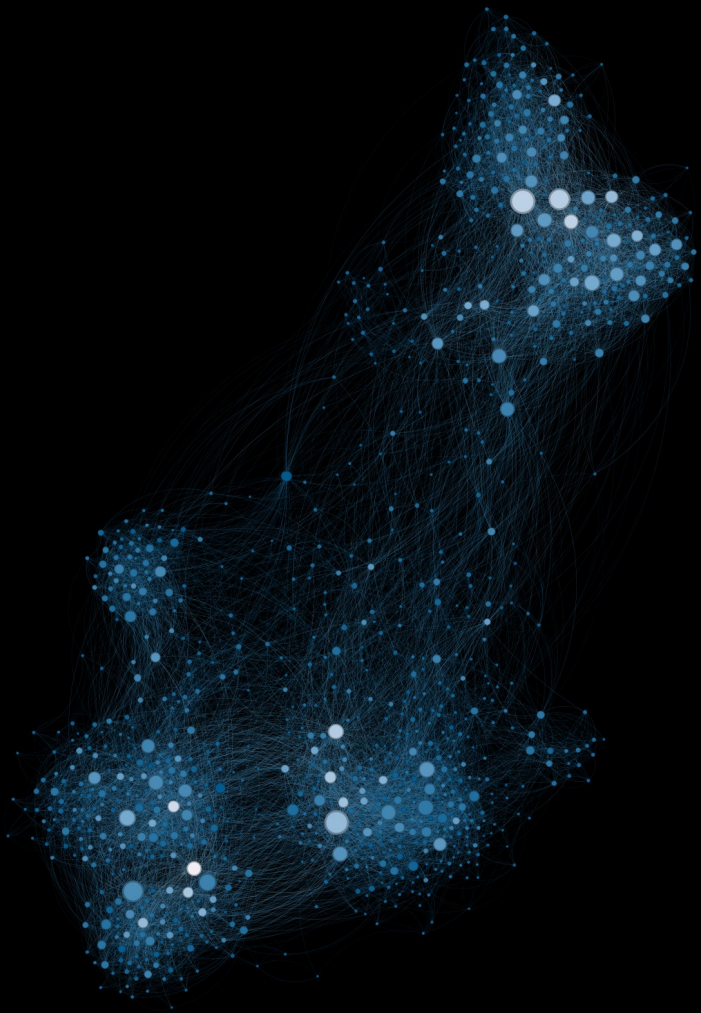
uOttawa

Introduction to Gephi

Jarno van der Kolk
University of Ottawa



Gephi

Visualize

Make complex networks understandable with a heavy focus on layouts of nodes and relations

Analyze statistics

Run statistical analyses with a single click and incorporate the results in visualizations

