Networks

Diving into networks

Gephi has the reputation of being a notoriously tricky and difficult program. But it is one of the most powerful network analysis tools and requires zero coding skills, so it is a really good idea to try to wrap our heads around it.

First of all, you need network data. Let's stop for a second and consider what networks actually are. You might remember from maths that it is all about a combination of points (aka nodes) and lines (aka edges), which describe relations.
Some of the more popular networks are the so-called social networks, (not to be mistaken with Twitter and Insta), that is: networks we create through our relations with other people.

Network theory is a huge field with a lot of studies, one good introduction is this book: https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199206650.001.0001/acprof-9780199206650 or an amazing DHSI course on Networks.

Obtaining stylometric network data

Stylometric networks are an outcome of "Bootstrap consensus" classification. The idea behind it is quite simple - stylo() runs a number of cluster analysis classifications and then combines its results to show only such connections that were regularly, above the chance level, pointed out as strong.

To put it simply, think of it this way: if we run the analysis for the range of 100-1000 MFW (most frequent words) with the iteration of 100 (so adding 100 at each new step, getting the analysis for 100, 200, 300 etc. features), we get 10 different classification scenarios - and quite possibly some texts change their nearest neighbors. The bootstrapping procedure involves setting a threshold for agreement in classification - by default 50% - which means that for the final outcome a given pair of texts need to be each other's nearest neighbors at least in 50% of the scenarios (so e.g. for the range of 100-600).

This approach is one of the ways of ensuring that we do not rely on a random one-off classification result, but rather analyze the connections that were found consistently and independently of the classification settings.

https://gephi.org/users/download/

Using Gephi

Basic start:

1. To create a network first you need a consensus tree, or more specifically - a csv table such analysis produces, one that we created would be: german_Consensus_*100-1000_MFWs_Culled_0__Classic Delta_C_0.5_EDGES.csv*.

2. Once you have it, select the **Data Laboratory** tab in Gephi.

3. In the ***Data laboratory*** tab:

   a. ***Data laboratory → Import Spreadsheet → Choose a csv file*** - you don't need to change anything here, so just keep pressing next until it finishes
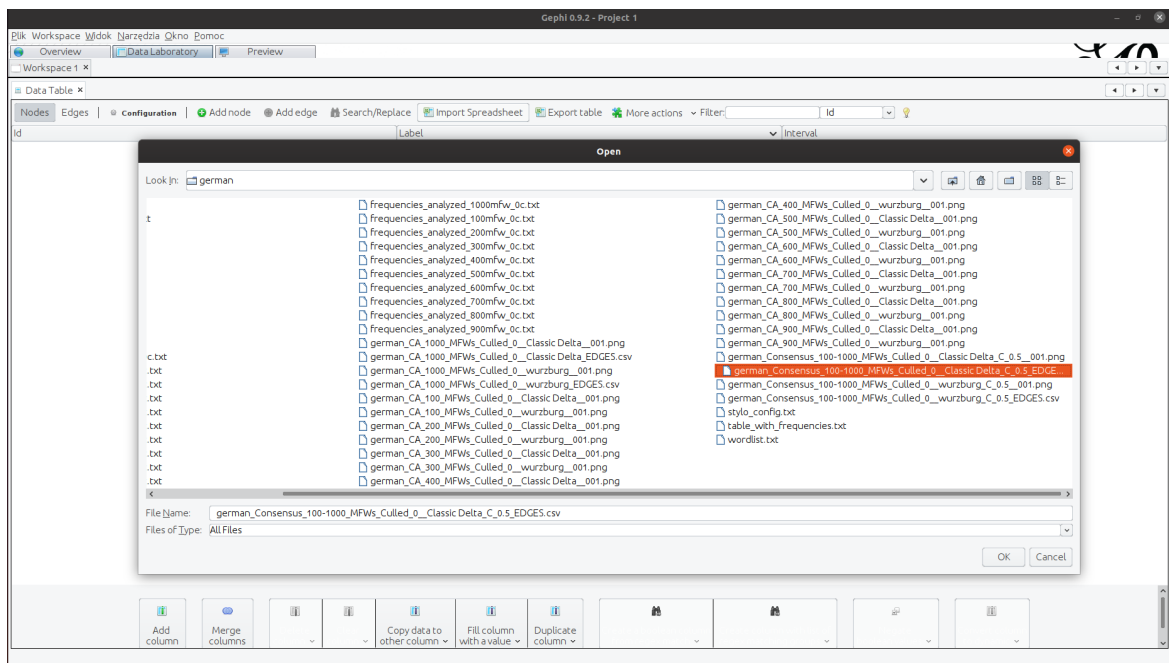
and gives you something like this:
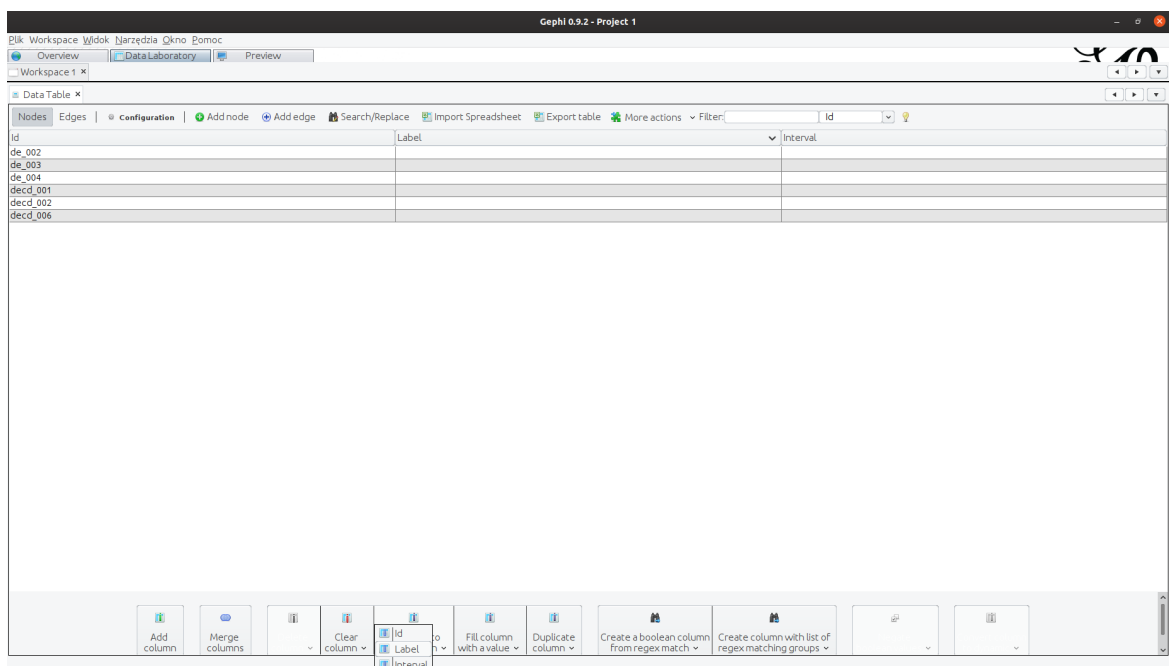
4. On the bottom of the window,

