Simple Gephi Project from A to Z

Clément Levallois

2017-01-20

Table of Contents

Description of the project

This project is for complete beginners to Gephi. It supposes you have Gephi installed and running on your computer. That is all.

When finishing this tutorial, you should be able to:

- understand the vocabulary to discuss networks

- download a network file for this exercise

- description of the file / the network

- open a network file

- read the report after opening a file

- show the labels of the nodes

- layout the network

- visualize attributes of the network

- prettify the network for enhanced readability

- compute the centrality of the nodes in the network

- visualize attributes created by Gephi

- export a visualization as a picture or pdf

know the terminology to discuss networks

Figure 1. terminology for networks

download a network file

download this zip file and unzip it on your computer.

You should find the file `miserables.gexf` in it.

Save it in a folder you will remember (or create a folder specially for this small project).

description of the file / the network

This file contains a network representing "who appears next to whom" in the 19th century novel *Les Misérables* by Victor Hugo [1: D. E. Knuth, The Stanford GraphBase: A Platform for Combinatorial Computing, Addison-Wesley, Reading, MA (1993)].

A link between characters A and B means they appeared on the same page or paragraph in the novel.

The file name ends with ".gexf", which just means this is a text file where the network information is stored (name of the characters, their relations, etc.), following some conventions.

open the network

- open Gephi. On the Welcome screen that appears, click on `Open Graph File`

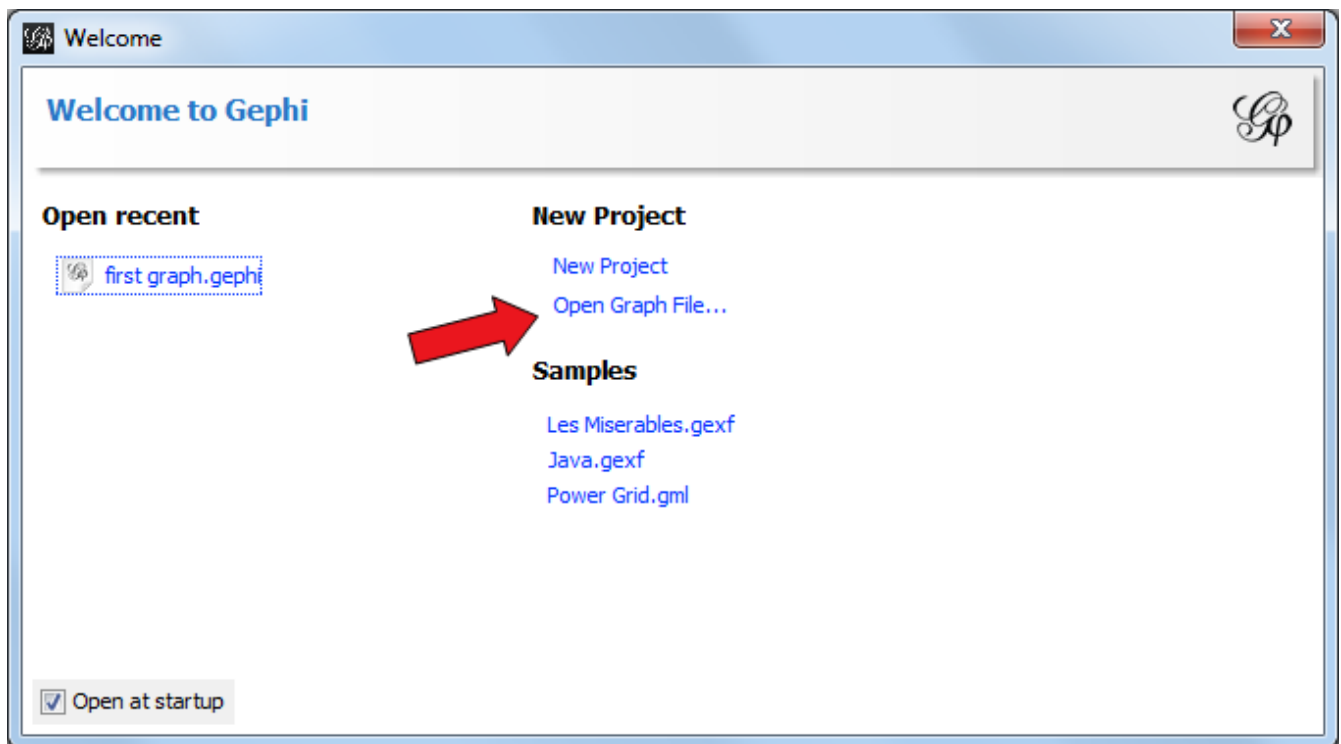- find `miserables.gexf` on your computer and open it