Extend

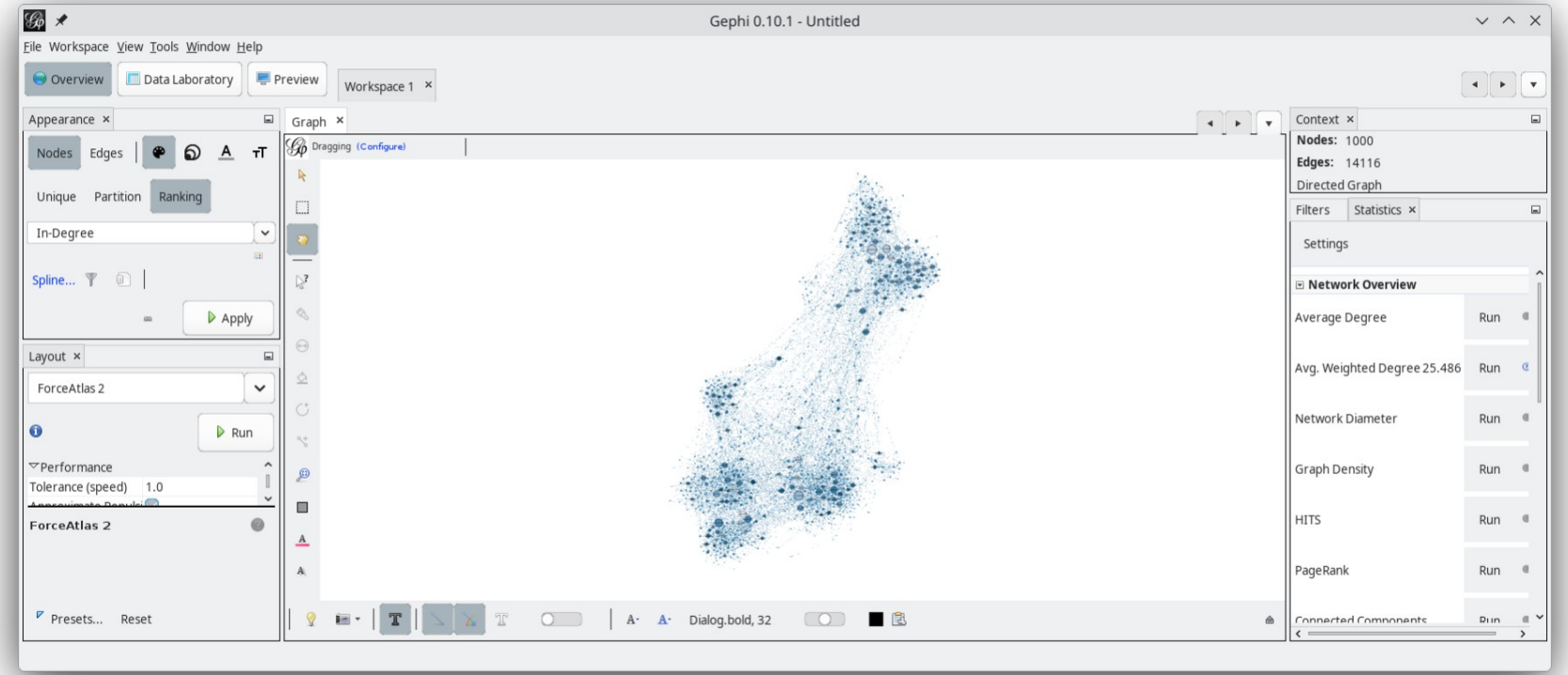
Free and open source as well as hundreds of available plugins