   a. select **Copy data to another column → Id** and select **Label**.
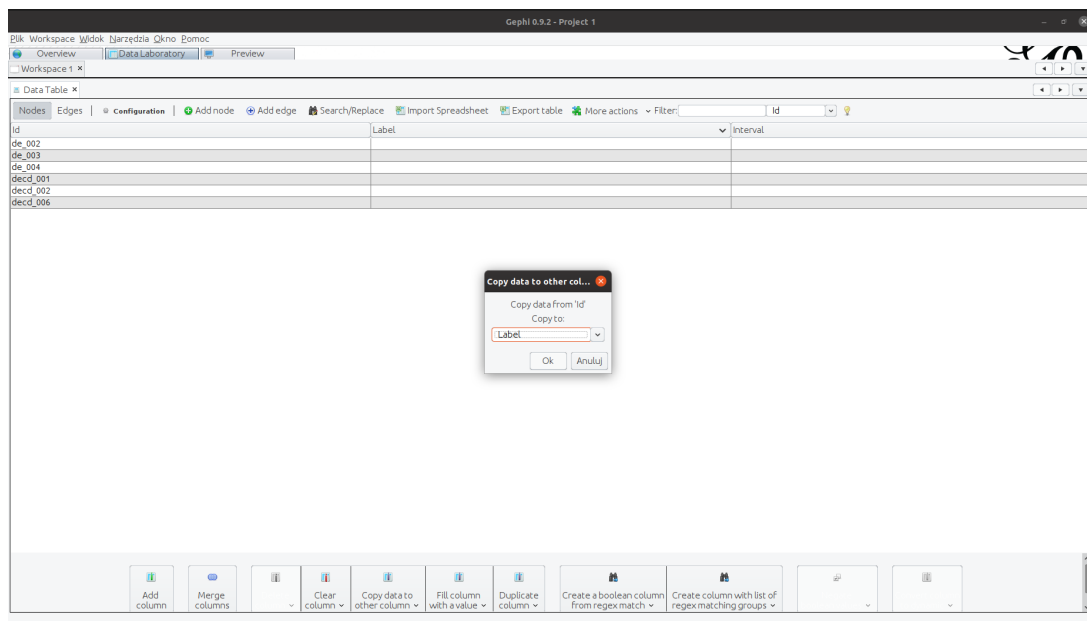
b. ***Create column with list of regex matching groups → ID***. Add **Title** of your choice, e.g. Author (or in my case TranslatorDiff), in **Regular Expression** enter the regular expression limiting what content should be extracted from the ID name, e.g. to extract just the content before the first underscore: ^[A-Za-z]+
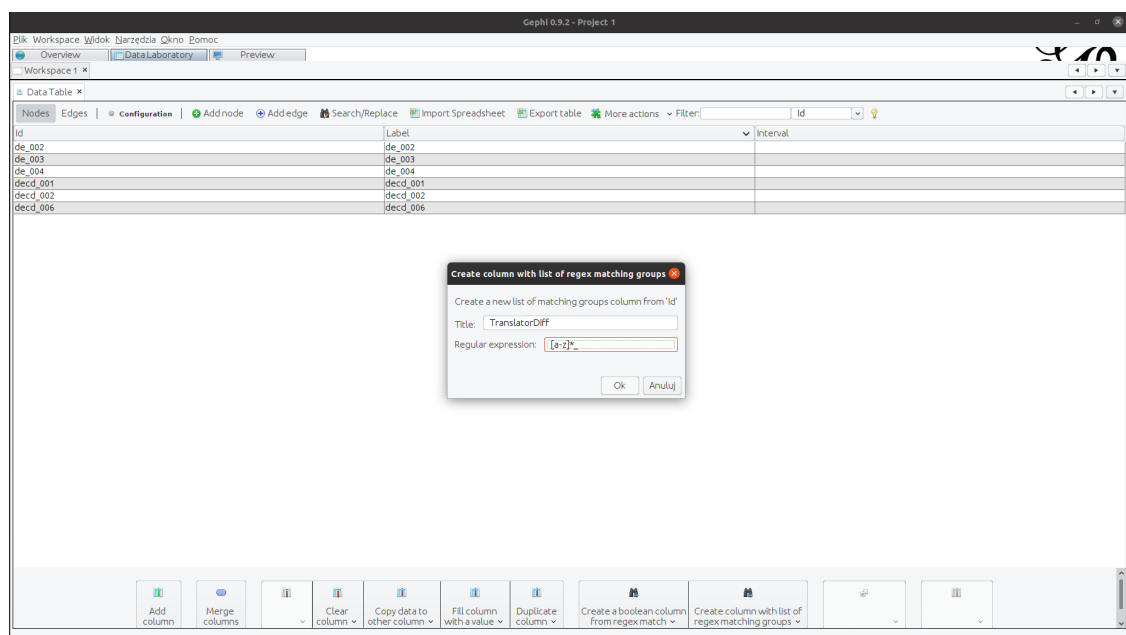

