Frontiers of Network Science
Fall 2023

Class 8:Introduction to Gephi:
Tools for Network Analysis and Visualization

Boleslaw Szymanski

Class Plan

Overview of tools for network analysis and visualization

Installing and using Gephi

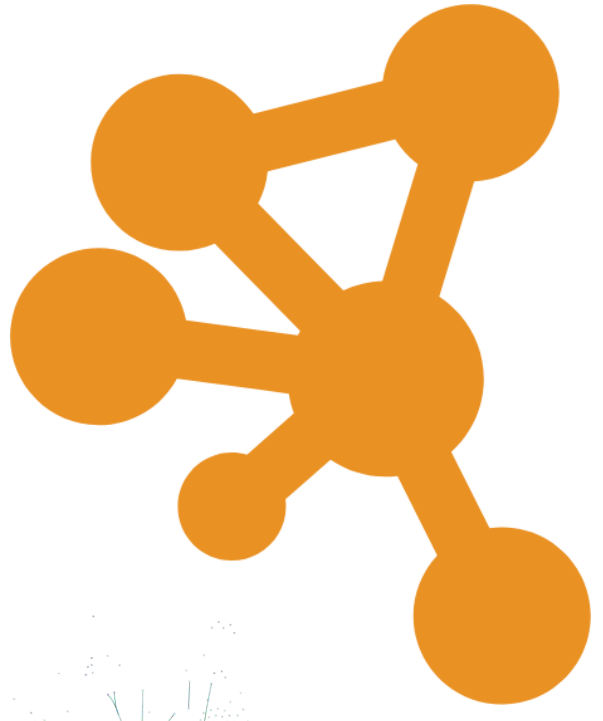
Gephi hands-on lab exercises

Tool for network analysis and visualization

- Computing model and interfaces
- GUI or API/code libraries
- Extensibility by the community (plugins)
- Source code availability

Cytoscape

- https://cytoscape.org/
- GUI application
- Allows visualization and analysis of large complex networks
- Originally developed for biological research
- Several layout algorithms
- Publication-quality images
- Written in Java
- Extensible through plugins
- Free and open source
- Windows, Linux, MacOS



Gephi

- https://gephi.org/
- GUI application
- Allows visualization and analysis of large complex networks and dynamic networks
- Clustering and community detection
- Layout algorithms supported
- Produces publication-quality images
- Written in Java
- Extensible through plugins
- Free and open source
- Windows, Linux, MacOS

Graphviz

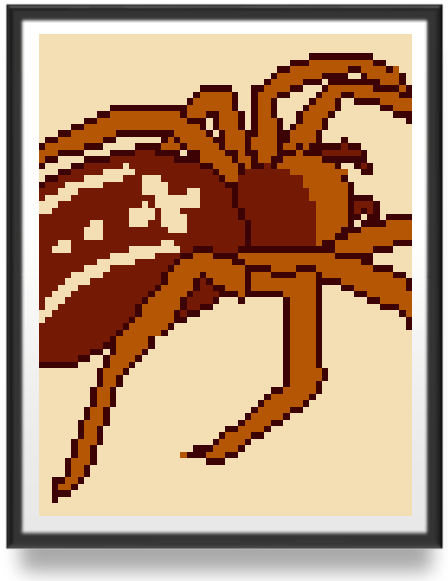
- https://graphviz.org/
- Graph description language (called DOT) + tools to handle DOT files
- Only performs drawing
- Wide range of graph layouts
- Written in C
- Extensible through a scripting API
- Free and open source

Neo4J

- https://neo4j.com/
- Graph Database
- Efficient management of semi-structured and network-oriented data
- Robust – ACID transactions, transaction recovery
- Highly scalable
- High performance – query optimizer, index-free adjacency
- Free

Pajek and Pajek-XXL

- http://mrvar.fdv.uni-lj.si/pajek/
- GUI application
- Network generator (Random, Bernoulli/Poisson, Scale-free, etc.)
- Capable of analyzing and visualizing large networks
- Free (noncommercial)
- Windows only

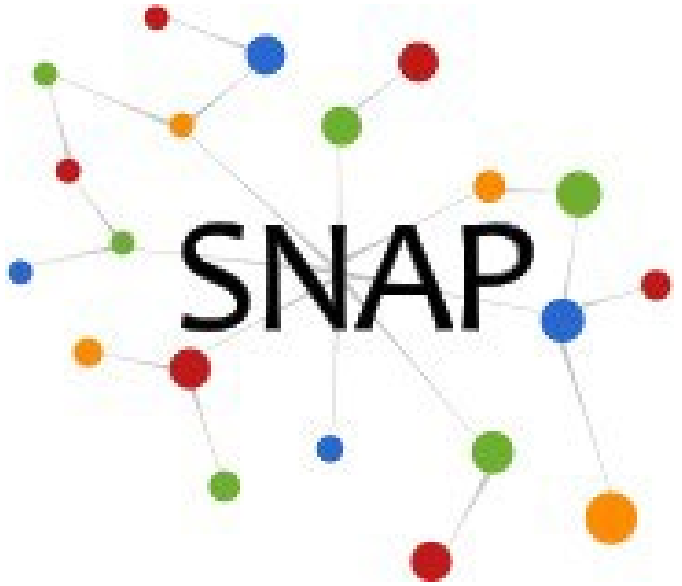


NetworkX

- https://networkx.github.io/
- Python library
- Creation, manipulation and analysis of networks
- Supports various network formats
- Provides standard graph algorithms
- Network analysis and generation
- Visualization with Matplotlib and Graphviz
- Free and open source