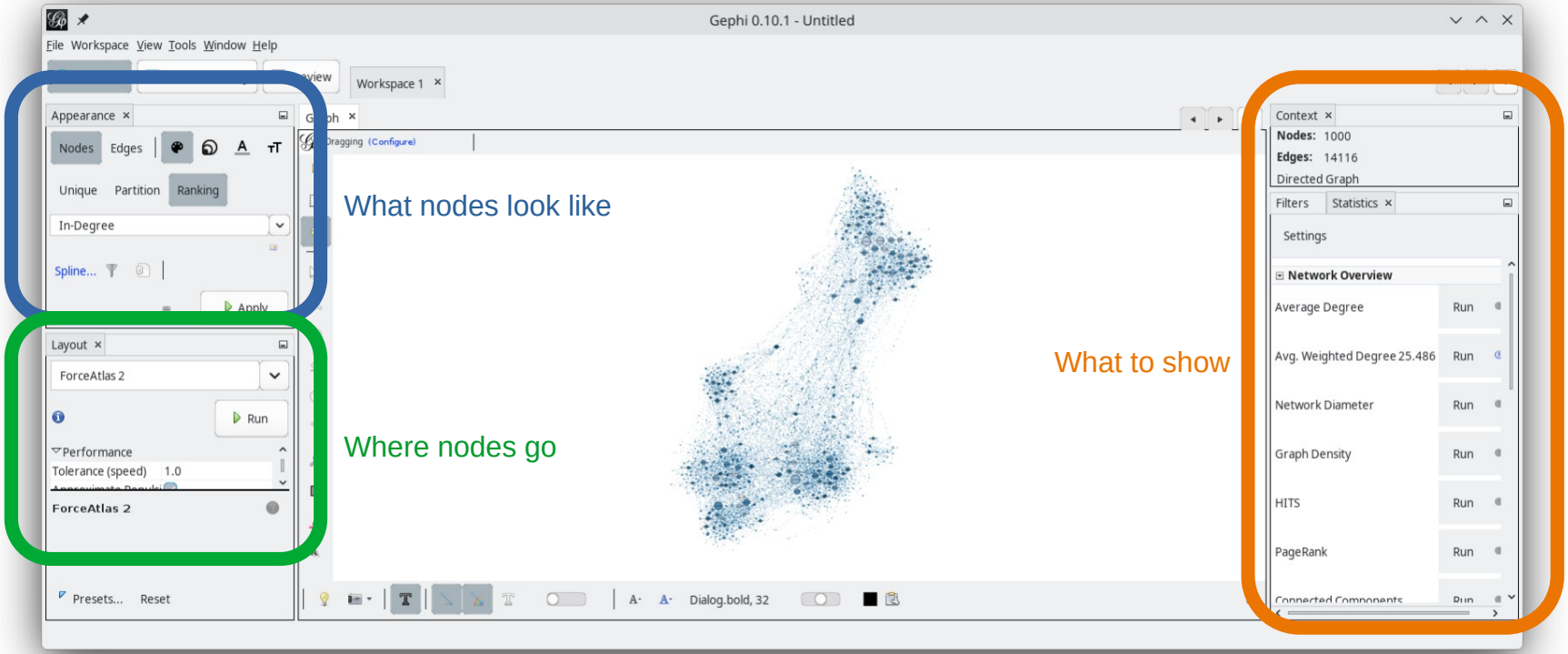
Examples

- Semantic linking of antonyms
  - https://allthingsgraphed.com/2015/04/09/a-matter-of-degrees/
- A computational network analysis of fake news on X
  - https://www.sciencedirect.com/science/article/pii/S2772503025000295

Interface

- Overview
  - Lay out your network and run statistical analyses
- Data Laboratory
  - Import and edit your data
- Preview
  - Export high-quality images


Gephi 0.10.1 - Untitled
File Workspace View Tools Window Help
Overview
Data Laboratory
Preview
Workspace 1
Appearance
Nodes
Edges
Unique
Partition
Ranking
In-Degree
Spline...
Apply
Layout
ForceAtlas 2
Run
Performance
Tolerance (speed) 1.0
ForceAtlas 2
Presets...
Reset
Graph
Dragging (Configure)
Dialog.bold, 32
Context
Nodes: 1000
Edges: 14116
Directed Graph
Filters
Statistics
Settings
Network Overview
Average Degree
Run
Avg. Weighted Degree 25.486
Network Diameter
Graph Density
HITS
PageRank
Connected Components


Gephi 0.10.1 - Untitled
File Workspace View Tools Window Help
Workspace 1
Appearance
Nodes
Edges
Unique
Partition
Ranking
In-Degree
Spline...
Apply
Layout
ForceAtlas 2
Run
Performance
Tolerance (speed)
1.0
ForceAtlas 2
Presets...
Reset
What nodes look like
Where nodes go
What to show
Context
Nodes: 1000
Edges: 14116
Directed Graph
Filters
Statistics
Settings
Network Overview
Average Degree
Run
Avg. Weighted Degree 25.486
Network Diameter
Graph Density
HITS
PageRank
Dialog.bold, 32

Loading data

- Multiple formats supported
  - GEXF (native)
  - GraphML
  - CSV
- Contains
  - Nodes
  - Edges (i.e. links)

https://codeberg.org/uOttawa-IT-Research/gephi-workshop

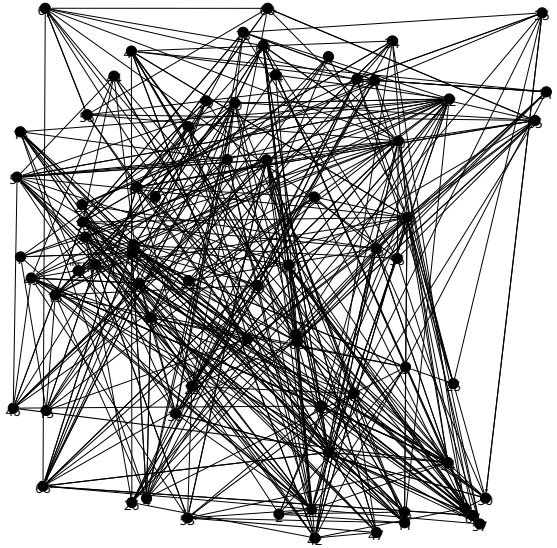
Loading data

- High school friendhips (**`hs.dl`**)
  - In the fall of 1957 boys in a small high school in Illinois were asked. "What fellows here in school do you go around with most often?"
- In Gephi: File → Open...
- Coleman, J. S. Introduction to Mathematical Sociology. New York: Free Press, pp.450-451.

