

# Using filters

Clément Levallois

2017-02-10

# Table of Contents

download a network file for practice .....	1
open the network in Gephi .....	1
getting a sense of the attributes in the data laboratory .....	3
discovering the filter panel .....	5
combining 2 filters .....	7
1. Case when the placement of filters makes no difference .....	7
2. Case when the placement of filters makes a difference .....	9
filter operators .....	13
1. The MASK operator .....	13
2. The UNION operator .....	15
3. The NOT operator .....	22
more tutorials on using filters in Gephi .....	25
the end .....	25



## download a network file for practice

[download this zip file](#) and unzip it on your computer.

or use this direct link: <https://tinyurl.com/gephi-tuto-3>

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).

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 in Gephi

- open Gephi. On the Welcome screen that appears, click on **Open Graph File**
- find `miserables.gexf` on your computer and open it

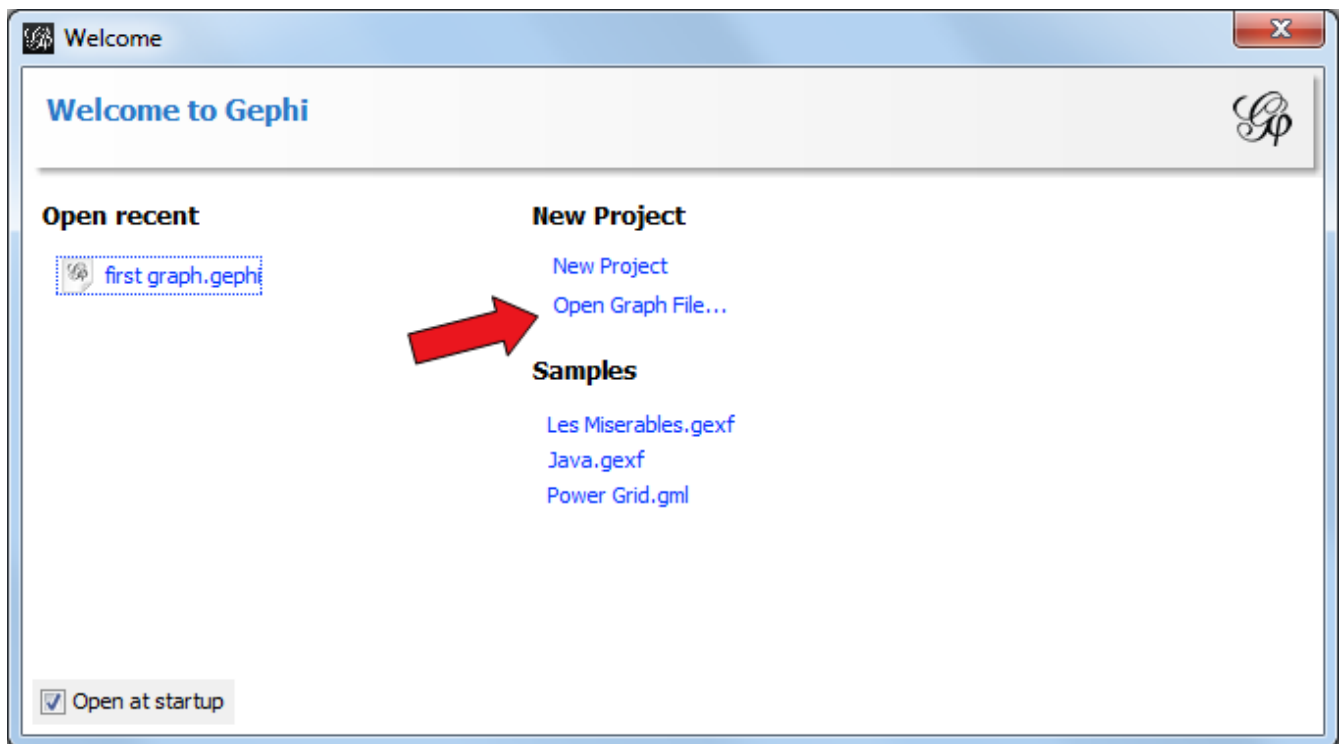


Figure 1. welcome screen

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