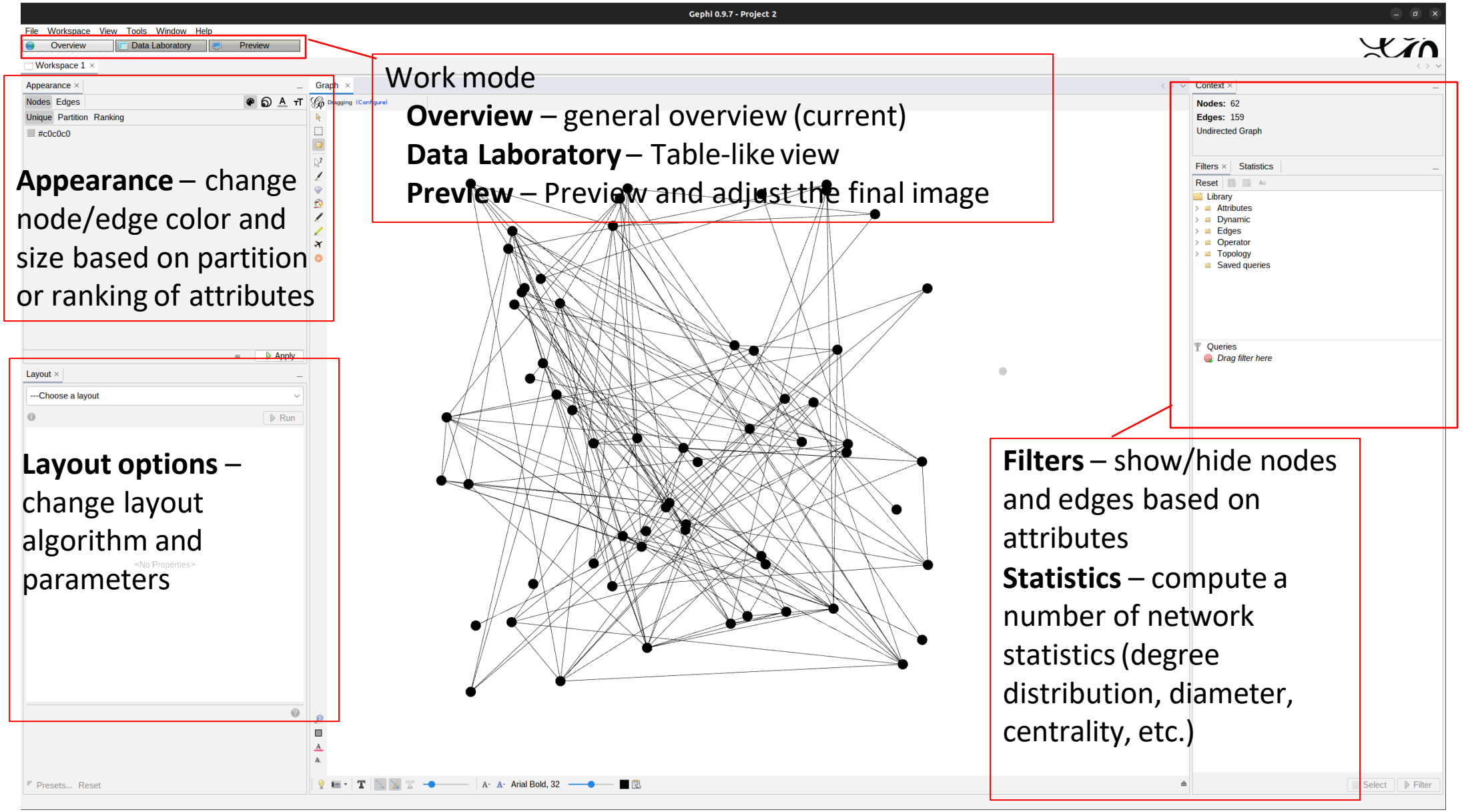
SNAP

- http://snap.stanford.edu/index.html
- C++ library – Python interface is available
- Scales to massive networks with hundreds of millions of nodes
- Efficiently manipulates large graphs
- Available through NodeXL – graphical front-end that integrates network analysis into Microsoft Office



Installing Gephi

- Download it from https://gephi.org/
- Current version is 0.9.5 (May 3 2022)
- Requires the Java Runtime Environment (JRE) 7 or 8
- **Important**: By default, Gephi is configured to start with 512 MB of memory allocated to the JVM. This might not be enough for larger graphs. To increase the memory allocated to the JVM, change the **–Xmx** option in file *gephi.conf* (e.g., set **–Xmx1024m**) then close and reopen Gephi for the new options to take effect.



Work mode

Overview – general overview (current)

Data Laboratory – Table-like view

Preview – Preview and adjust the final image

Appearance – change node/edge color and size based on partition or ranking of attributes

Layout options – change layout algorithm and parameters

Filters – show/hide nodes and edges based on attributes

Statistics – compute a number of network statistics (degree distribution, diameter, centrality, etc.)