5. Now, go to ***Overview***,

a. In the menu ***Appearance*** on the left select ***Nodes→Partition*** and on the rolling list find the name you gave your column in step 2, e.g. Author, click on ***Apply***. Nodes should now have some colors.
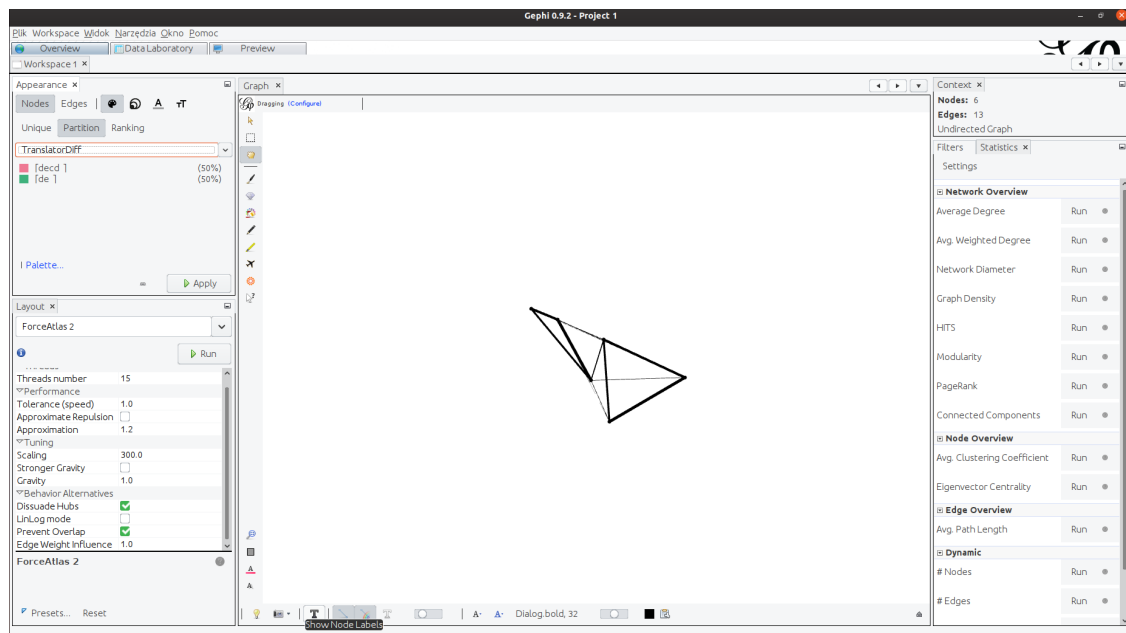
In the graph field you can change some display settings, it's a good idea to add IDs to nodes - in the bottom line select **Show Node Labels** (3rd from the left) and in **Attributes** (first from the right) **ID** rather than **Label**.