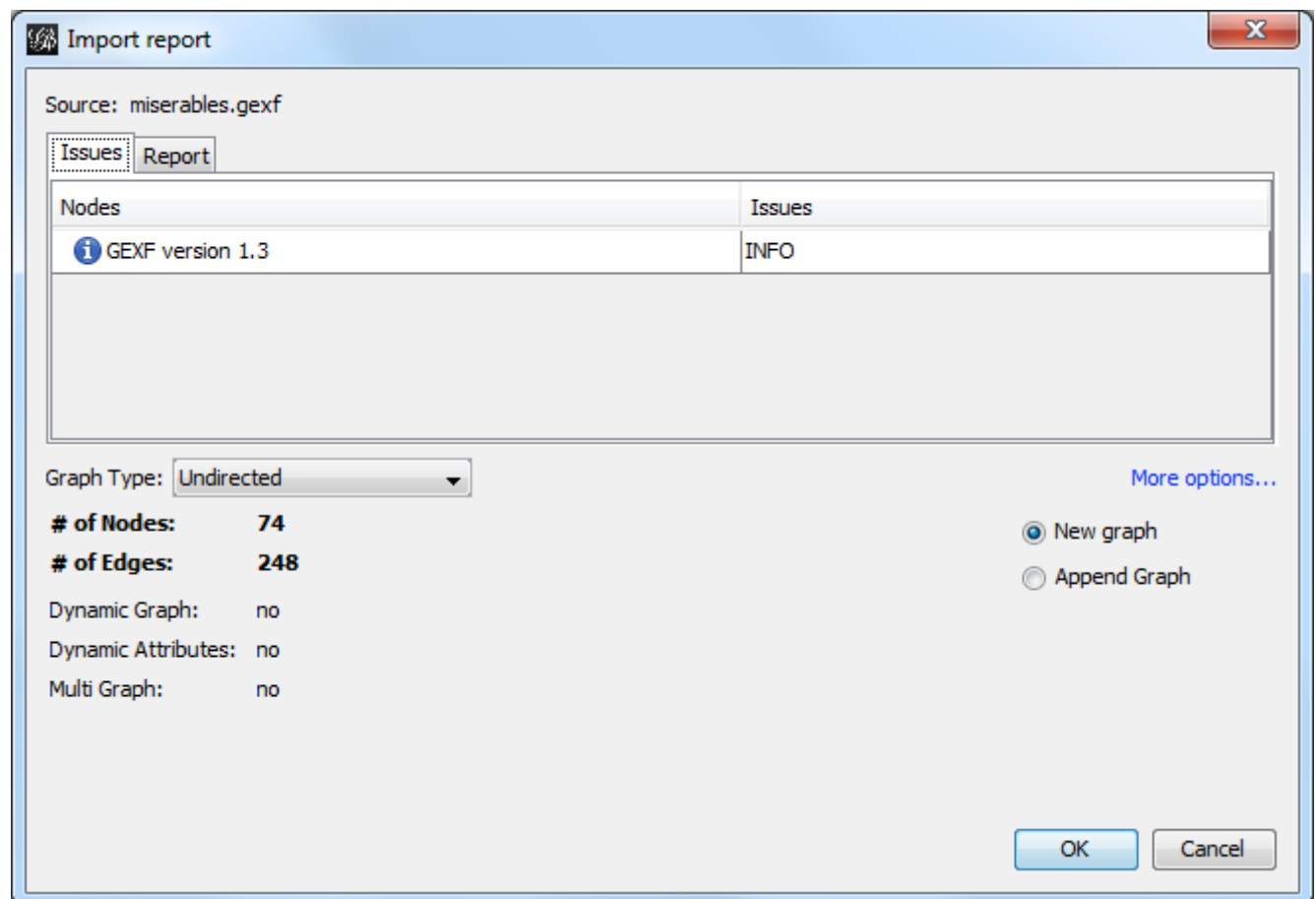


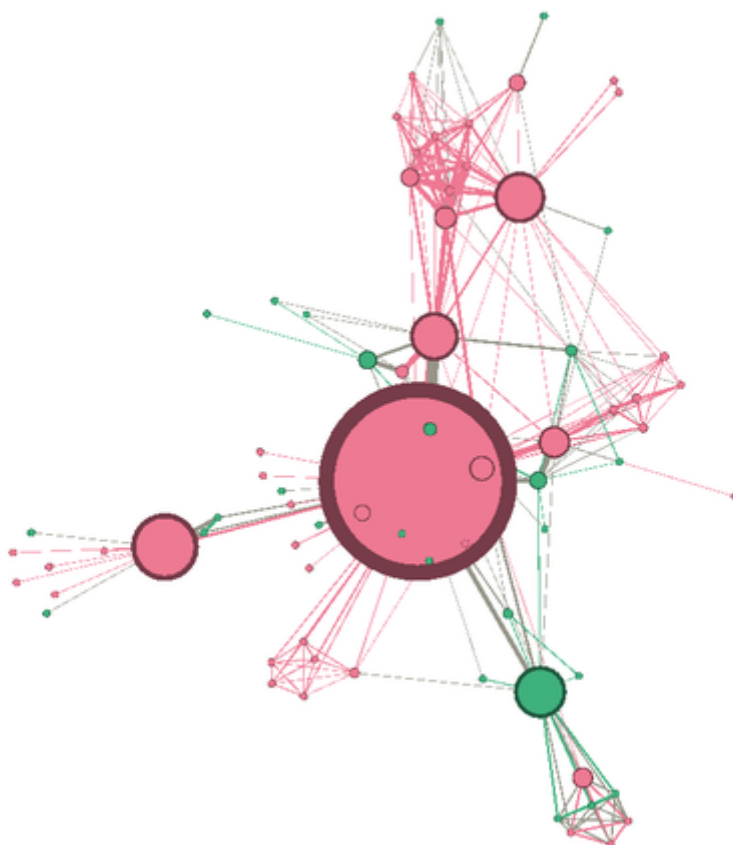
Figure 2. report window

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.



*Figure 3. The network we will use*

## getting a sense of the attributes in the data laboratory

We can switch to the data laboratory to see the underlying data:

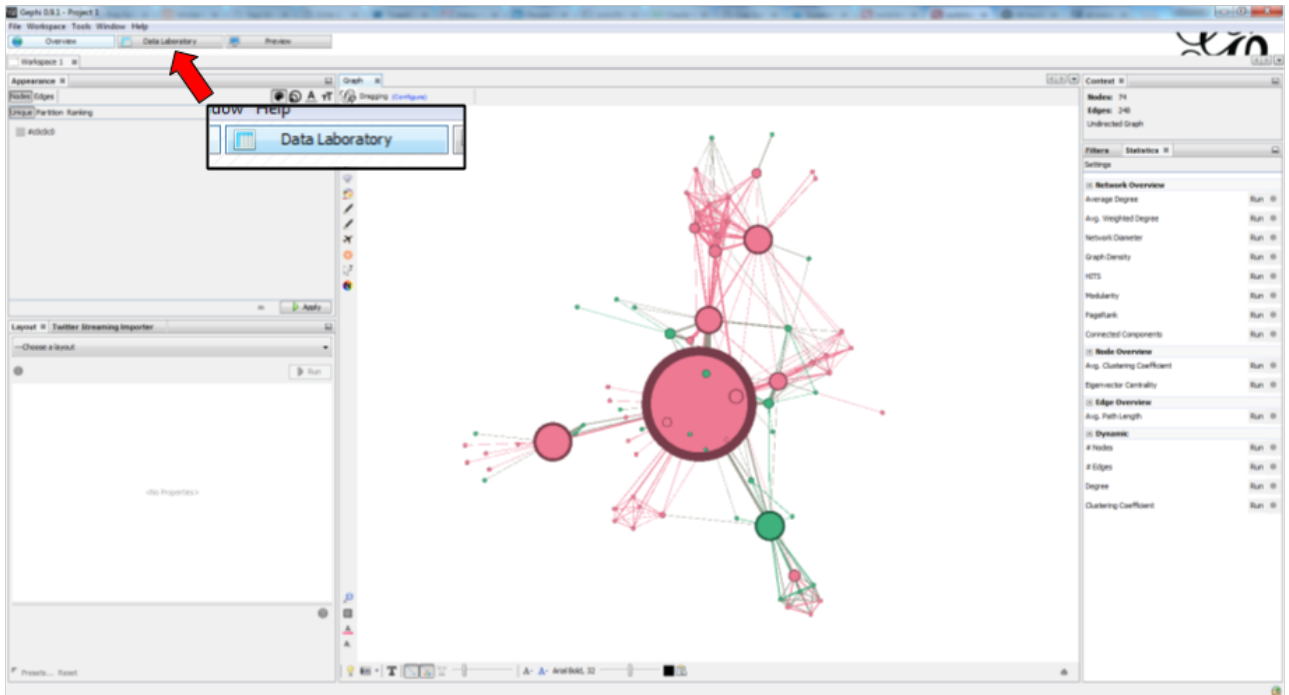


Figure 4. Switching to the data laboratory

We see that the nodes of the network have many attributes. In particular, each have a Gender and a measure of how central they are:

ID	Label	Interval	Gender	Eccentricity	Betweenness Centrality	Harmonic Closeness Centrality	Closeness Centrality
1	Muriel		M	4.0	0.0	0.437126	0.437126
2	Napoleon		M	5.0	0.0	0.305429	0.305429
3	Midnapoline		F	4.0	0.0	0.41954	0.41954
4	Herzogin		F	4.0	0.0	0.41954	0.41954
5			F	5.0	0.0	0.305429	0.305429
6			F	5.0	0.0	0.305429	0.305429
7			M	5.0	0.0	0.305429	0.305429
8			M	5.0	0.0	0.305429	0.305429
9			M	5.0	0.0	0.305429	0.305429
10			M	4.0	0.0	0.305429	0.305429
11			M	5.0	0.0	0.305429	0.305429
12			F	5.0	0.0	0.305429	0.305429
13			F	4.0	0.0	0.305429	0.305429
14			M	4.0	0.0	0.305429	0.305429
15			M	4.0	0.0	0.305429	0.305429
16			M	4.0	0.0	0.305429	0.305429
17			M	4.0	0.0	0.305429	0.305429
18			M	4.0	0.0	0.305429	0.305429
19	Blanche		F	4.0	0.0	0.305429	0.305429
20	Parasite		F	4.0	0.0	0.305429	0.305429
21	Dalia		F	4.0	0.0	0.305429	0.305429
22	Josephine		F	4.0	0.0	0.305429	0.305429
23	Parine		F	4.0	0.0	0.305429	0.305429
24	Herzogin		F	5.0	0.0	0.305429	0.305429
25	Cosette		F	5.0	0.0	0.305429	0.305429
26	Javert		M	5.0	0.0	0.305429	0.305429
27	Pauchonnet		M	5.0	0.0	0.305429	0.305429
28	Blanche		F	5.0	0.0	0.305429	0.305429
29	Parasite		F	5.0	0.0	0.305429	0.305429
30	Josephine		F	5.0	0.0	0.305429	0.305429
31	Parine		F	5.0	0.0	0.305429	0.305429
32	Herzogin		F	5.0	0.0	0.305429	0.305429
33	Blanche		F	5.0	0.0	0.305429	0.305429
34	Parasite		F	5.0	0.0	0.305429	0.305429
35	Josephine		F	5.0	0.0	0.305429	0.305429
36	Parine		F	5.0	0.0	0.305429	0.305429
37	Herzogin		F	5.0	0.0	0.305429	0.305429
38	Blanche		F	5.0	0.0	0.305429	0.305429
39	Parasite		F	5.0	0.0	0.305429	0.305429
40	Josephine		F	5.0	0.0	0.305429	0.305429
41	Parine		F	5.0	0.0	0.305429	0.305429
42	Herzogin		F	5.0	0.0	0.305429	0.305429

Figure 5. Nodes attributes.

This is the list of edges (relations) in the network. Notice that they have a "weight" (a "strength").

We can then go back to the "Overview" and start using Filters.

The "Weight" is a number representing how many times the 2 characters are associated, in the novel "Les Miserables" by Victor Hugo.