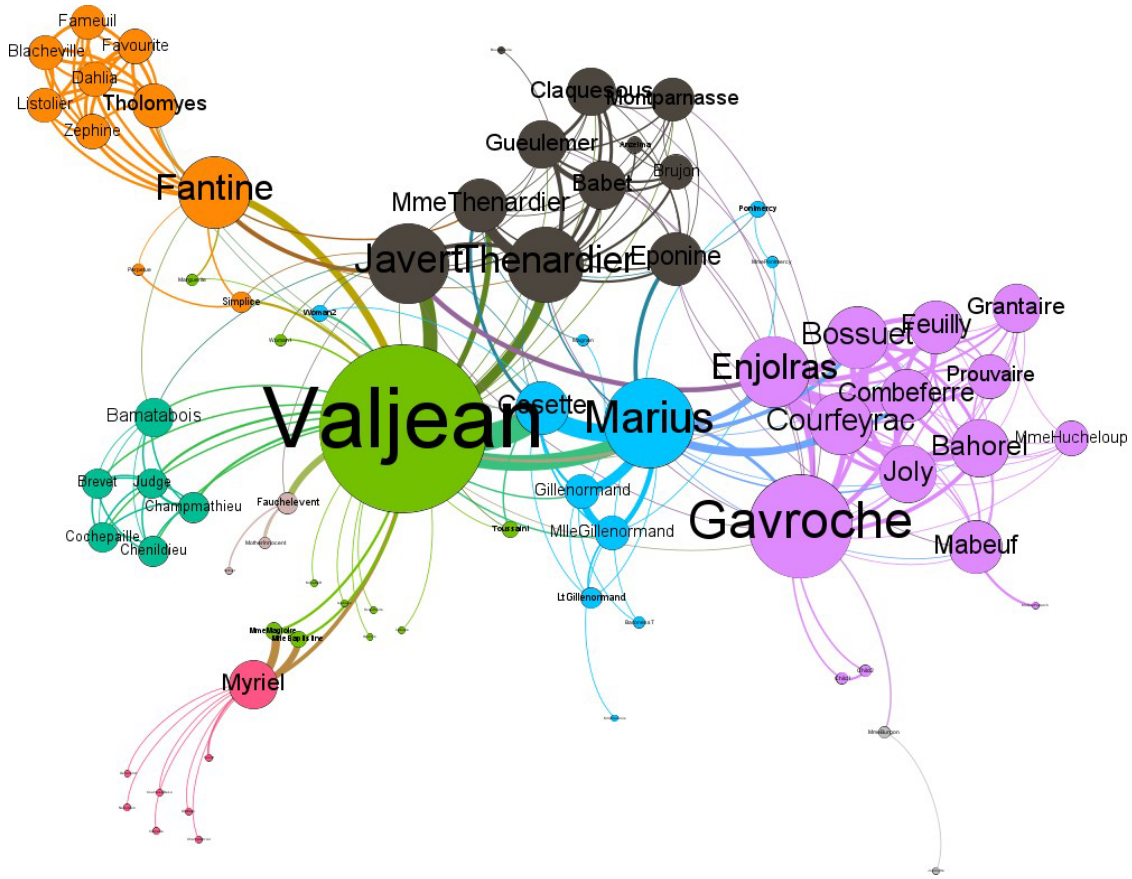
Sample Network: Les Misérables

- This network comes with Gephi version 0.9.5*
- Based on Victor Hugo's novel *Les Misérables*
- Nodes represent characters, edge represent the co-occurrence of characters in the same chapter
- Directed or undirected?
- Could there be edge weights?

* Also available from http://www.casos.cs.cmu.edu/computational_tools/datasets/external/lesmis/

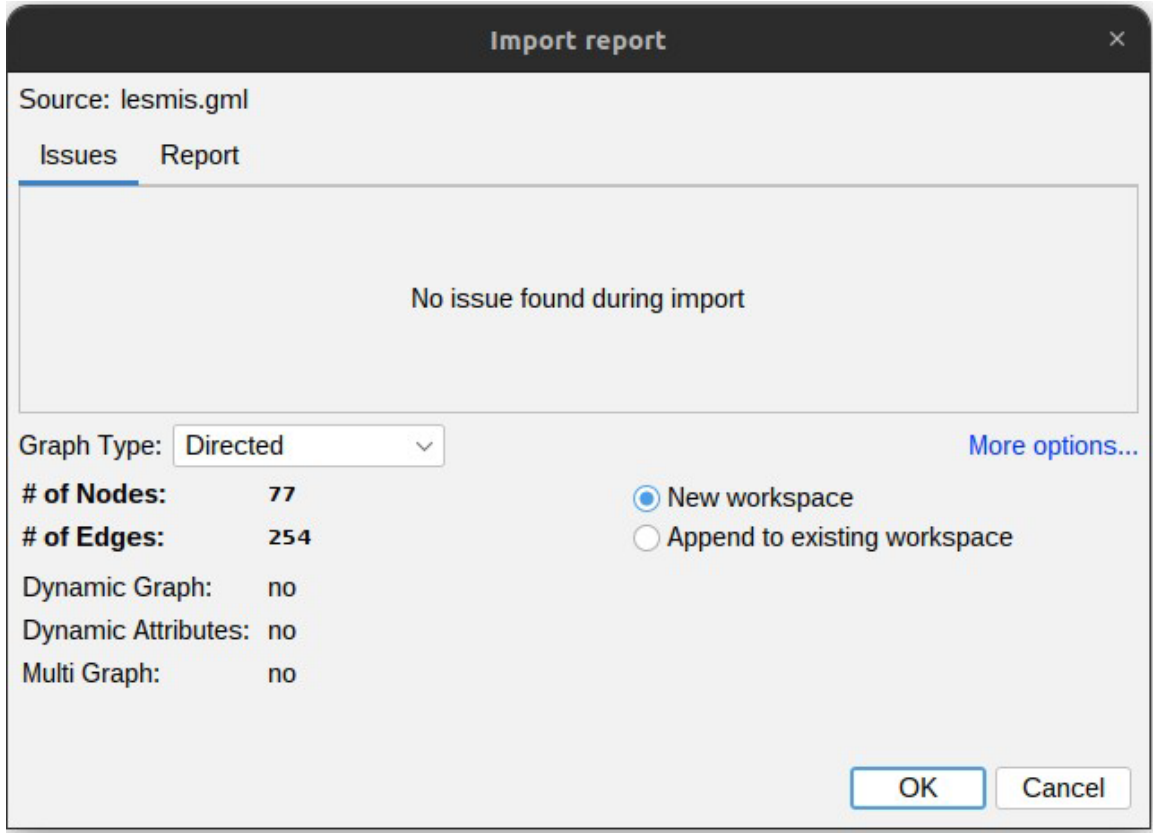
Step by step for creating a network visualization in Gephi

- Node color: partition (modularity)
- Node size: degree
- Edge width: co-occurrence
- What is modularity?
  - The novel was published in 5 volumes