b. On the bottom left, in the **Layout** menu, choose **Force Atlas 2.**
for better readability of the network, set
Scaling to 300, and turn **Dissuade Hubs** and **Prevent Overlap** on. Sometimes, if the nodes are very far / you want to force gephi to make one network instead of a few, it makes sense to switch on **Stronger Gravity**. Select **Run**, and stop it when the network doesn't seem to change much anymore.
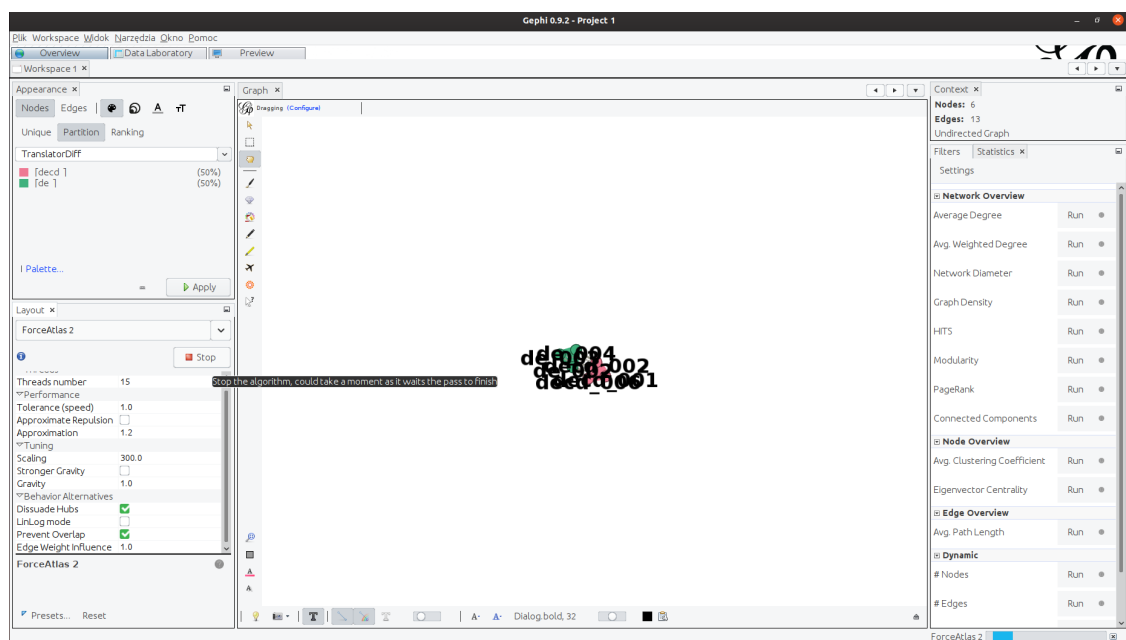
c. In the **Layout** menu, choose **Expansion**, run it once or a few times to stretch the graph and make it more readable.