Figure 2. welcome screen

read the report after opening a file

A report window will open, giving you basic info on the network you opened:
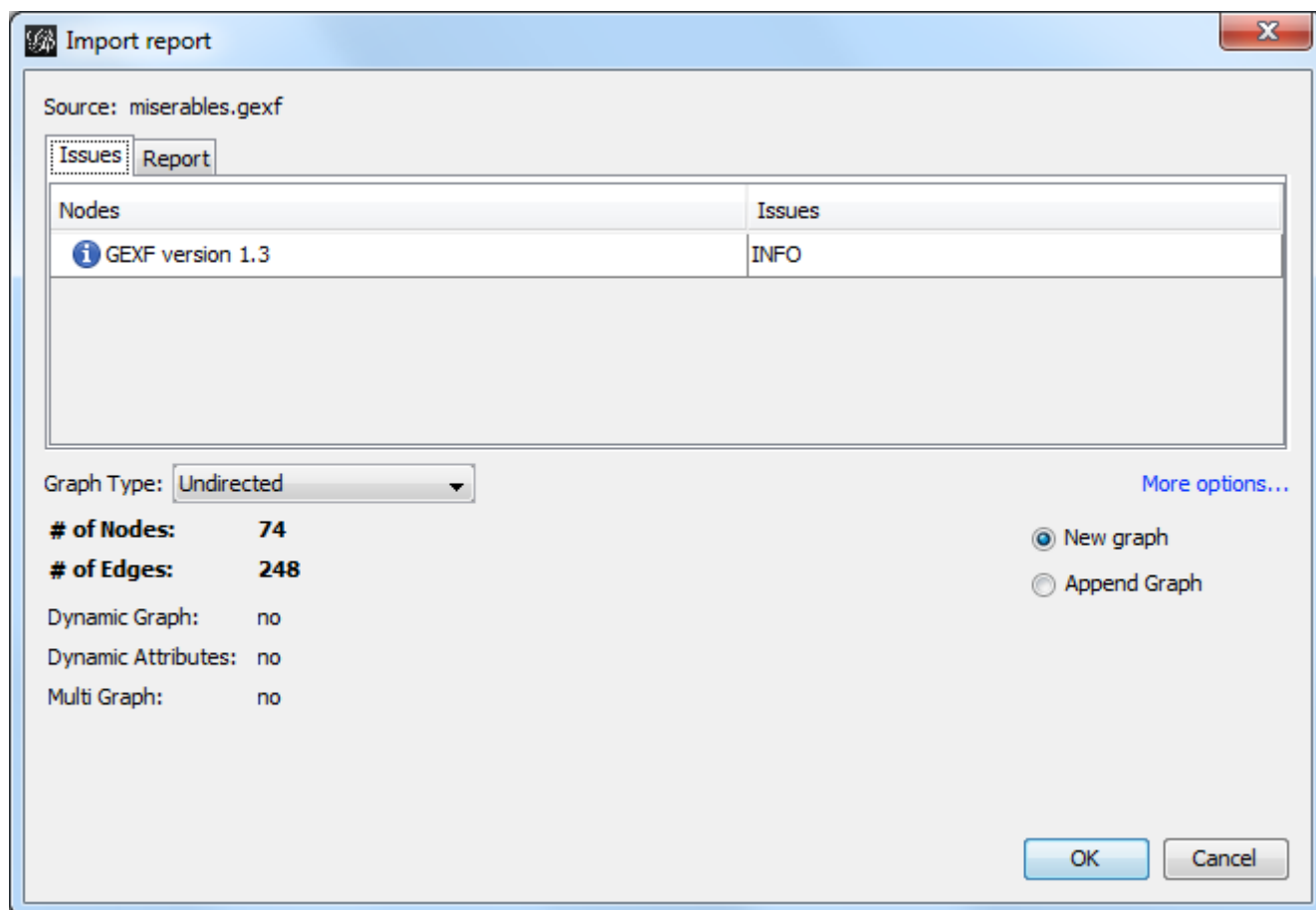
Figure 3. report window

This tells you that the network comprises 74 characters, connected by 248 links.

Links are undirected, meaning that if A is connected to B, then it is the same as B connected to A.

The report also tells us the graph is not dynamic: it means there is no evolution or chronology, it won't "move in time".

Click on OK to see the graph in Gephi.

initial view

Figure 4. initial view when opening a graph

This is how the network appears in Gephi. Not very useful! Let's examine what we have here.