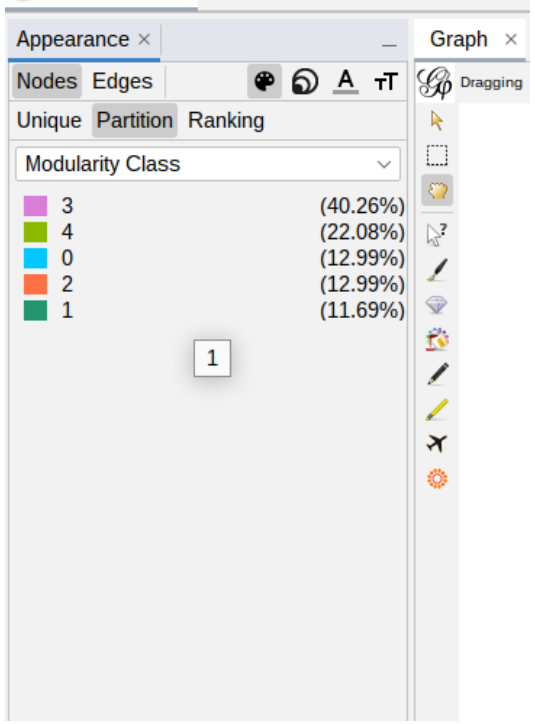
Download and open the network file

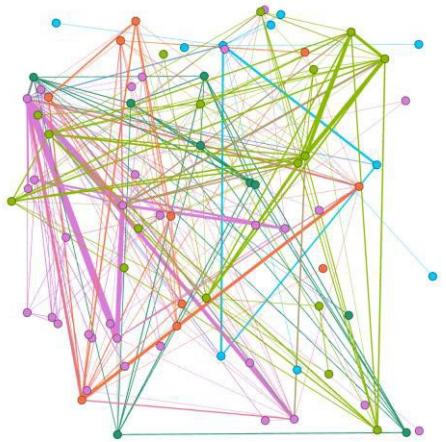
- Download the GML file from http://www.casos.cs.cmu.edu/computational_tools/datasets/external/lesmis/
- In Gephi, go to **File > Open** and find the downloaded file.
- Check the import report to see if inferred attributes are correct
- **Is it?**



Changing the color and size of nodes

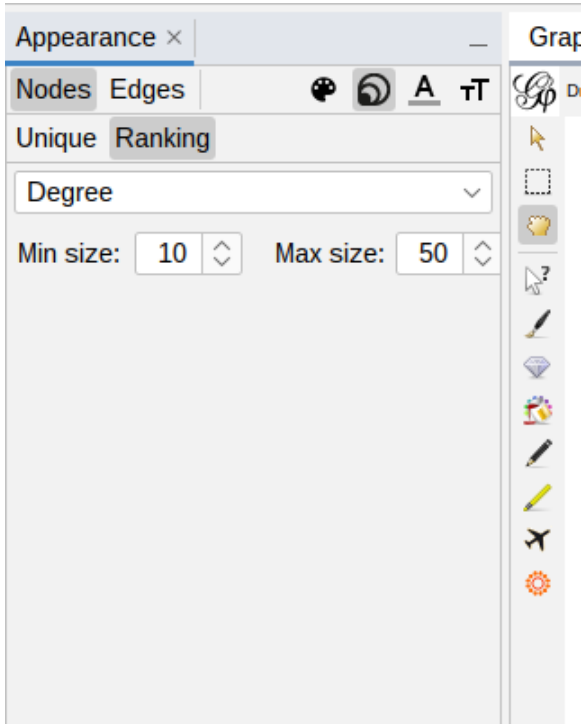
- First, we need to compute some attributes
- Go to the **Statistics** menu and run **Average Degree** and **Modularity** calculations
- In the menu, select the **Partition** tab and choose **Modularity Class** from the dropdown menu
- Click on the **Apply** button
- Why Partition?
  - Partition is normally used to distinguish between *categorical* variables, when the values are not in a scale

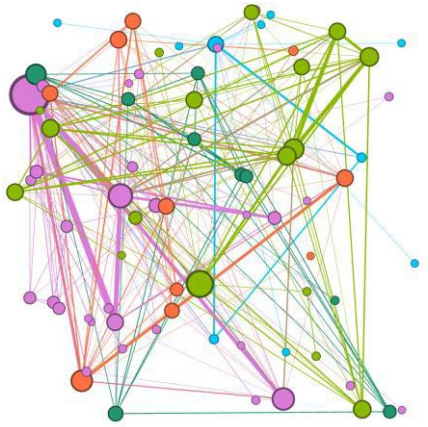




Changing the color and size of nodes

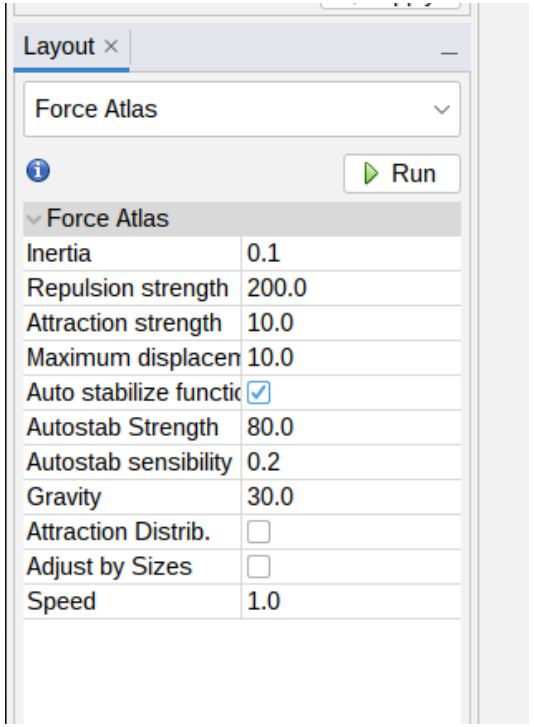
- Switch to the node size **Appearance** options
- Select the **Ranking** tab and choose **Degree** from the dropdown menu
- Set the Min size to 10 and Max size to 50
- Click on **Apply**
- Notice the change in the node size
- Why Ranking?
  - When values are not categorical and have an ordering, ranking can be used to represent node size