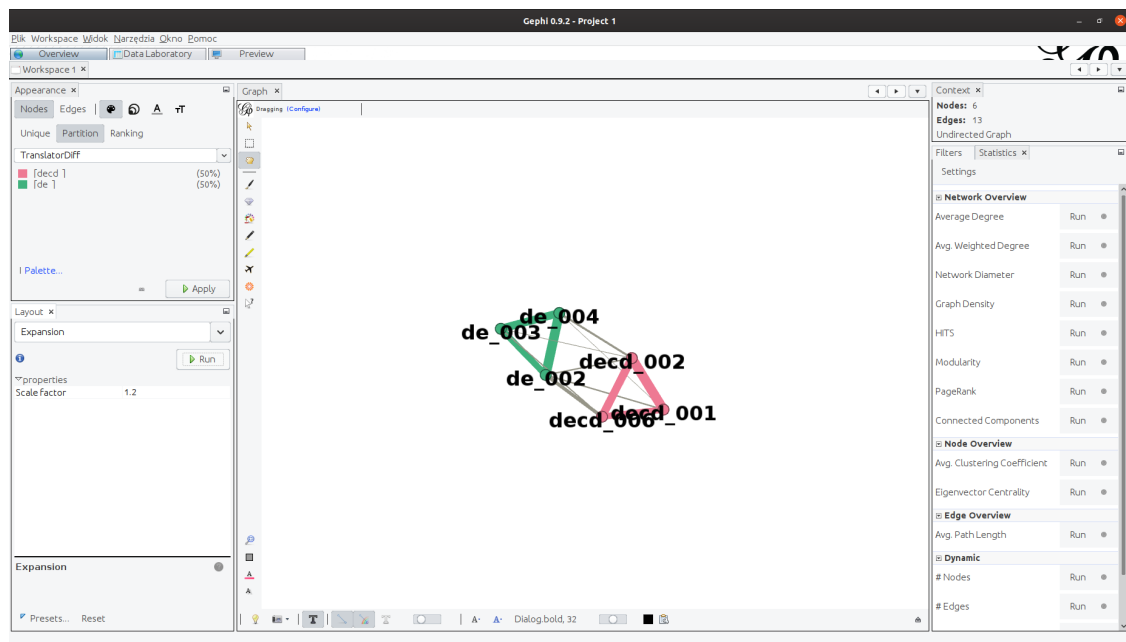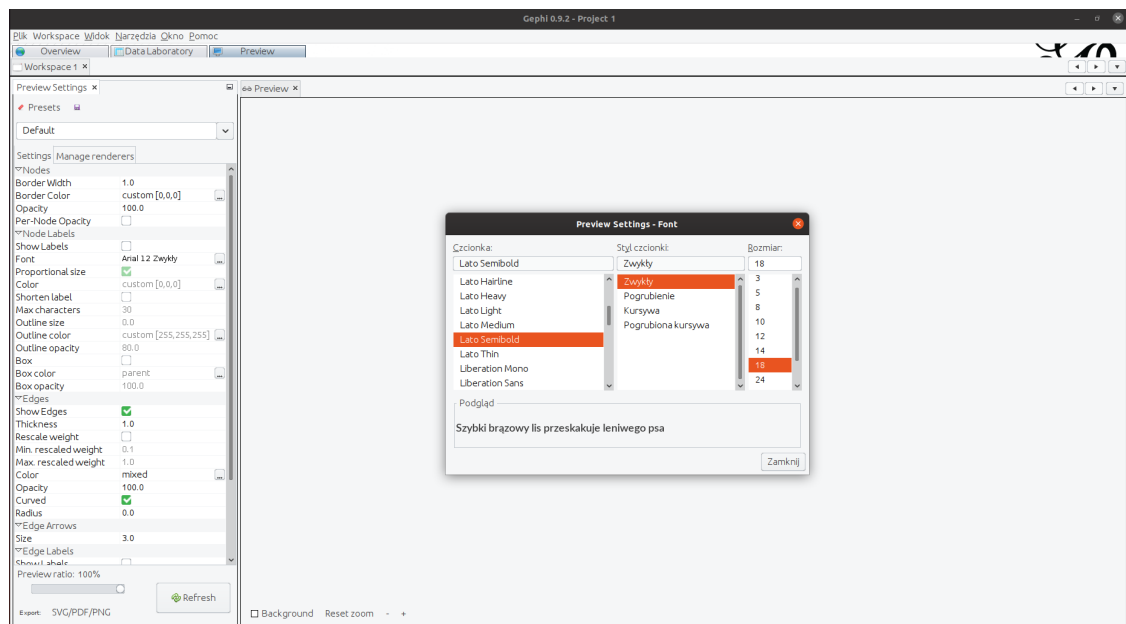
d. Again in the **Layout** menu, choose **Label Adjust**, run it once so that names of the nodes don't overlap.
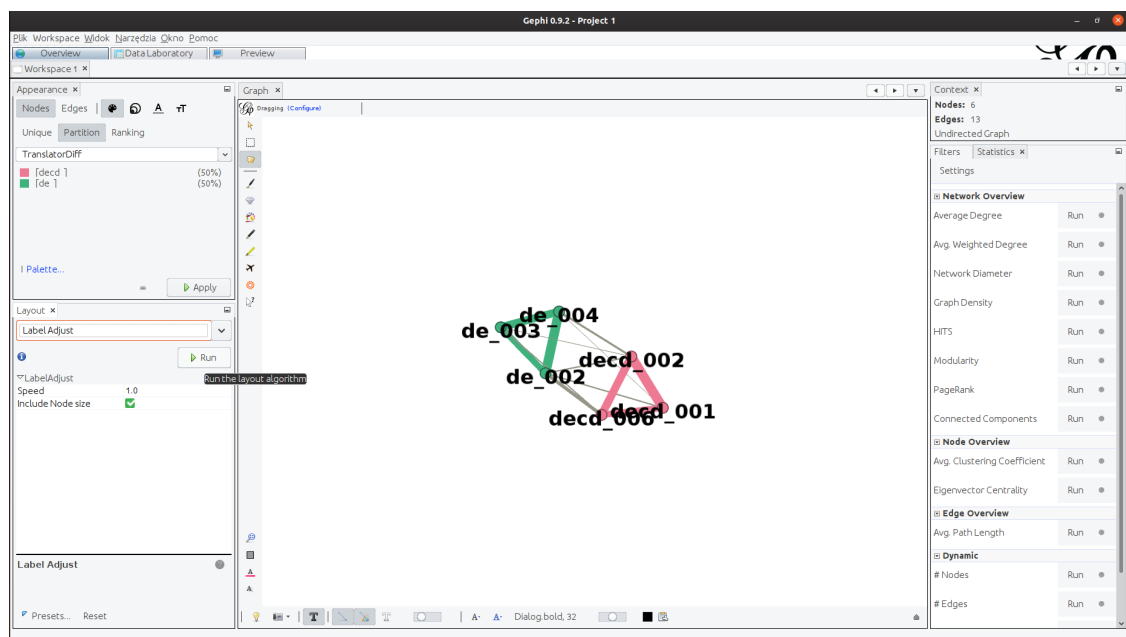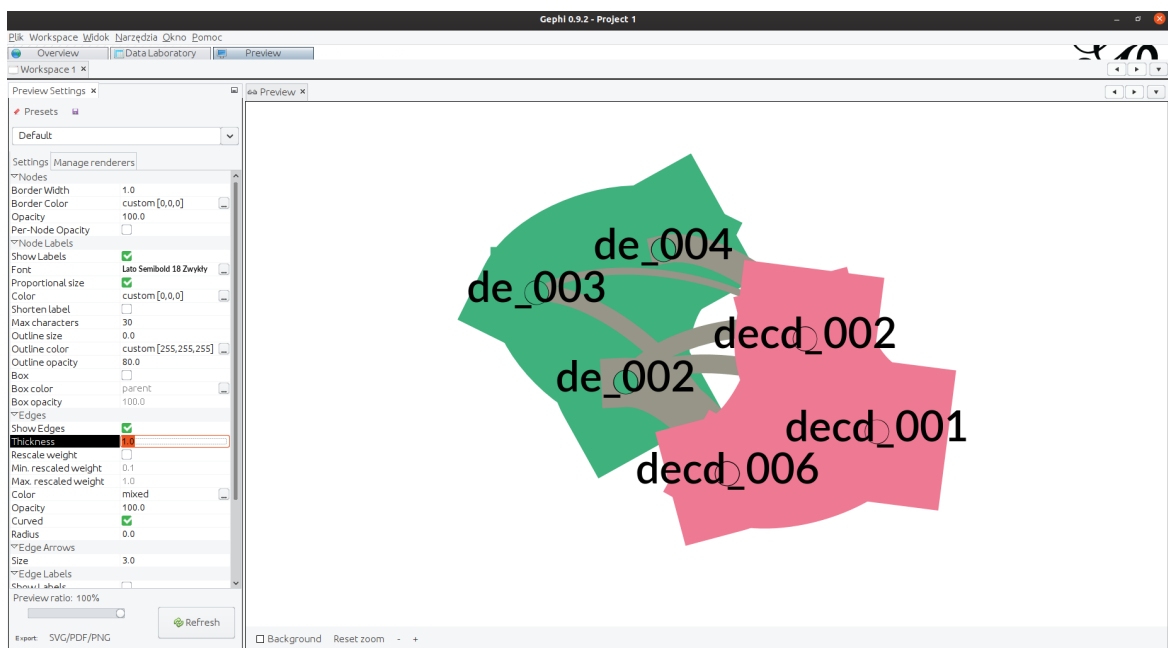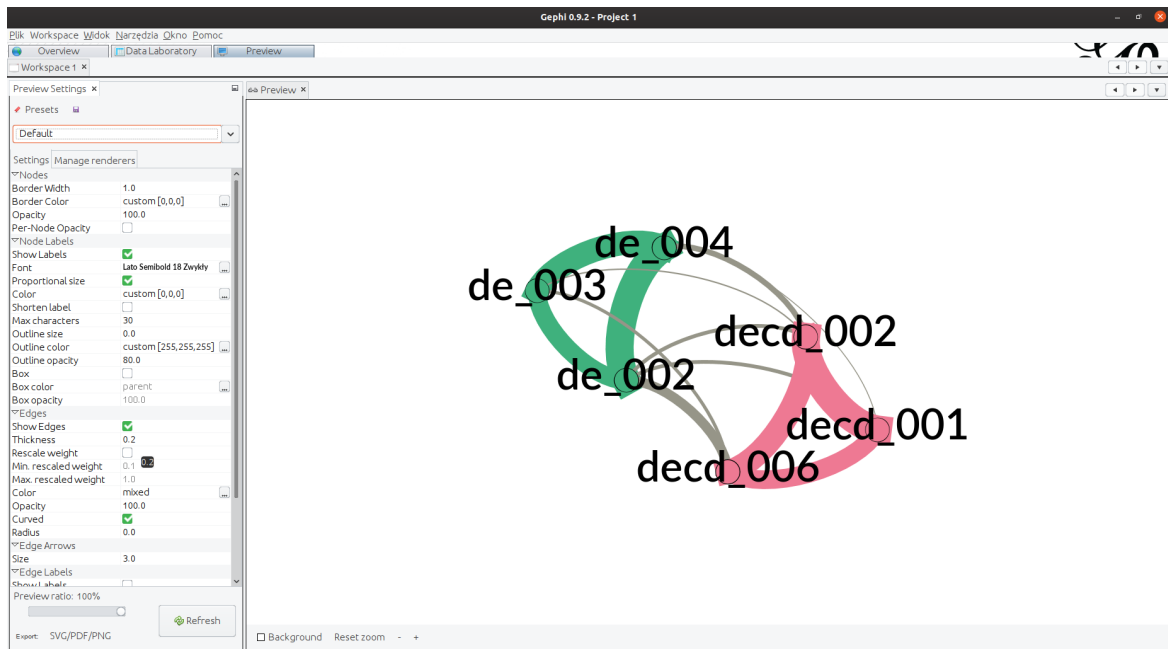

