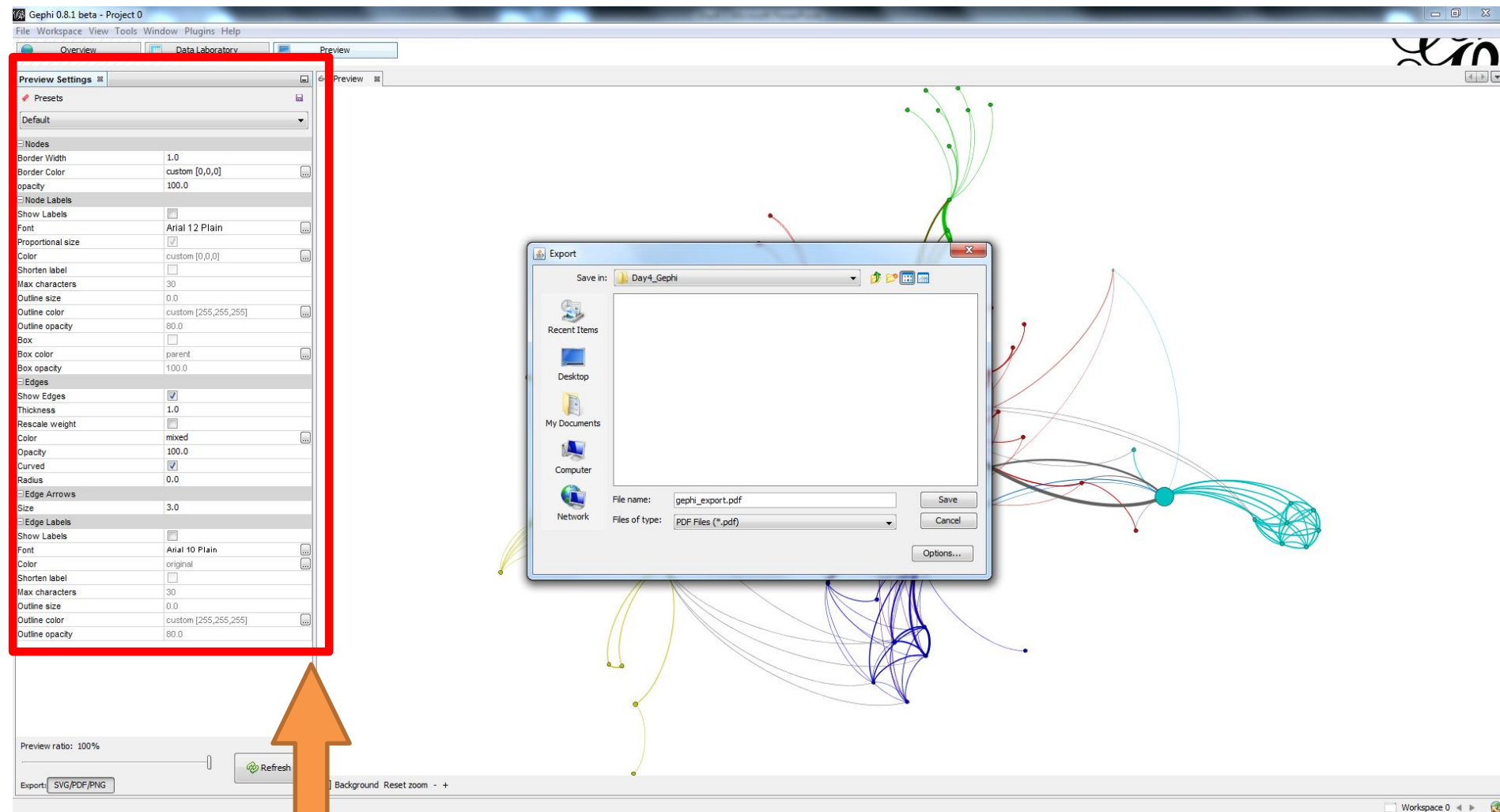
The screenshot displays the Cytoscape software interface. At the top, the 'Preview' tab is selected and highlighted with a red box. A red arrow points to this tab. Below the tab bar, the 'Preview Settings' panel is visible on the left, containing various configuration options for nodes, edges, and labels. The 'Refresh' button in this panel is also highlighted with a red box. The main workspace on the right shows a complex network graph with numerous nodes and edges, rendered in a colorful, multi-layered style. The status bar at the bottom indicates the export format as 'SVG/PDF/PNG' and includes a 'Background Reset zoom' option.

Export Graph

- At the bottom of the panel, click



Export Graph



You can find more options, here.
Before exporting graph, you can
apply these options.

Dynamic Networks

- We have a network at three different year 2007, 2008 and 2009 with three nodes.



```
<nodes>
  <node id="n1" label="Node 1" start="2007" endopen="2009" />
  <node id="n2" label="Node 2" start="2007" end="2009" />
  <node id="n3" label="Node 3" start="2008" end="2009" />
</nodes>
```

More information:

<https://github.com/gephi/gephi/wiki/Import-Dynamic-Data>

Reference

- Gephi Tutorial, <https://gephi.org/>