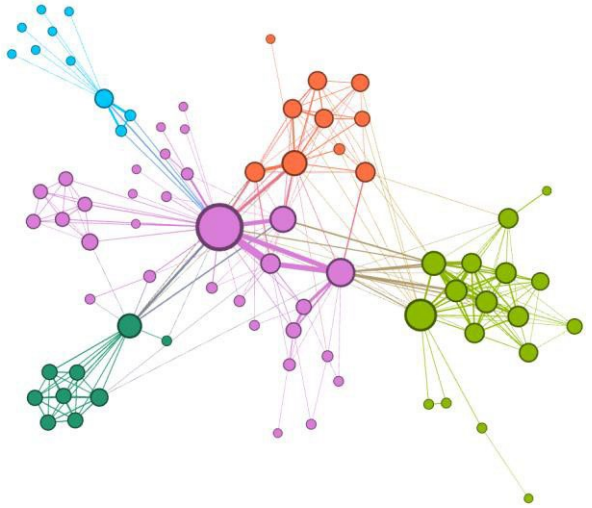




Applying a layout

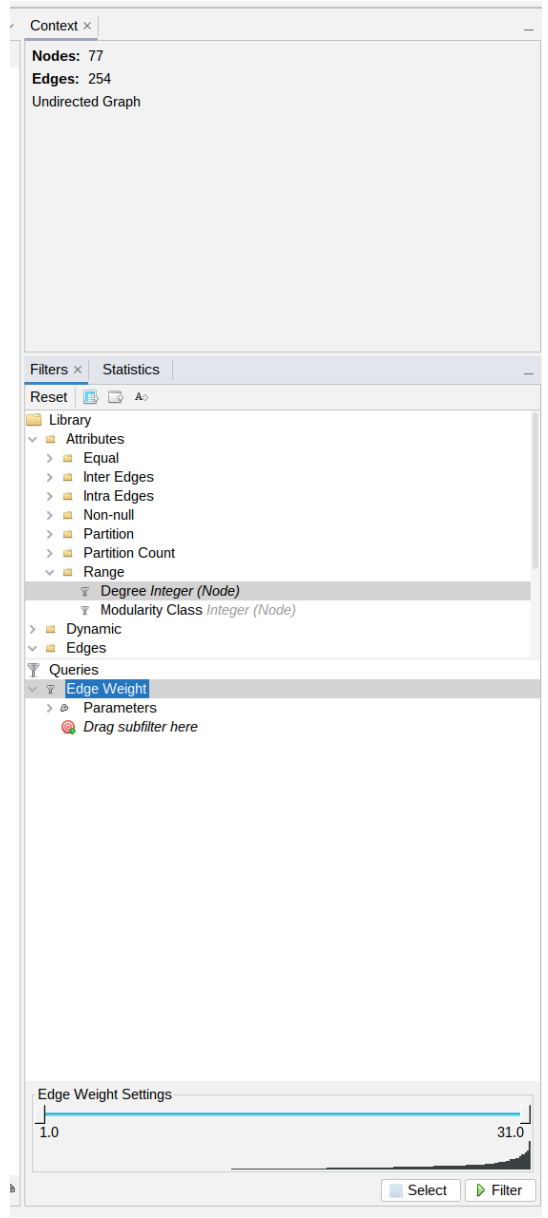
- Nodes do not inherently have a position in a 2D plot
- Layouts use node and edge attributes to produce more meaningful displays of the network
- In the **Layout** menu, select the **Force Atlas 2** algorithm from the dropdown list and click on **Run**
- Click on **Stop** to pause the algorithm as it may not converge to a solution
- Repeat it but now set the **Scaling** attribute to 100 and select the **Prevent Overlap** option