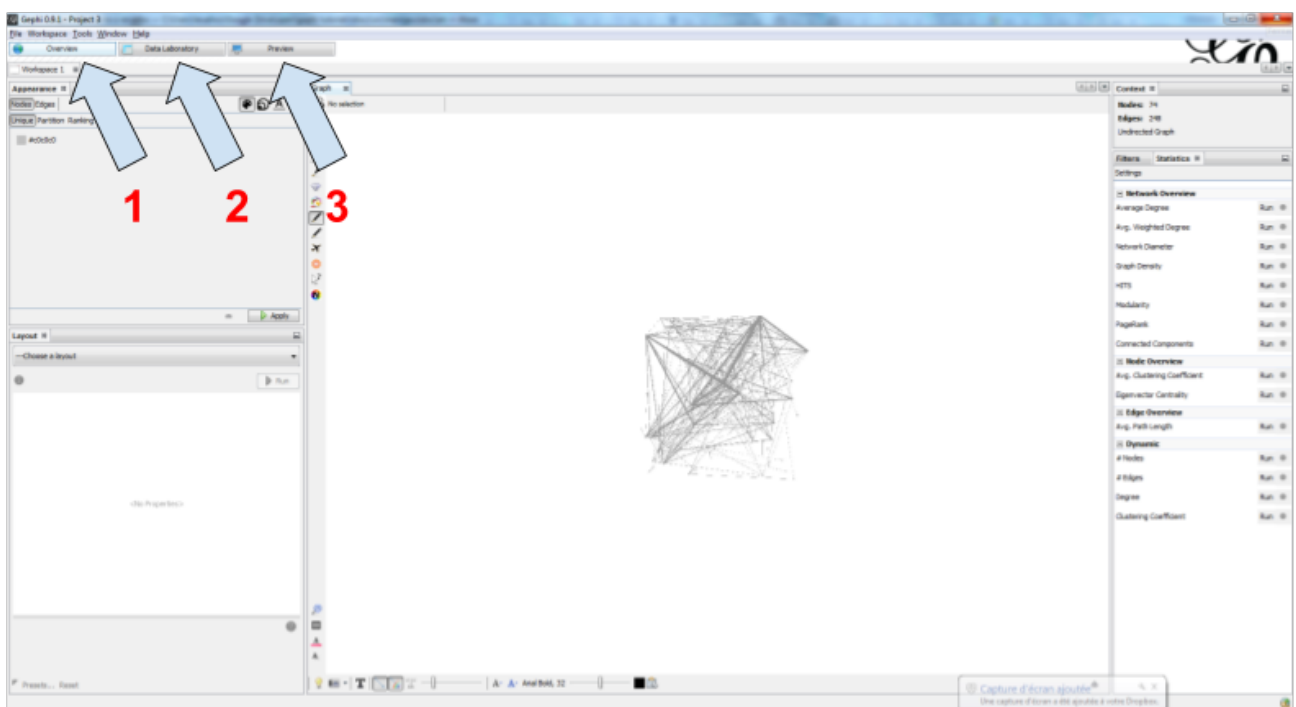# basic view of Gephi's interface

Figure 5. the 3 main screens in Gephi

Gephi has 3 main screens:

1. Overview: where we can explore the graph visually

2. Data Laboratory: provides an "Excel" table view of the data in network

3. Preview: where we polish the visualization before exporting it as a pictue or pdf
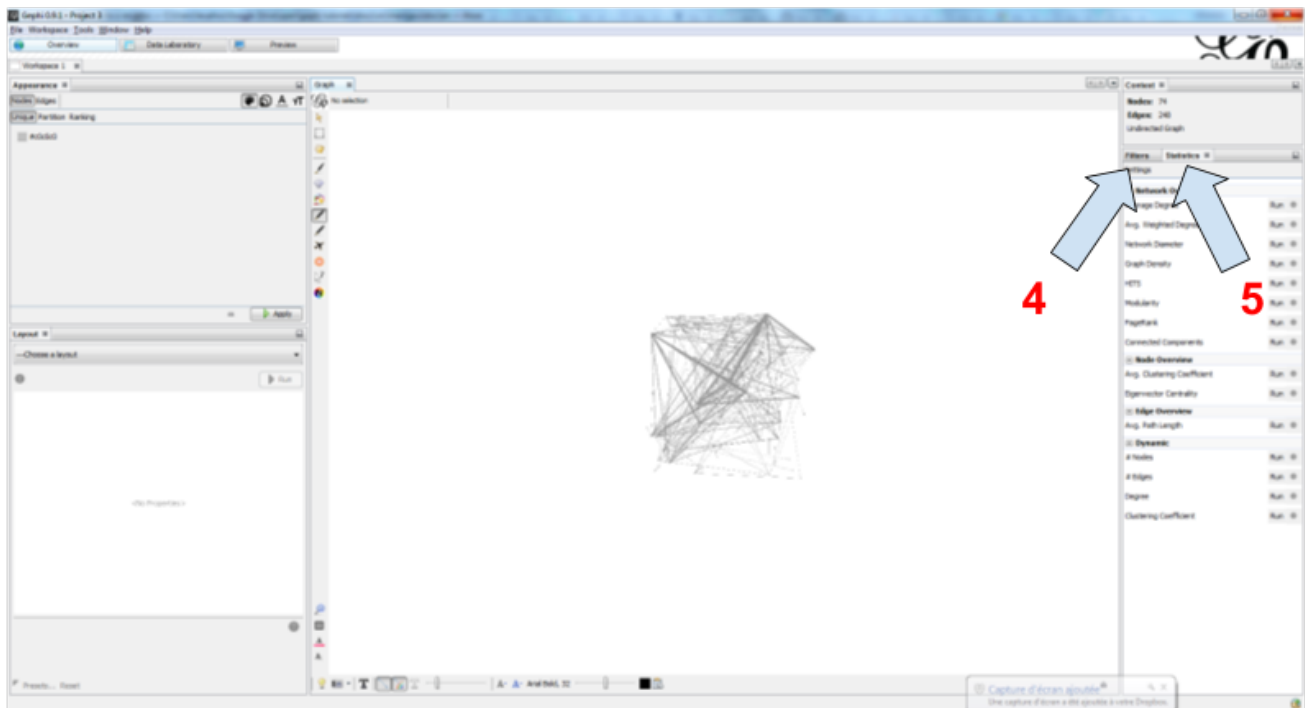
What we see here is the Overview.

