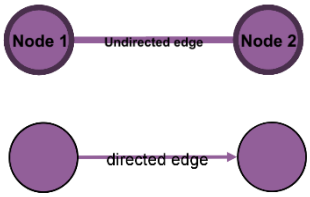
Exploring networks with Gephi

Vocabulary



Nodes (vertices).

Edges (ties, links).

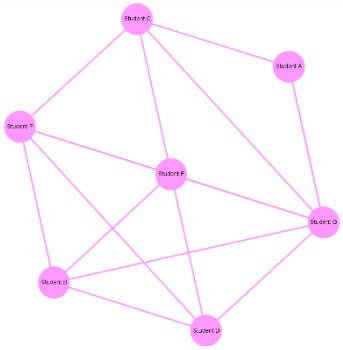
Directed and **undirected** networks.

In the exercise session, you will work with an undirected network.

Two-mode networks (affiliation/bipartite network) and **one-mode networks**.

Two types of nodes

One type of nodes

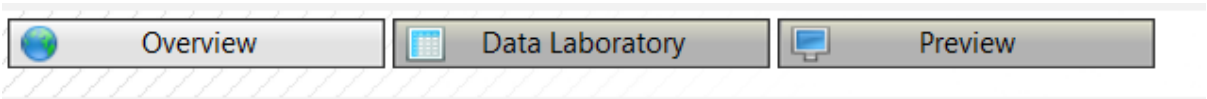


Degree. A node's number of edges. In a directed network, number of ingoing edges is *in-degree* and outgoing edges *out-degree*.

Gephi

1. Open graph file (.gexf) in new workspace. (File → Open…)

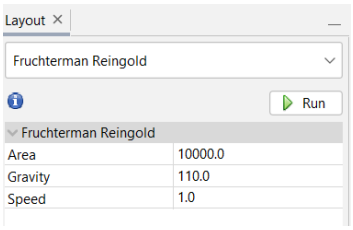
Workspace. A workspace consists of three tabs:



- ***Overview.*** In the overview, you can do most operations: changing color and size of nodes, Running *layout algorithms,* running statistics (e.g. *degree*, *modularity*), apply statistics as a filter, change network aesthetics (labels, size of labels for edges and nodes).
- ***Data Laboratory.*** In the data laboratory, you have access to the *node list* and the *edge list.* The *node list* is a collection of all nodes in the network together with their attributes. If you run a statistic, a new column will be created containing the result of the calculation. You can add a column yourself, edit color, label color etc. for nodes you select. The *edge list* contains information on the source and the target of a link connecting two nodes. An edge list and node list is all you need for a network. Here, you can import and export edge list and node list spreadsheets.

- ***Preview.*** The preview is used to export the network. Some settings you make in the Overview will not necessarily have an effect for the Preview, e.g. label size. You need to click on "Refresh" before you see you network. Click on refresh every time you change something in the Preview settings. Click on "SVG/PDF/PNG" to export your network.

2. Run a layout algorithm



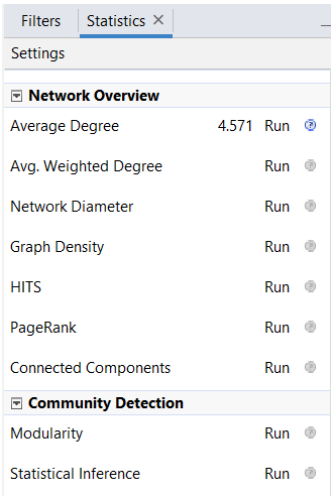
Layout algorithms determine the position of nodes in relation to each other weighting *attraction* and *repulsion* according to the nodes' ties and the ties' weights. *There is no true or false representation of the network, but a more or less useful one to identify patterns (visual analysis) or tell a story (visualization).* Experiment with layout algorithms.

Those can be recommended: **ForceAtlas2** and **Fruchterman Reingold**. Especially with large networks: you don't have to wait until an algorithm finishes the calculation – this can take a long time and does often not change the visualization noticeably. After you click "run", click "stop" if you are satisfied with the visualization or if you see that not much is changing anymore. You are working with a small network, this probably takes less than a second. **Label Adjust** algorithm will adapt the position of overlapping labels so that all labels are readable.

> *You cannot find your nodes after running an algorithm? Use the magnifier to recenter.* 

Fine-tune the layout algorithm: Nodes are too close to each other or too far apart, experiment with the forces. Increase *scaling* and/or reduce *gravity* if they are too close. You can also *dissuade hubs* and *prevent overlap* with ForceAtlas2.

3. Run statistics



On the right hand in the Gephi workspace you can run statistics. Running ***average degree*** calculates the average degree of the network.

What is the average degree of your network? What does that mean in your case?

Apart from calculating the average degree, this action created a new column in your *Data Laboratory.*

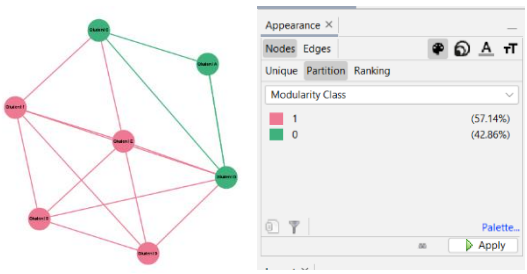
Go to the Data Laboratory and check out the new column. Which are the most connected students? Sort by degree. Right click on a single student and "Select on Overview". Go back to Overview and you will see that you have identified in the context.

The ***modularity*** statistics identifies communities in you network. Changing parameters will create more or less communities.

Running statistics adds attributes to nodes that you can further use when changing the color, size or label of nodes.

> ***Filters***: *If you want to apply a filter, you need to run a statistic first. For instance, after running degree statistics, you can apply a filter and only keep nodes above or below a certain degree threshold.*

4. Change appearance



Changing the appearance of nodes can enable you to see patterns in larger networks which are otherwise hard to identify. In Gephi you have various options to adapt the appearance of the size and color of your nodes, edges, and labels.

If you have run a degree or modularity statistics before, you can now use the results to color your nodes. A common way is to display the node size according to degree. It can make sense to use other node features when changing the appearance of nodes.

5. Describe and interpret your network

What are the ***regions*** in the network with a concentration of actors?

Are there ***centers*** in the network, i.e. star-like actors that are much more connected than others?

How ***dense*** is the network? Are all the nodes connected or does the network have separate ***components***?

Can you identify ***interfaces***? This is, areas where different regions of the network are connected by ***boundary actors?***

> *Comment: The notions used here are those of visual network analysis (VNA)[1], a qualitative approach for exploring networks. VNA's uses network analysis as a method to understand the relations of actors and the relational formation. In contrast, social network analysis (SNA) and graph theory (GT) focus on the analysis of the network as such. Don't be surprised finding other concepts used for related ideas in SNA and GT.*

5. Additional Info

Clicking on the somehow hidden button on the right hand, you can make changes to the appearance of colors and labels in the Overview mode. It can be that these changes are limited to the *Overview* mode and will not be effective in the *Preview*. Try it out.

Short introduction to Gephi for new MSc Social Data Science students.
Simon Ullrich (simon.ullrich@sodas.ku.dk). 8/31/2020.

[1] Decuypere, M. (2020). Visual Network Analysis: A qualitative method for researching sociomaterial practice. Qualitative Research, 20(1), 73–90. https://doi.org/10.1177/1468794118816613