Source: http://moreno.ss.uci.edu/data.html#high

Layouts

- Initially all nodes are placed randomly
- Bit of a mess!
- Layouts let you organize the nodes

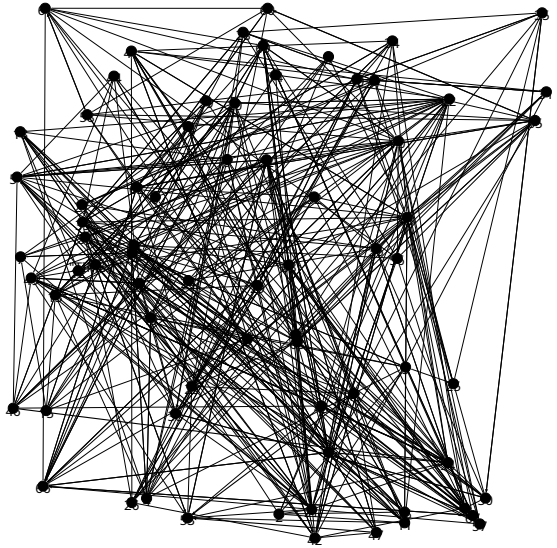


Layouts

- Reorders points based on different algorithms
  - Force atlas (2): All points repels but bonds are stiff
  - Fruchterman-Reingold: Attempts layout in disk with bonds the same length
  - Expand/Contract: moves all points further/closer compared to centre
- Often running one after the other is helpful

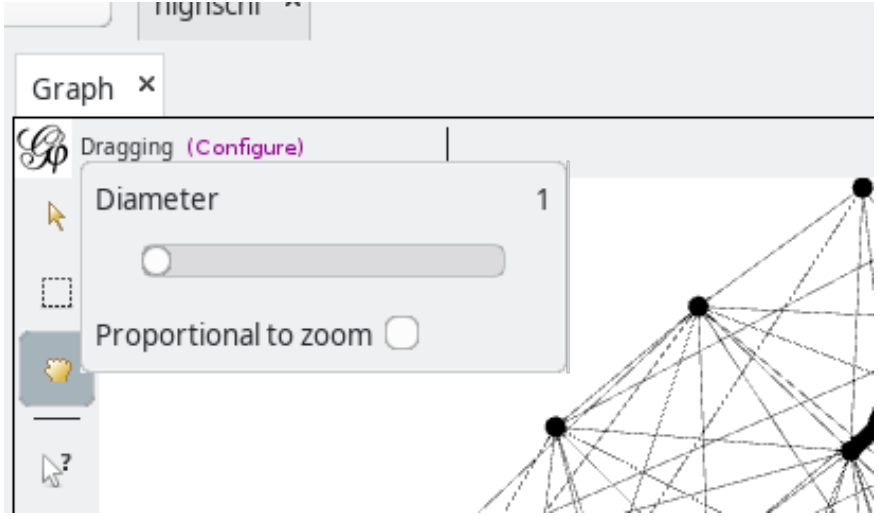
Layouts

- Try them out
- Also play with the parameters



Layouts

- Positions can be further tweaked
- Make sure drag icon is selected
- Press Configure to increase drag radius