Source	Target	Type	Weight
1 - Napoleon	2 - Millaudaine	Undirected	3
2 - Millaudaine	3 - Millaudaine	Undirected	3
3 - Millaudaine	4 - Courtenay	Undirected	2
4 - Courtenay	5 - Courtenay	Undirected	4
5 - Courtenay	6 - Courtenay	Undirected	5
6 - Courtenay	7 - Courtenay	Undirected	6
7 - Courtenay	8 - Courtenay	Undirected	7
8 - Courtenay	9 - Courtenay	Undirected	8
9 - Courtenay	10 - Courtenay	Undirected	9
10 - Courtenay	11 - Courtenay	Undirected	10
11 - Courtenay	12 - Courtenay	Undirected	11
12 - Courtenay	13 - Courtenay	Undirected	12
13 - Courtenay	14 - Courtenay	Undirected	13
14 - Courtenay	15 - Courtenay	Undirected	14
15 - Courtenay	16 - Courtenay	Undirected	15
16 - Courtenay	17 - Courtenay	Undirected	16
17 - Courtenay	18 - Courtenay	Undirected	17
18 - Courtenay	19 - Courtenay	Undirected	18
19 - Courtenay	20 - Courtenay	Undirected	19
20 - Courtenay	21 - Courtenay	Undirected	20
21 - Courtenay	22 - Courtenay	Undirected	21
22 - Courtenay	23 - Courtenay	Undirected	22
23 - Courtenay	24 - Courtenay	Undirected	23
24 - Courtenay	25 - Courtenay	Undirected	24
25 - Courtenay	26 - Courtenay	Undirected	25
26 - Courtenay	27 - Courtenay	Undirected	26
27 - Courtenay	28 - Courtenay	Undirected	27
28 - Courtenay	29 - Courtenay	Undirected	28
29 - Courtenay	30 - Courtenay	Undirected	29
30 - Courtenay	31 - Courtenay	Undirected	30
31 - Courtenay	32 - Courtenay	Undirected	31
32 - Courtenay	33 - Courtenay	Undirected	32
33 - Courtenay	34 - Courtenay	Undirected	33
34 - Courtenay	35 - Courtenay	Undirected	34
35 - Courtenay	36 - Courtenay	Undirected	35
36 - Courtenay	37 - Courtenay	Undirected	36
37 - Courtenay	38 - Courtenay	Undirected	37
38 - Courtenay	39 - Courtenay	Undirected	38
39 - Courtenay	40 - Courtenay	Undirected	39
40 - Courtenay	41 - Courtenay	Undirected	40
41 - Courtenay	42 - Courtenay	Undirected	41
42 - Courtenay	43 - Courtenay	Undirected	42
43 - Courtenay	44 - Courtenay	Undirected	43
44 - Courtenay	45 - Courtenay	Undirected	44
45 - Courtenay	46 - Courtenay	Undirected	45
46 - Courtenay	47 - Courtenay	Undirected	46
47 - Courtenay	48 - Courtenay	Undirected	47
48 - Courtenay	49 - Courtenay	Undirected	48
49 - Courtenay	50 - Courtenay	Undirected	49
50 - Courtenay	51 - Courtenay	Undirected	50
51 - Courtenay	52 - Courtenay	Undirected	51
52 - Courtenay	53 - Courtenay	Undirected	52
53 - Courtenay	54 - Courtenay	Undirected	53
54 - Courtenay	55 - Courtenay	Undirected	54
55 - Courtenay	56 - Courtenay	Undirected	55
56 - Courtenay	57 - Courtenay	Undirected	56
57 - Courtenay	58 - Courtenay	Undirected	57
58 - Courtenay	59 - Courtenay	Undirected	58
59 - Courtenay	60 - Courtenay	Undirected	59
60 - Courtenay	61 - Courtenay	Undirected	60
61 - Courtenay	62 - Courtenay	Undirected	61
62 - Courtenay	63 - Courtenay	Undirected	62
63 - Courtenay	64 - Courtenay	Undirected	63
64 - Courtenay	65 - Courtenay	Undirected	64
65 - Courtenay	66 - Courtenay	Undirected	65
66 - Courtenay	67 - Courtenay	Undirected	66
67 - Courtenay	68 - Courtenay	Undirected	67
68 - Courtenay	69 - Courtenay	Undirected	68
69 - Courtenay	70 - Courtenay	Undirected	69
70 - Courtenay	71 - Courtenay	Undirected	70
71 - Courtenay	72 - Courtenay	Undirected	71
72 - Courtenay	73 - Courtenay	Undirected	72
73 - Courtenay	74 - Courtenay	Undirected	73
74 - Courtenay	75 - Courtenay	Undirected	74
75 - Courtenay	76 - Courtenay	Undirected	75
76 - Courtenay	77 - Courtenay	Undirected	76
77 - Courtenay	78 - Courtenay	Undirected	77
78 - Courtenay	79 - Courtenay	Undirected	78
79 - Courtenay	80 - Courtenay	Undirected	79
80 - Courtenay	81 - Courtenay	Undirected	80
81 - Courtenay	82 - Courtenay	Undirected	81
82 - Courtenay	83 - Courtenay	Undirected	82
83 - Courtenay	84 - Courtenay	Undirected	83
84 - Courtenay	85 - Courtenay	Undirected	84
85 - Courtenay	86 - Courtenay	Undirected	85
86 - Courtenay	87 - Courtenay	Undirected	86
87 - Courtenay	88 - Courtenay	Undirected	87
88 - Courtenay	89 - Courtenay	Undirected	88
89 - Courtenay	90 - Courtenay	Undirected	89
90 - Courtenay	91 - Courtenay	Undirected	90
91 - Courtenay	92 - Courtenay	Undirected	91
92 - Courtenay	93 - Courtenay	Undirected	92
93 - Courtenay	94 - Courtenay	Undirected	93
94 - Courtenay	95 - Courtenay	Undirected	94
95 - Courtenay	96 - Courtenay	Undirected	95
96 - Courtenay	97 - Courtenay	Undirected	96
97 - Courtenay	98 - Courtenay	Undirected	97
98 - Courtenay	99 - Courtenay	Undirected	98
99 - Courtenay	100 - Courtenay	Undirected	99