6. Now go to **Preview**:

   a. In **Nodes** select **Show Labels**, adjust the font and refresh the graph,

b. Sometimes, if the connections are really hard to read, you need to go to **Edges** make **Thickness** smaller


And they should look like this:

7. To save the graph as a picture go to **File→Export→SVG/PNG/PDF file** and choose your preferred option. Note that PDF gives you high quality and enables you to copy names from the picture, however it sometimes cuts the graph, so go to **Options** and give it bigger margins - e.g. = 1 and preferably, choose landscape orientation.
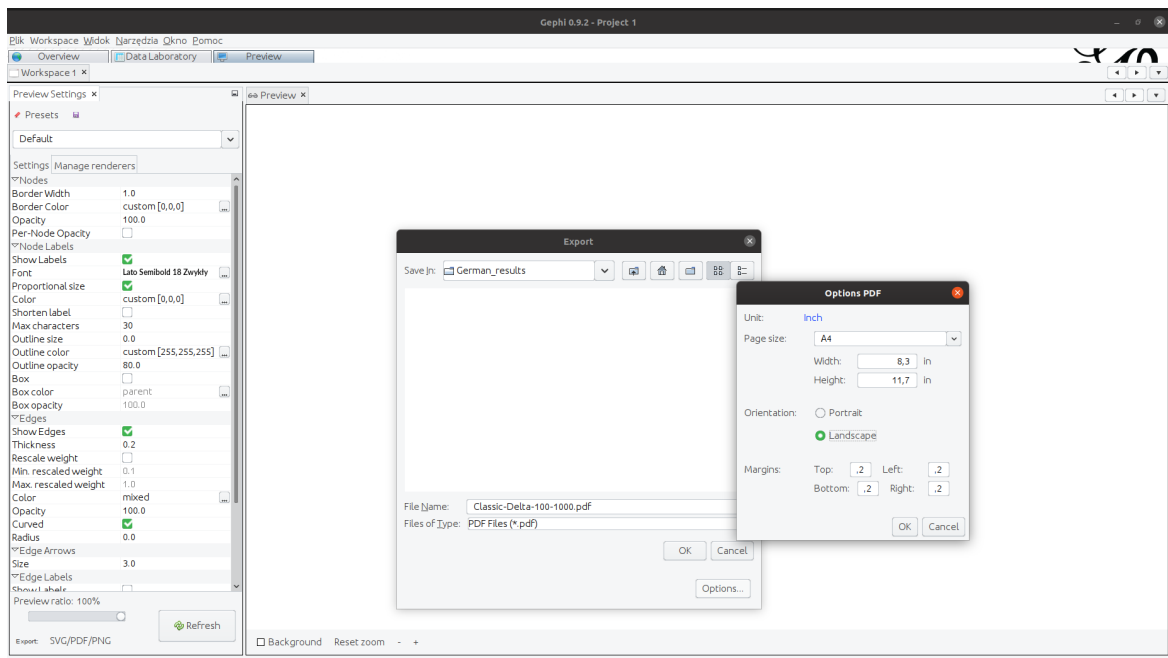
8. To save the whole project (with its latest settings) go to **File→Save as** and in the rolling list **Files of Type** choose **Gephi Files** rather than **All Files** - note, it likes to crash between different computers otherwise.