Figure 6. Filters and statistics panels in Gephi

In the Overview, the graph is shown at the center. Around it, several panels help us fine tune the visualization.

4. "Filters", where we can hide different parts of the network under a variety of conditions
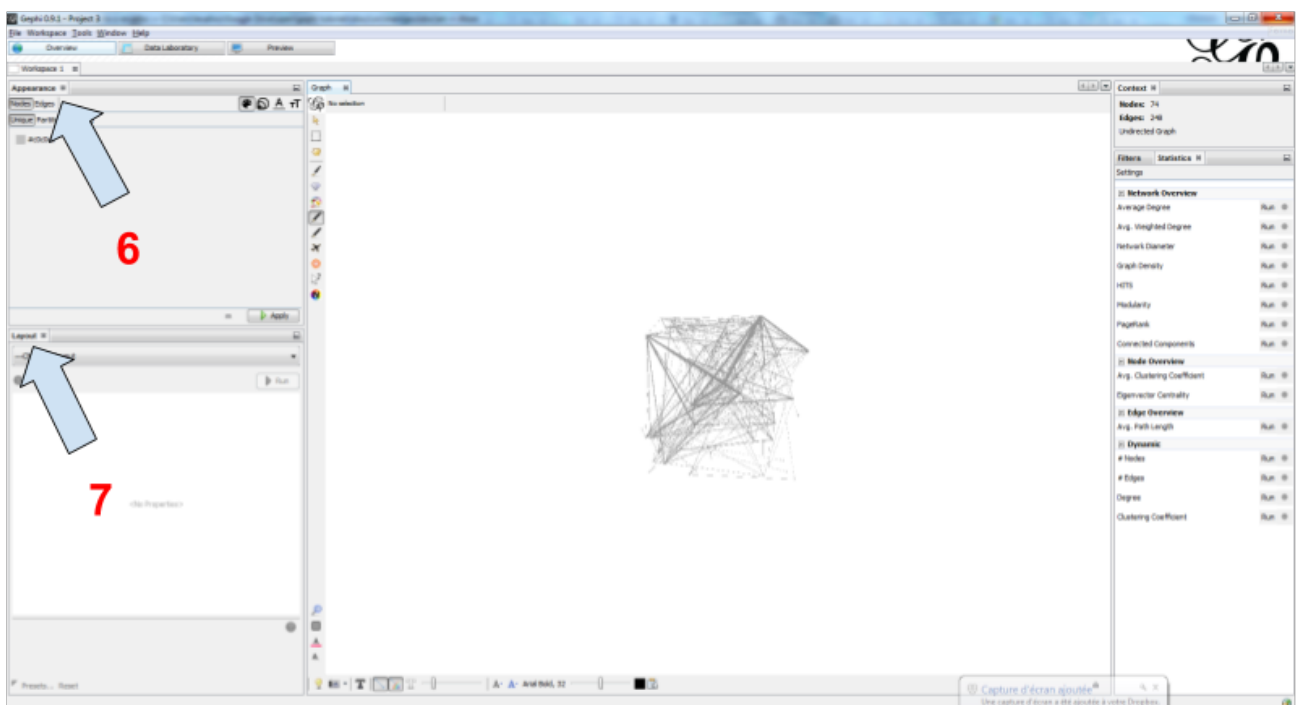5. "Statistics", where we can compute metrics on the network

Figure 7. Appearance and layout panels in Gephi

6. "Appearance", where we can change colors and sizes in interesting ways

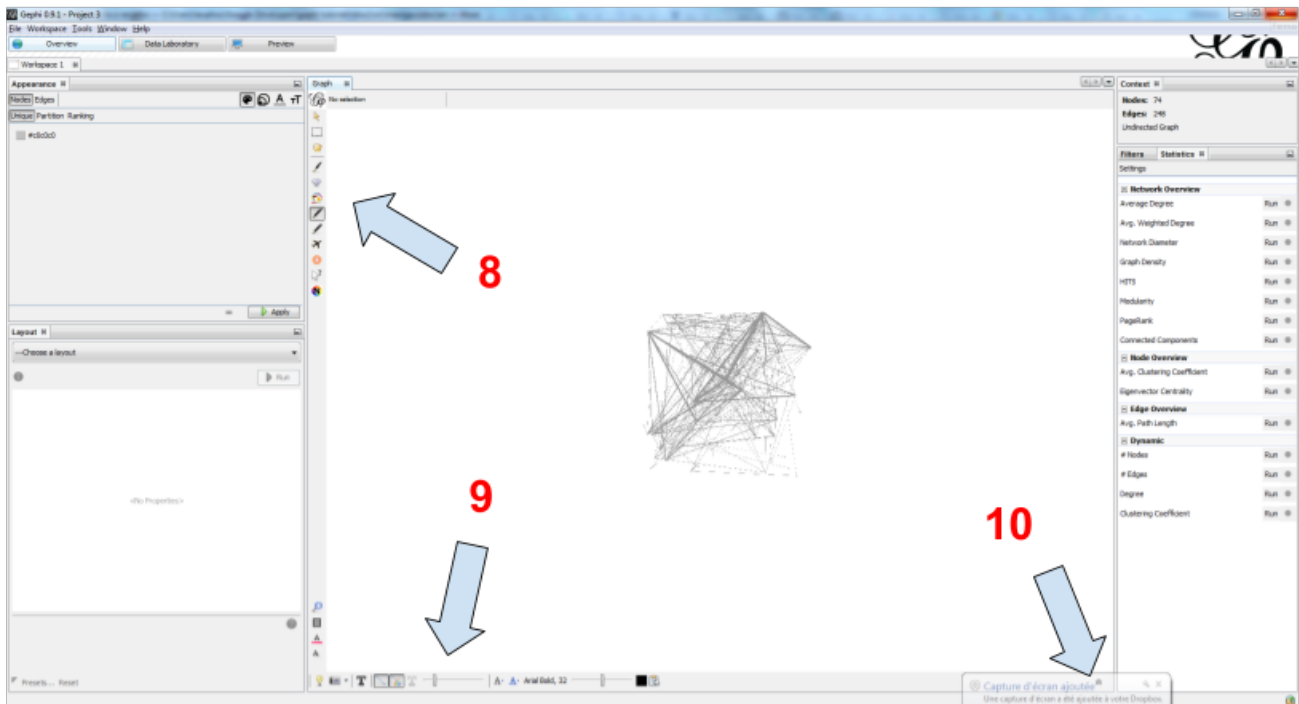7. "Layouts", where we can apply automated procedures to change the position of the network