Figure 6. Edges attributes

## discovering the filter panel

In the overview, make sure the Filter panel is displayed:

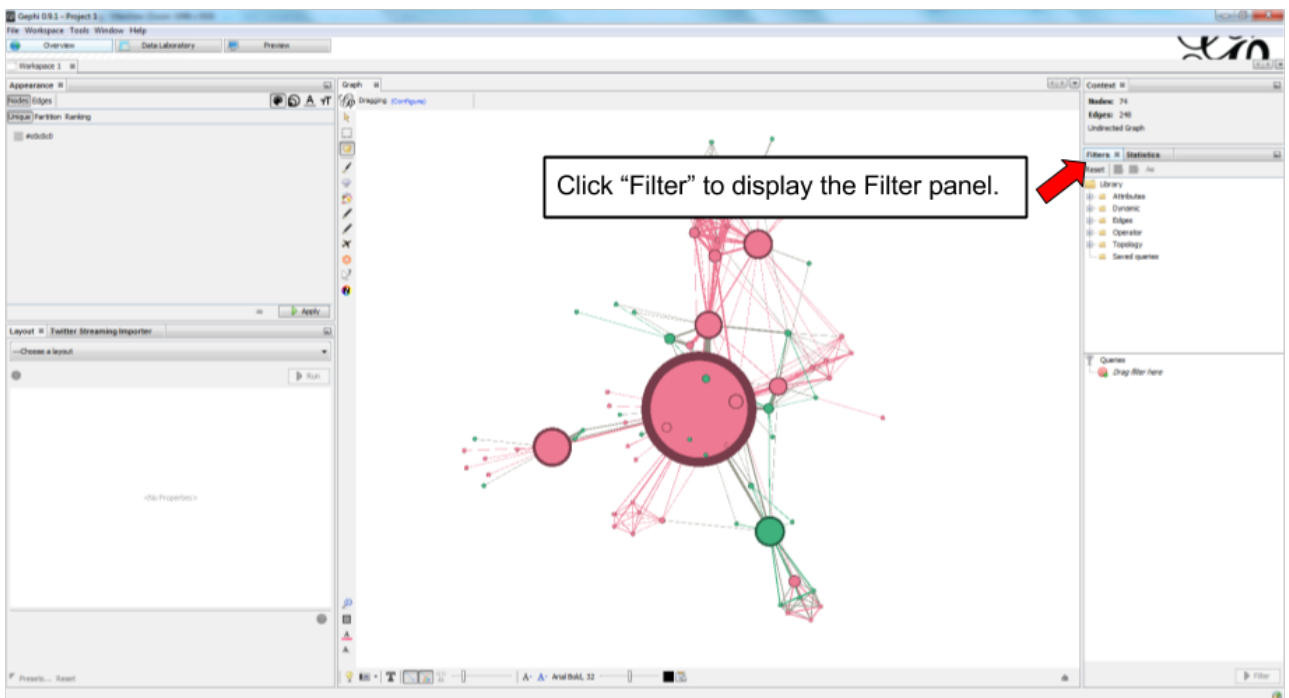


Figure 7. Making the Filter panel visible.