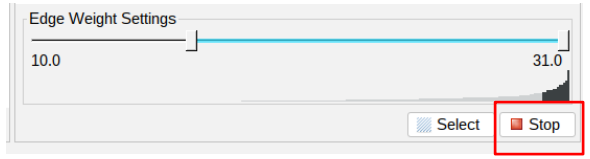




Filtering nodes and edges

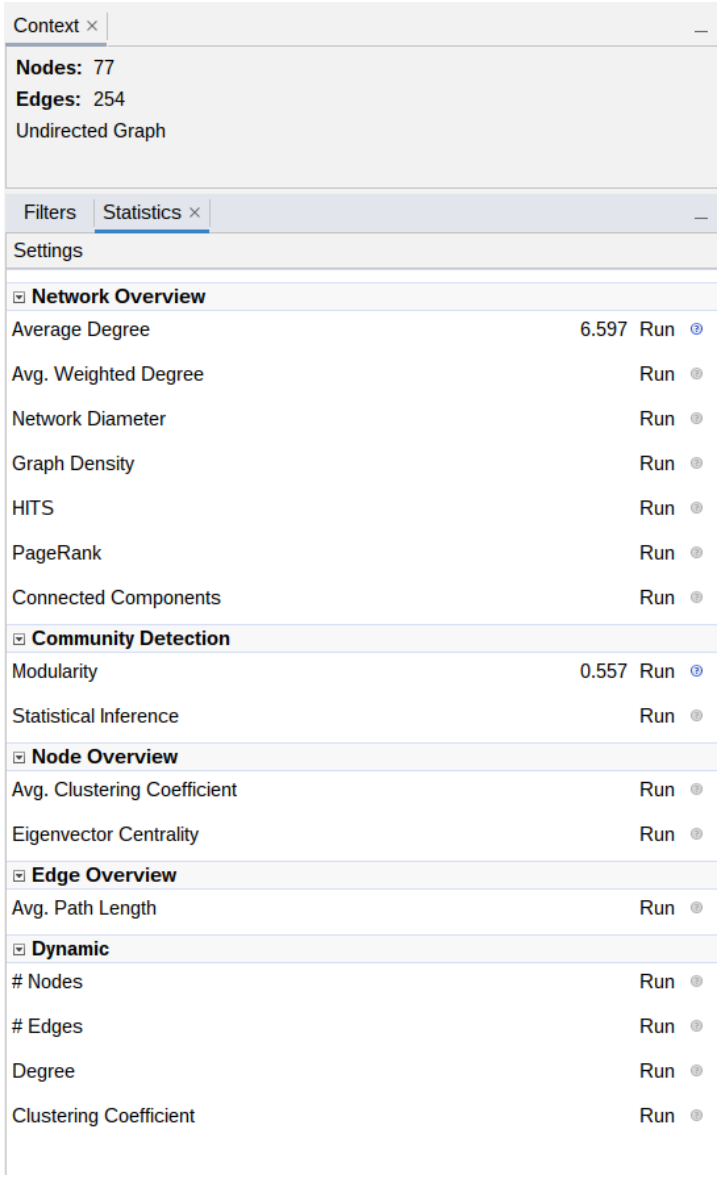
- Nodes and edges can be filtered based on their attributes to provide clearer visualization
- Examples of filter options are: node degree, edge weight, modularity class
- To create a filter, go to the Filters menu, select a filtering option and an attribute, for example, Edge Weight
- Adjust the filter to include the range of interest and click on the **Filter** button
- To stop the filter and revert back, click on **Stop**
- Multiple filters can be combined to create complex queries





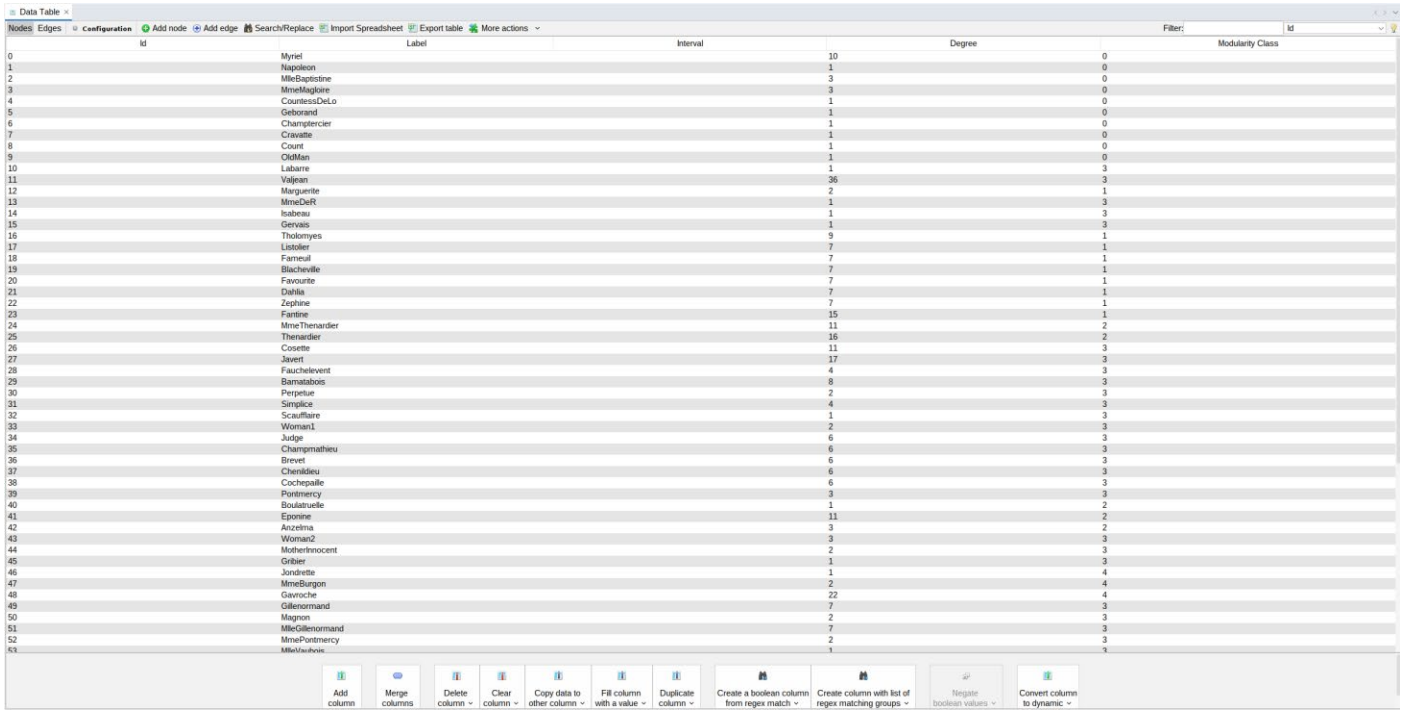
Network statistics

- **Degree** – the number of edges incident to a node (in/out degree available for directed graphs)
- **Betweenness centrality** – the number of shortest paths from all nodes to all other nodes that pass through that node
- **Closeness centrality** – sum of the shortest paths from a node to all other nodes
- **Eccentricity** – the greatest distance between a node and any other node in the network
- **Diameter** – the longest shortest path between any nodes
- Many other metrics are available, and even more can be loaded as plugins