Figure 8. 3 groups of icons

8. A series of icons to add / colorize nodes and links manually, by cliking on them

9. Options and sliders to change the size of all nodes, links, or labels

10. More options become visible if we click on this **little arrow head pointing up**
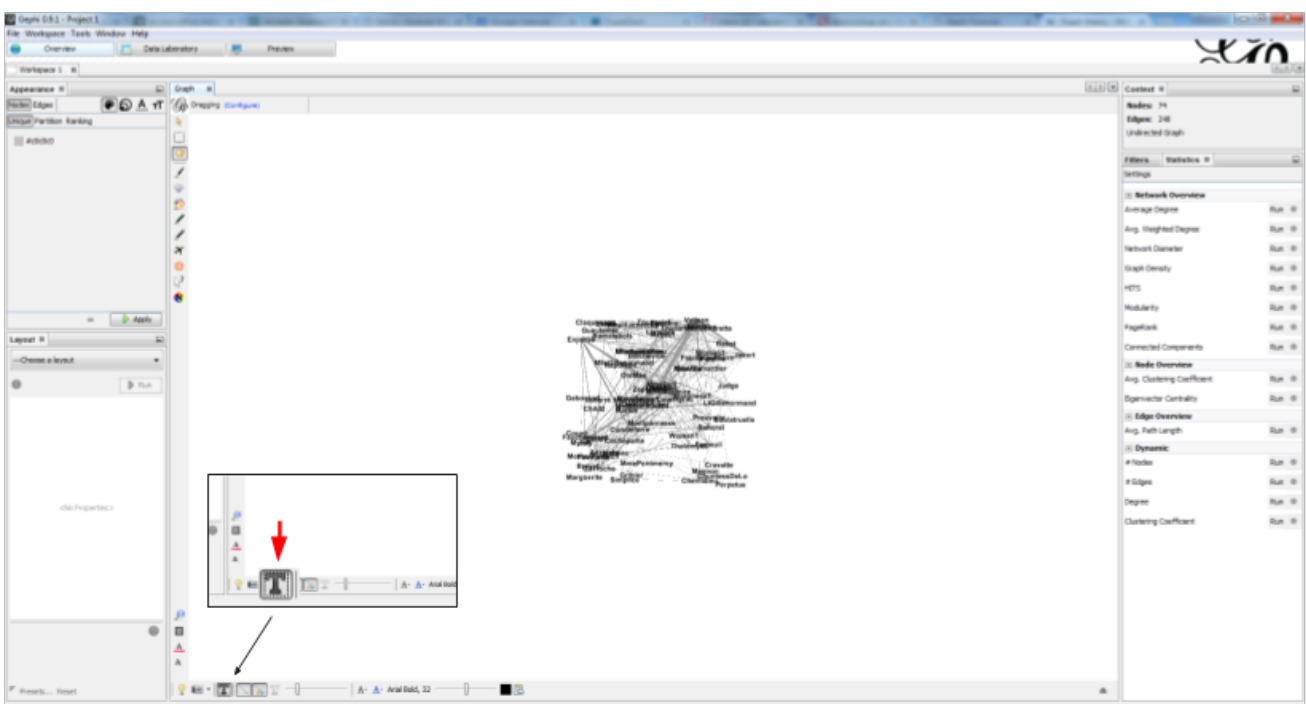
showing labels of the nodes