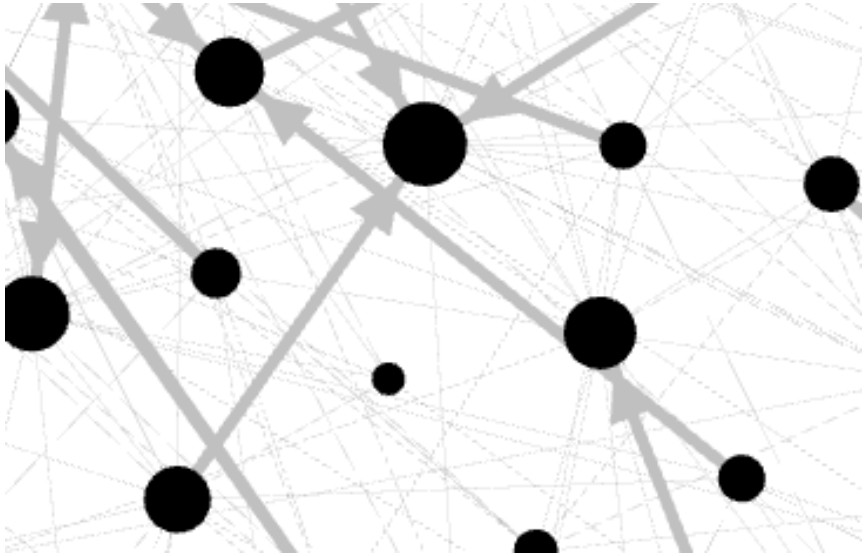


Styling

- The nodes and edges can be styles to convey more information
- Colour & size for nodes
- Colour for edges (size is governed by edge weight)

Styling

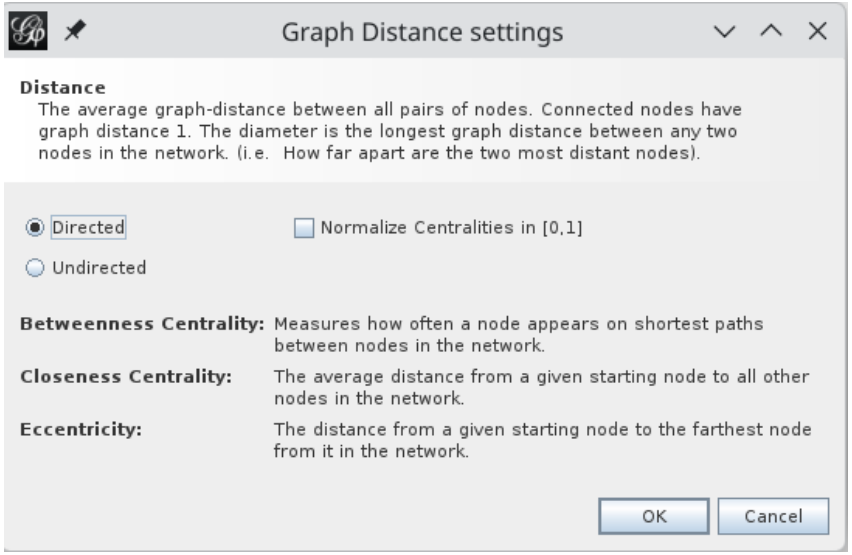
- Nodes
  - Size
    - Ranking
      - Degree
      - Min: 10
      - Max: 40



Statistics

- Not a lot of properties to choose from for styling
- However the built-in statistics lets us calculate more
- Even more statistics can be added through plugins

Average path length

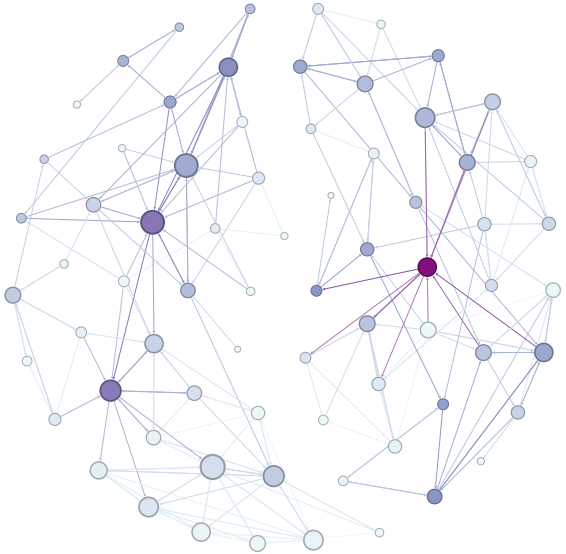


Betweenness Centrality

- Measures how often a node appears on shortest paths between nodes in the network
- Nodes with high “betweenness” are middle man for other nodes

Betweenness Centrality

- New properties available now
- Colour by betweenness



Modularity

- Under Community Detection, run the Modularity statistics
- Detects groups of nodes that form friend groups

Modularity

Other statistics

- Documentation isn’t great...
- But
  - All filters have a short description when you click on them
  - Report show references to the papers

Data Laboratory

- Spreadsheet-like interface
- Tabs for Nodes and Edges
- The “Label” column can be use to label nodes and edges
  - Merge Column can be used to create labels combining other columns