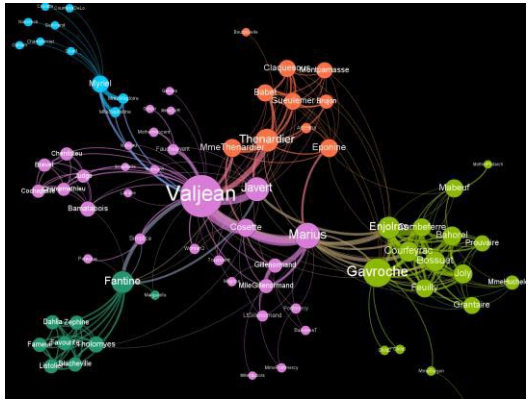
Data Table

- A table view of the data can be seen in the **Data Table** menu
- Data can be edited
- Nodes and edges can be created or deleted
- Can import data from spreadsheets

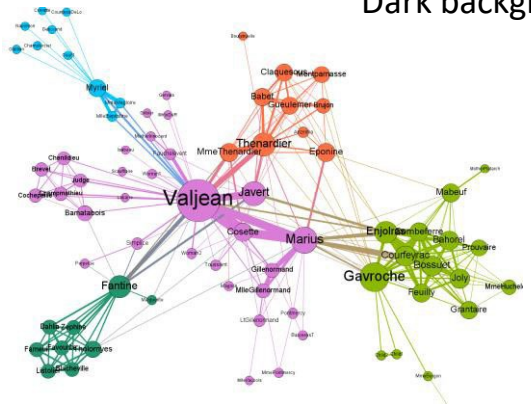


Rendering a graph to export

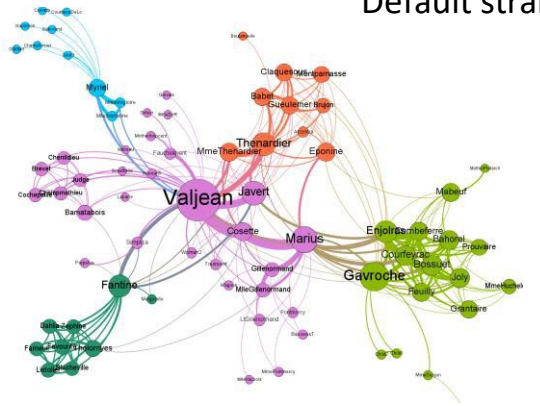
- The Preview menu allows you to do some final adjustments to the final product
- Default presets are available, custom presets can be created
- Possible adjustments: show node labels, draw borders, show edges, define edge shape (straight/curved), etc.
- Remember to click on **Refresh** to update the preview
- To export the figure go to **File > Export > SVG/PDF/PNG**...
- For publication, export vector images such as SVG and PDF