Figure 9. showing node labels
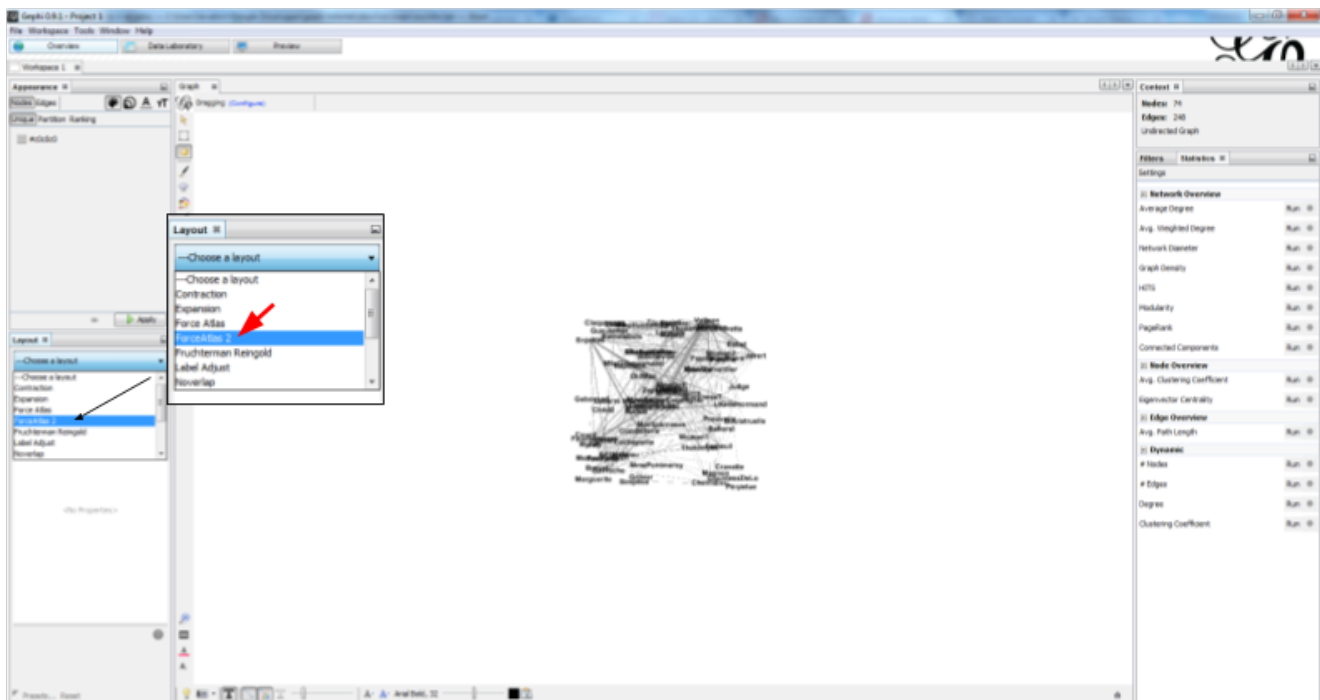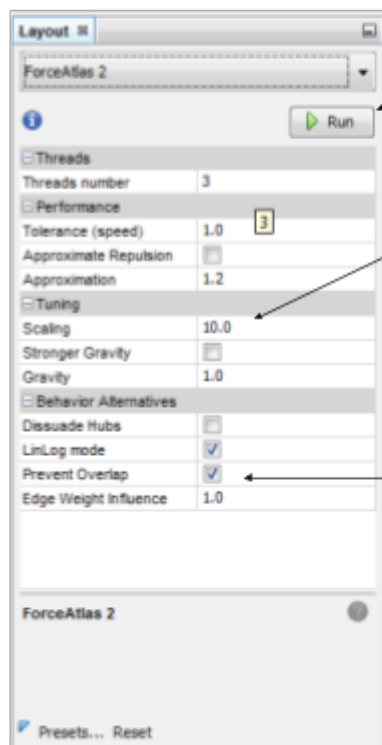
layout ("spatialize") the network