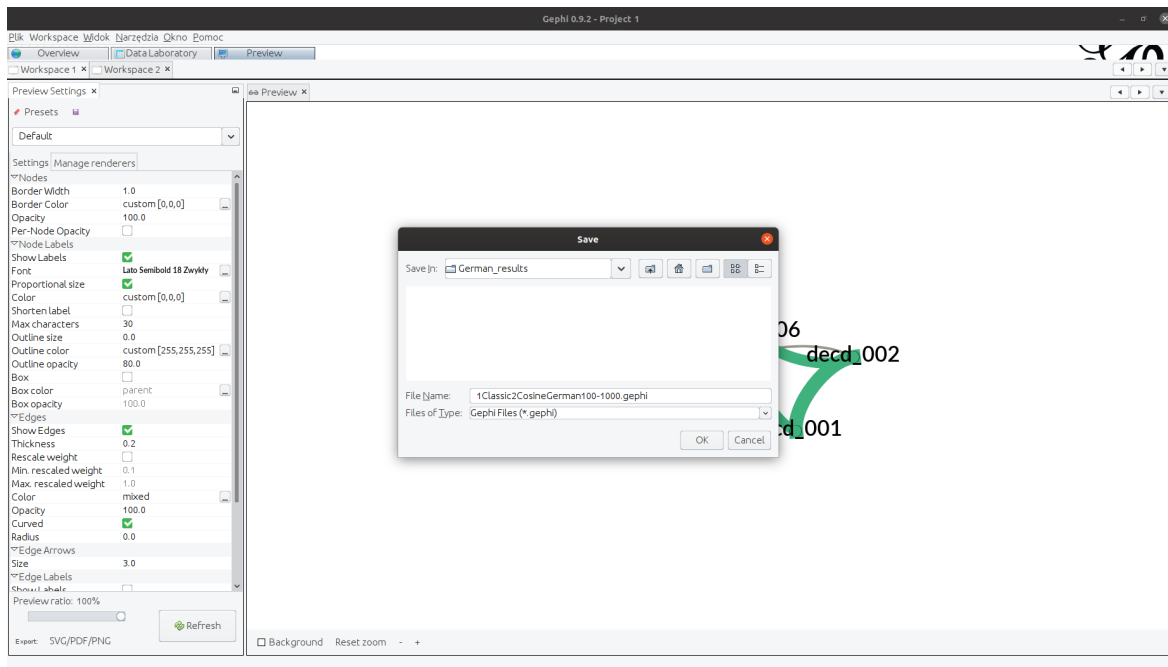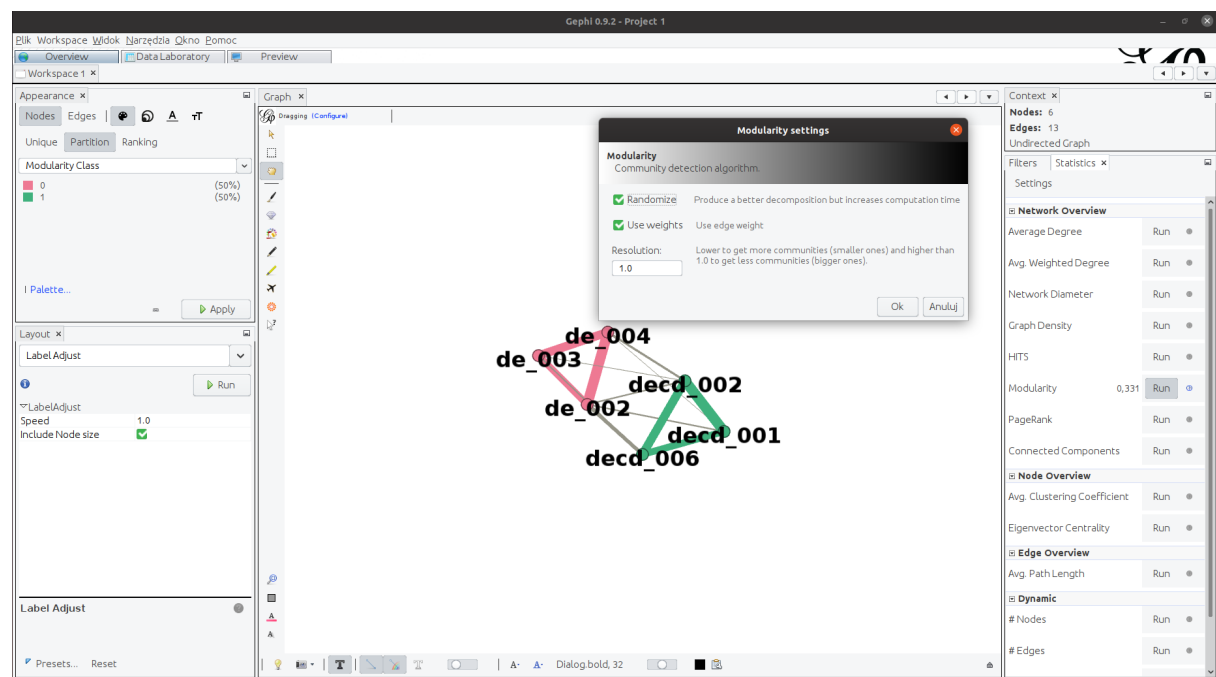
Community Detection

Modularity:

In Overview, go to the menu on the right and select
Statistics → Network Overview → Modularity

After running it you will get a notice on how many groups were found, you can choose to color your network according to Modularity Class in **Appearance→Nodes→ModularityClass**

Final notes

Gephi has quite a lot of extra tutorials on specific issues, you can find them here:

https://gephi.org/users/

https://gephi.org/tutorials/gephi-tutorial-layouts.pdf ←more information on layouts, I found that quite useful

Learn networks with DraCor cards: https://dracor.org/doc/merch

https://kateto.net/network-visualization