Dark background



Default straight

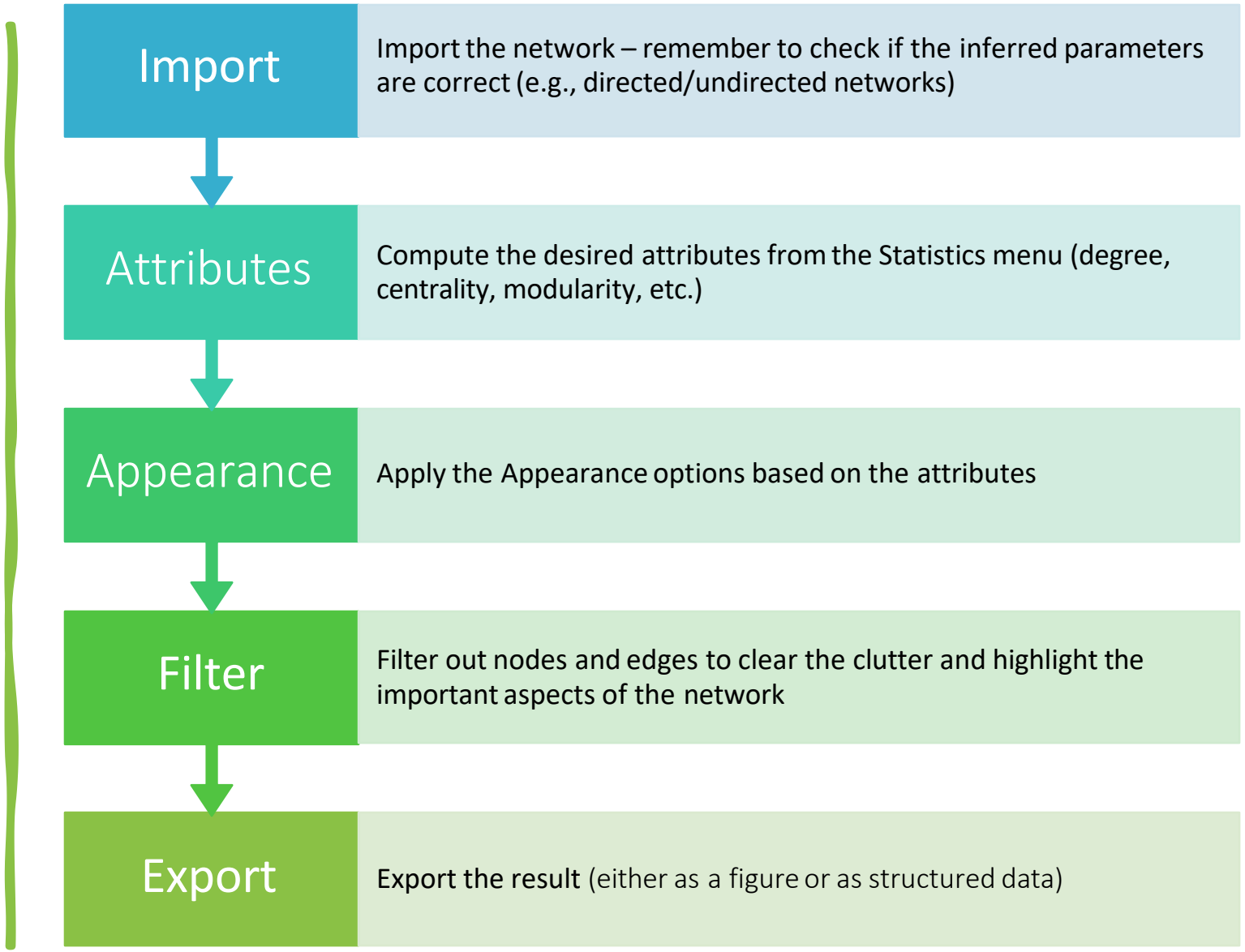


Default curved

Export Formats - Gephi

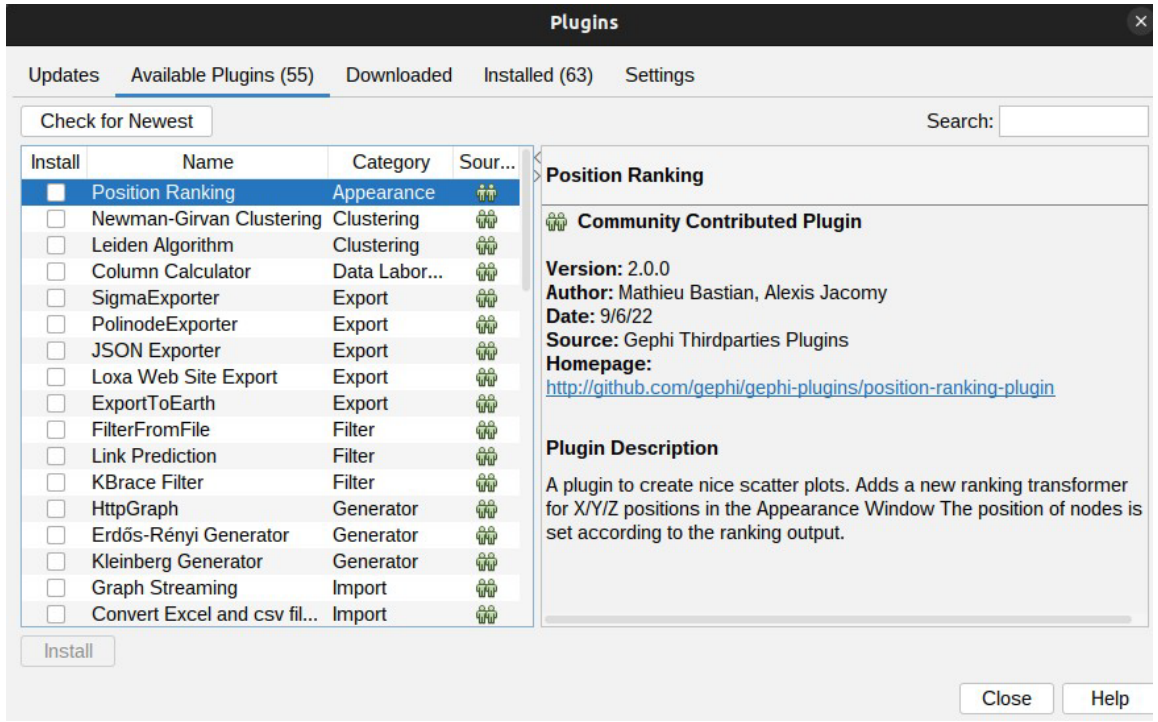
- PDF
- SVG
- PNG
- CSV
- GDF
- GEXF
- GML
- GraphML
- UCINET DL
- Pajek NET
- Netdraw VNA
- Spreadsheet

Gephi workflow



Plugins

- A number of plugins are available for Gephi
- These can be found and installed in Tools > Plugins



References

- Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. "Gephi: an open source software for exploring and manipulating networks." *ICWSM* 8 (2009): 361-362.
- Cherven, Ken. *Network graph analysis and visualization with Gephi*. Packt Publishing Ltd, 2013.
- Cherven, Ken. *Mastering Gephi network visualization*. Packt Publishing Ltd, 2015..
- Gephi additional tutorial slides are available at: https://www.slideshare.net/GraceBenefield/basics-gephi-tutorial
- More publications are listed at: https://gephi.org/users/publica
- Khokhar, Devangana. *Gephi Cookbook*. Packt Publishing Ltd, 2015
- Sample Gephi datasets are available at: https://github.com/gephi/gephi/wiki/Datasets

Exercises

Exercise 1

In this exercise you will:

- Review network statistics in Gephi
- Practice setting up network analysis using different metrics

Task: Compute three measures for each of the following networks:

- Dolphins Social Network in New Zealand: http://www-ersonal.umich.edu/~mejn/netdata/dolphins.zip
- Erdős Collaboration Network: http://vlado.fmf.uni-lj.si/pub/networks/data/

Network	# of nodes	# of edges	Density	# of triangles	Diameter
Dolphins	62	159			
Erdős	6,927	11,850			

Exercise 2

Using the following networks (five in total), compute the attributes

1. https://snap.stanford.edu/data/oregon1.html
2. https://snap.stanford.edu/data/oregon2.html
3. https://snap.stanford.edu/data/email-Enron.html

Network	# of nodes	# of edges	Average node degree (<k>)	# of triangles (Δ)	Diameter (D)
Autonomous systems - Oregon-1-010407[1]	10,729	21,999			
Erdős Collaboration Network	6,927	11,850			
Autonomous systems - Oregon-1-010331[1]	10,670	22,002			
Autonomous systems - Oregon-2-010331[2]	10,900	31,180			
Enron Giant Component[3]	33,696	180,811			

Exercise 3

Review the layout algorithm and visualization parameters for the following networks, then, create a publication quality vector image (SVG/PDF) for them

1. http://vlado.fmf.uni-lj.si/pub/networks/data/ucinet/ucidata.htm#zachary
2. http://www-personal.umich.edu/~mejn/netdata/
3. http://snap.stanford.edu/data/amazon0302.html

Network	# of nodes	# of edges
Zachary Karate Club[1]	34	78
American College Football[2]	115	613
Amazon product co-purchasing[3]	262,111	1,234,877