Figure 10. selecting the force atlas 2 layout

Figure 11. changing a few parameters and launching the layout
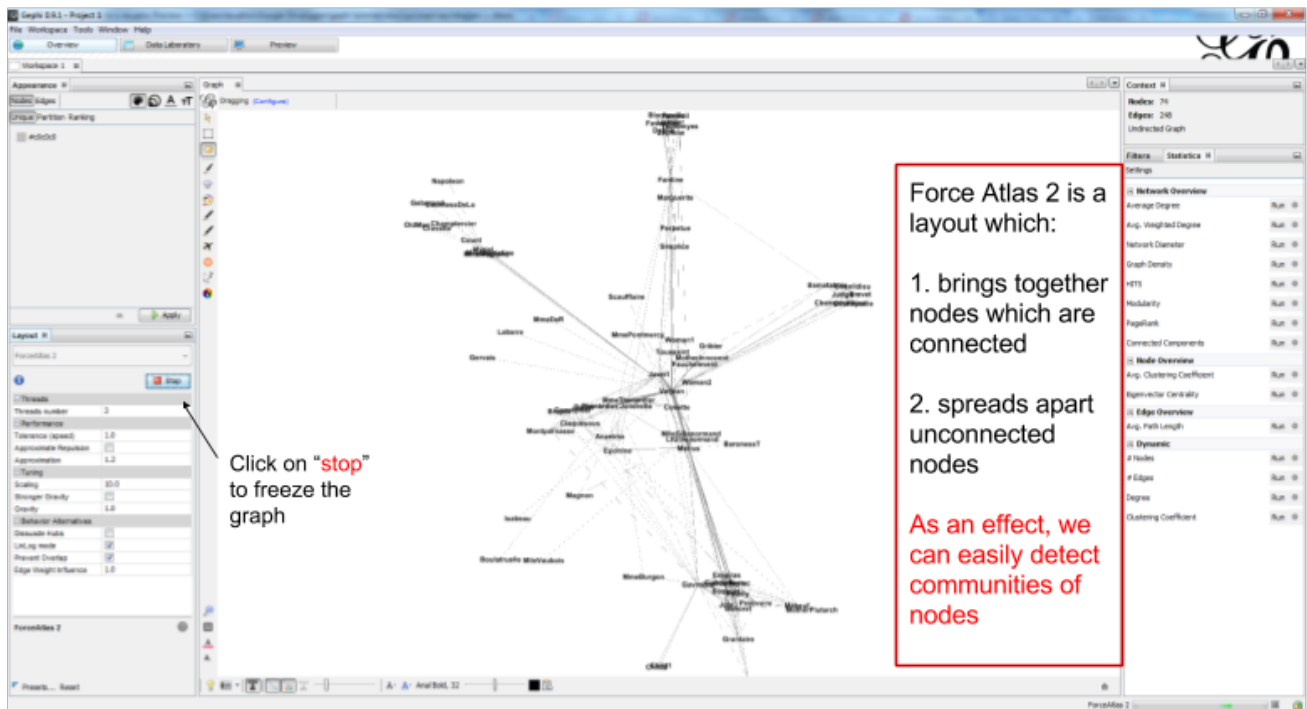
Figure 12. result of Force Atlas 2 layout

visualize the properties of the nodes

A network consists in entities and their relations. This is what we just visualized. Yet, the properties of these entities remain invisible.

For instance: the characters in the novel "Les Misérables" are male or female. Are males more likely to be connected to males, or females? Just looking at the network in Gephi, we can't tell.

Now, we will see how to make this property ("Gender") visible.

First, let's switch to the `Data Laboratory`'" view:
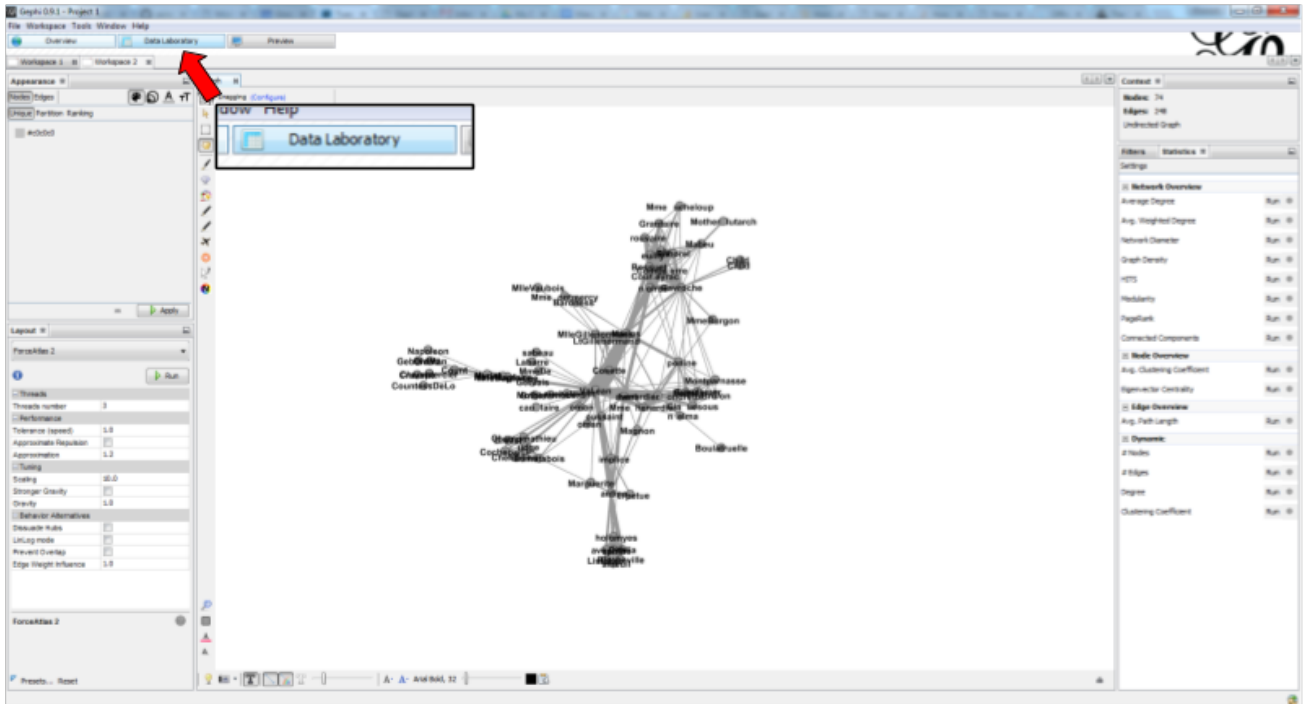
Figure 13. Switching the view to the data laboratory