Preview

- Final image generation
- Choice of many presets
- Styling of
  - Nodes and labels
  - Edges and labels
  - Background colour
- Export to image

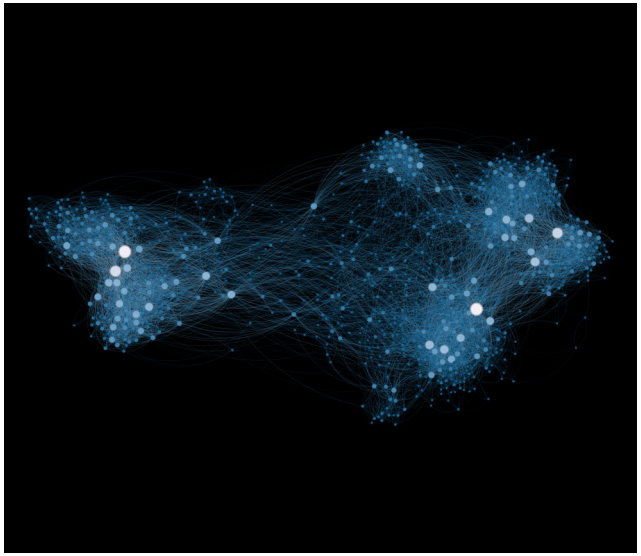
https://codeberg.org/uOttawa-IT-Research/gephi-workshop

Mapping letters over Europe

- Based on https://www.martingrandjean.ch/gephi-introduction/
- Load Nodes1.csv and Edges1.csv
  - **Important:** Append to existing workspace for Edges1.csv!
- Layout: Force Atlas 2 with scaling = 5.0
- Colour node by Degree, inverted colour scale
- Size node by Degree, min = 5, max = 30

Mapping letters over Europe

- Preview
  - Default Curved
  - Show Labels = Off
  - Edges, Opacity = 20.0
  - Background = black



Collaboration networks

- Requires more advance import
- Tools → Plugins → Available Plugins → “Convert Excel and csv files to networks”
- New entry in File → Import... appears!

Collaboration networks

- Data generated from Web of Science keyword search
- Export → Tab delimited file
- Can only export 1000 records at a time so I had to employ some scripting to automate it (outside of scope of workshop)

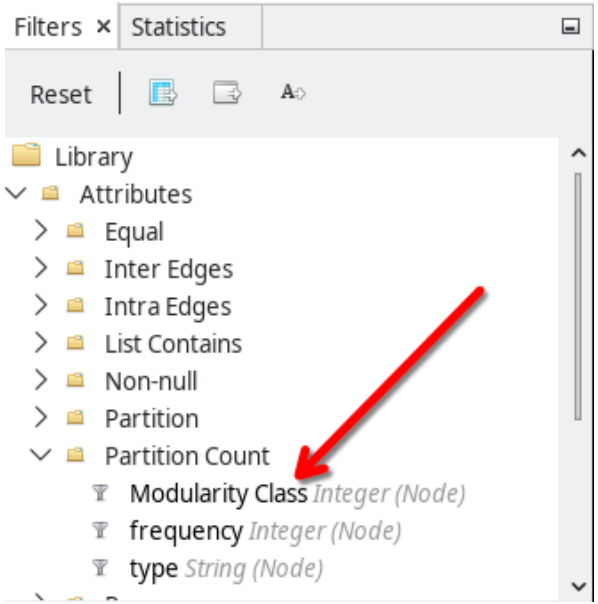
https://codeberg.org/uOttawa-IT-Research/gephi-workshop

Collaboration networks

- File → Import…
- Select "bundled_symplectic_geometry.tsv"
- Field delimiter: tab
- AU is connected to AU (AU=author)
- Subfields delimiter is semicolon
- Ignore dynamic
- Links between agents and remove self-loops

Collaboration networks

- 4,735 nodes (unique authors)
- 6,093 edges (people on same paper)
- Too many, so filter by Modularity Class
  - set minimum to 80



Collaboration networks

- Play with the layouts
- Try to find a layout that shows how there are distinct groups of researchers but that also shows how they work together a little bit

Not covered

- Gephi also support dynamic networks, but you need data with time intervals. No easy way to export animations anyway.

Python

- Gephi does have limitation in how big the input data can be.
- For really large datasets, it might be better to use the networkx package in Python.

QUESTIONS?