We see the list of the characters in the network. There is a column `Gender`, where each character is either "M" or "F". Click on `Overview` to come back to the visual version.
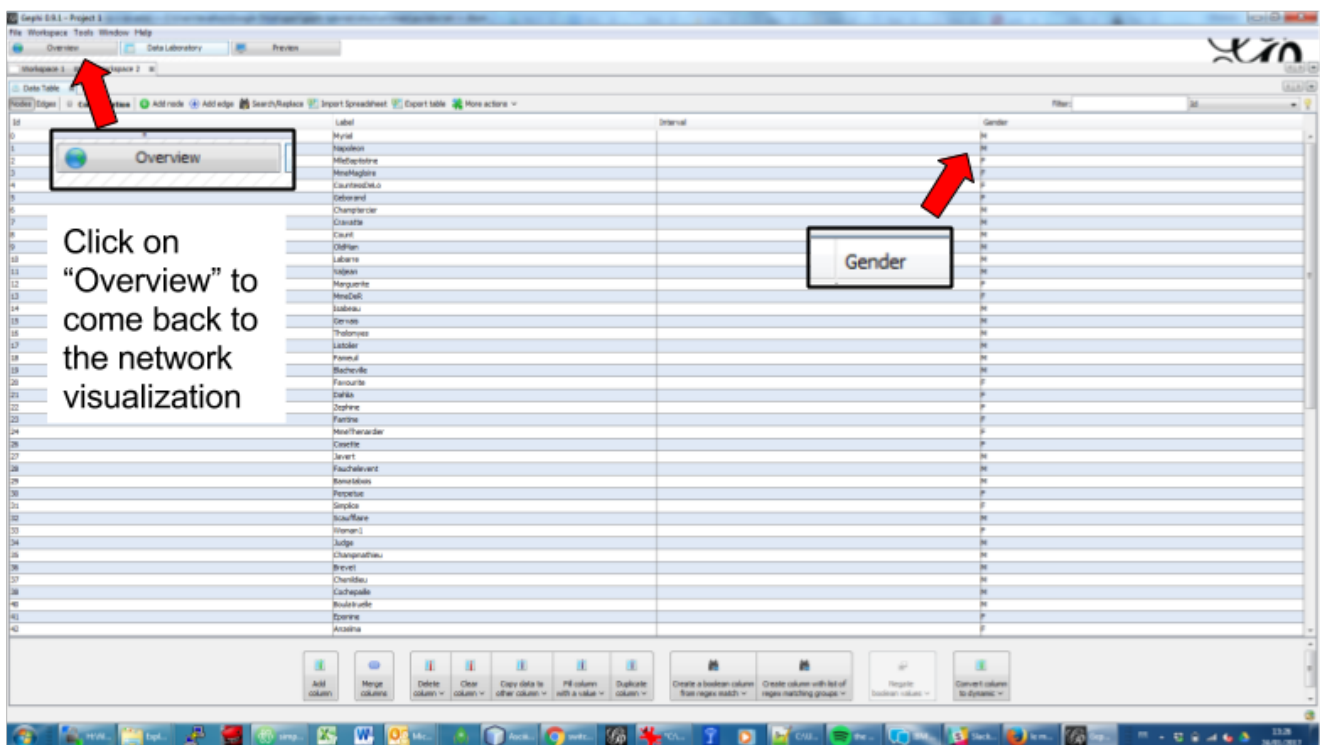
Figure 14. We see there is a Gender attribute for each character.

We will color the nodes based on their gender. To do that, we select `Gender` in the `Appearance` panel:

Figure 15. Coloring nodes according to their gender

The result:

Figure 16. After coloring characters according to their gender

prettify the network for enhanced readability

There are a number of issues with the result we get:

1. the network is too big or too small, it is hard to read

2. the labels of the characters overlap

3. the size of the labels might be too big / small

4. the links are sometimes too large

Let's fix these issues.

To enlarge or shrink the network, we have 2 options:

- either we use the "scale" parameters of the layout, as we have seen here.

- or the scale is fine, it is just that we need to zoom it or out. If so, you can use the scrolling wheel of your mouse.

(to be continued)

questions and exercises

1. Open the file `miserables.gexf` with a text editor (here is how to do it on a Mac, and on Windows). See how the nodes and the links are written in the file. Can you find the character Javert?

2. Our network of Les Miserables characters was undirected. Can you think of networks which are directed?

Imagine how undirected and directed networks differ when computing centrality, for example.

3. Force Atlas 2 is a layout which brings together connected nodes, and spreads out unconnected nodes. We might have nodes with no relation at all with other nodes (called "isolated nodes").

How will these isolated nodes move on screen?

4. When applying the Force Atlas 2 layout, the network moves quickly, then stabilizes, and then keeps moving a bit.

Can you guess why this is happening?

5. In the list of layouts, Force Atlas 2 is just one of many options you can choose. Try "Fruchterman Reingold" and "Yfan Hu".

These are layouts which follow the same logic as Force Atlas 2, but with slight variations. Explore how these algorithms result in similar, yet specific layouts.