How the Filter panel works:

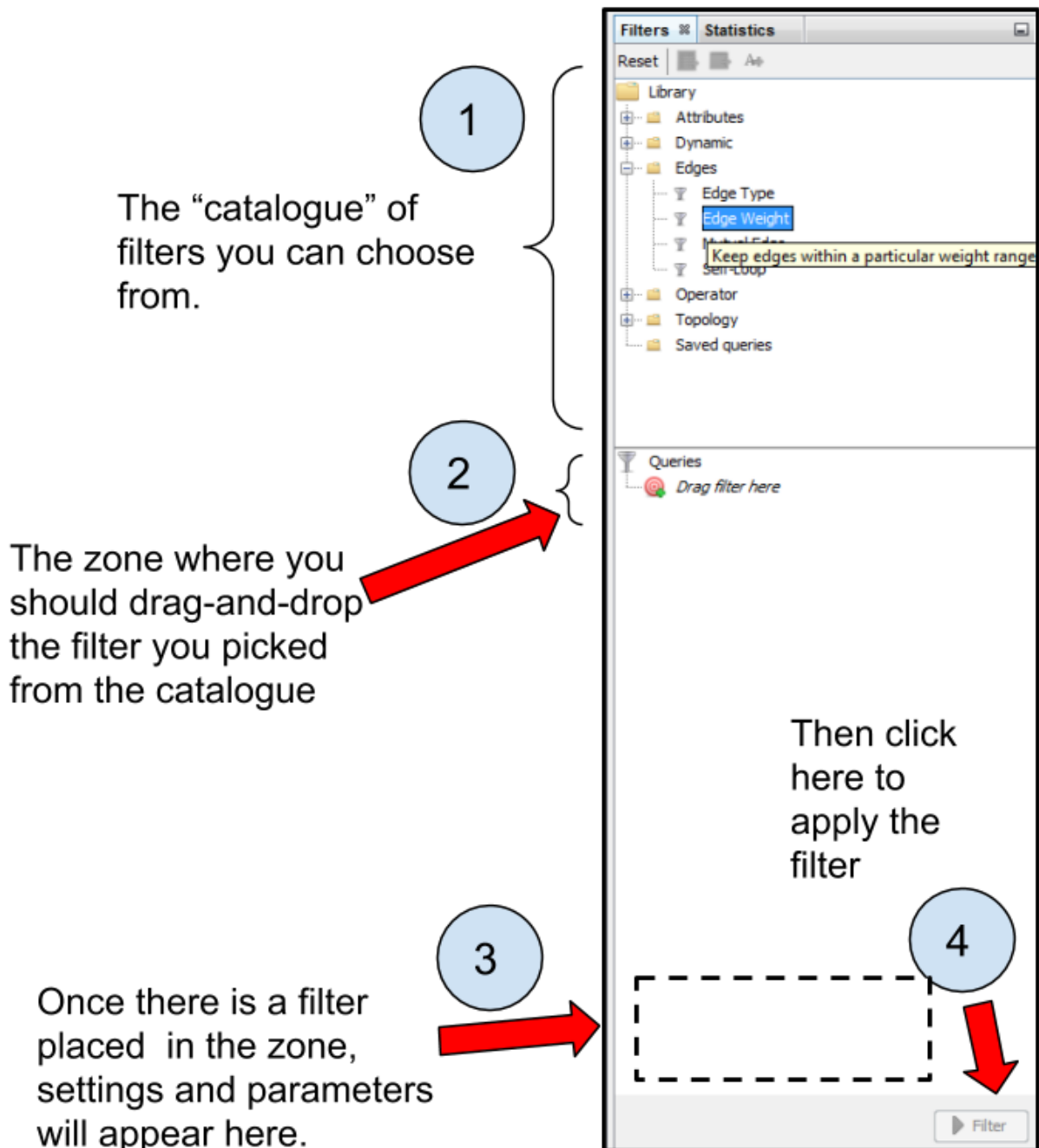


Figure 8. Workflow of filters

An example: hiding edges with weight lower than 2



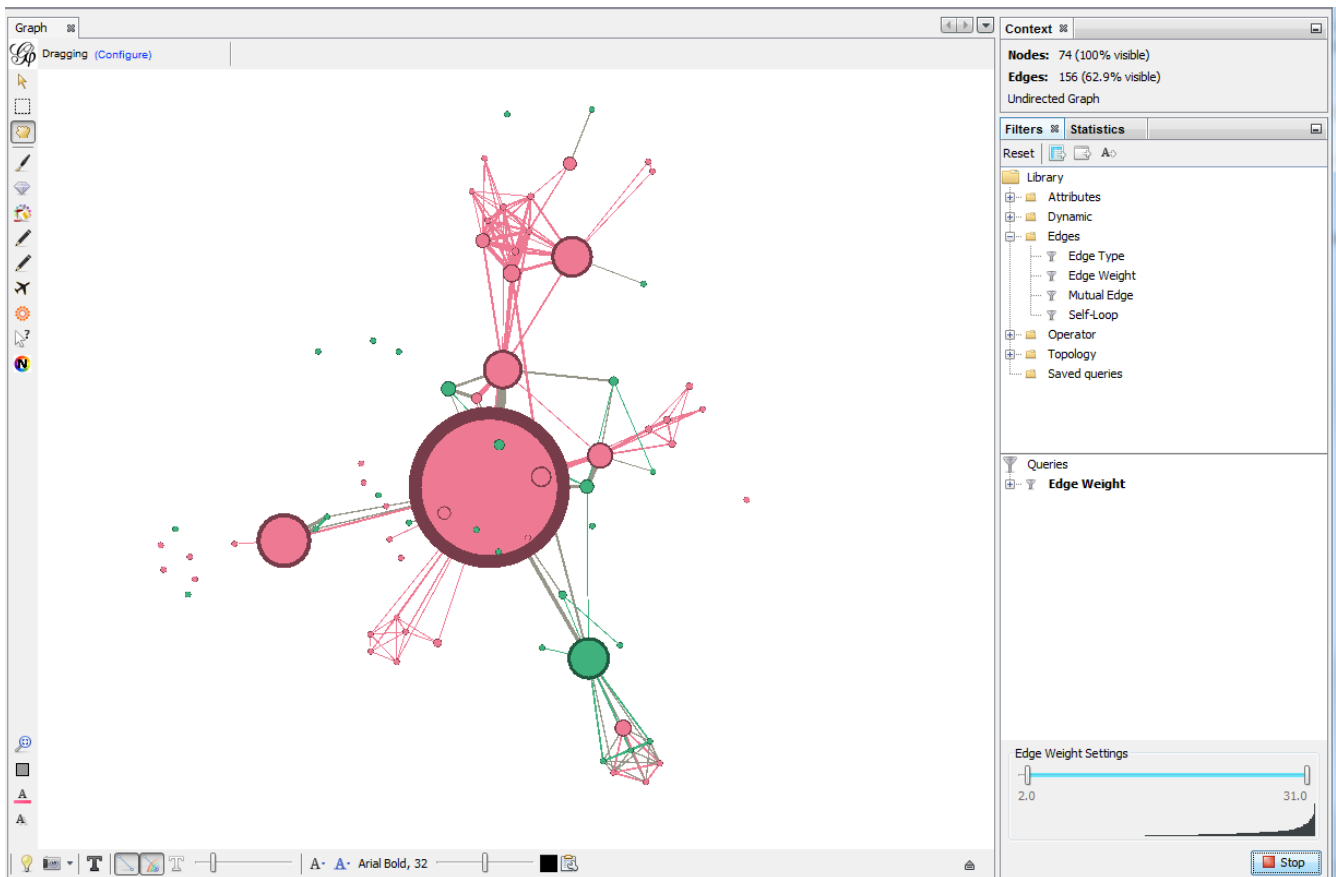


Figure 9. Filtering out edges with weight lower than 2.

[view online animation](https://tinyurl.com/gephi-tuto-2) - link: <https://tinyurl.com/gephi-tuto-2>

When you are finished using a filter in the zone, right click on it and select "remove".

## combining 2 filters

One filter is applied AFTER this other:

**The first filter to be applied is NESTED (placed inside) the second one as a "subfilter"**

Which filter should be placed inside which? Let's look at different examples:

### 1. Case when the placement of filters makes no difference

Goal: Keeping on screen only the female characters which have a tie (an edge, a relation) of at least strength 2.

→ place the filter "edge weight" inside the filter "Gender":

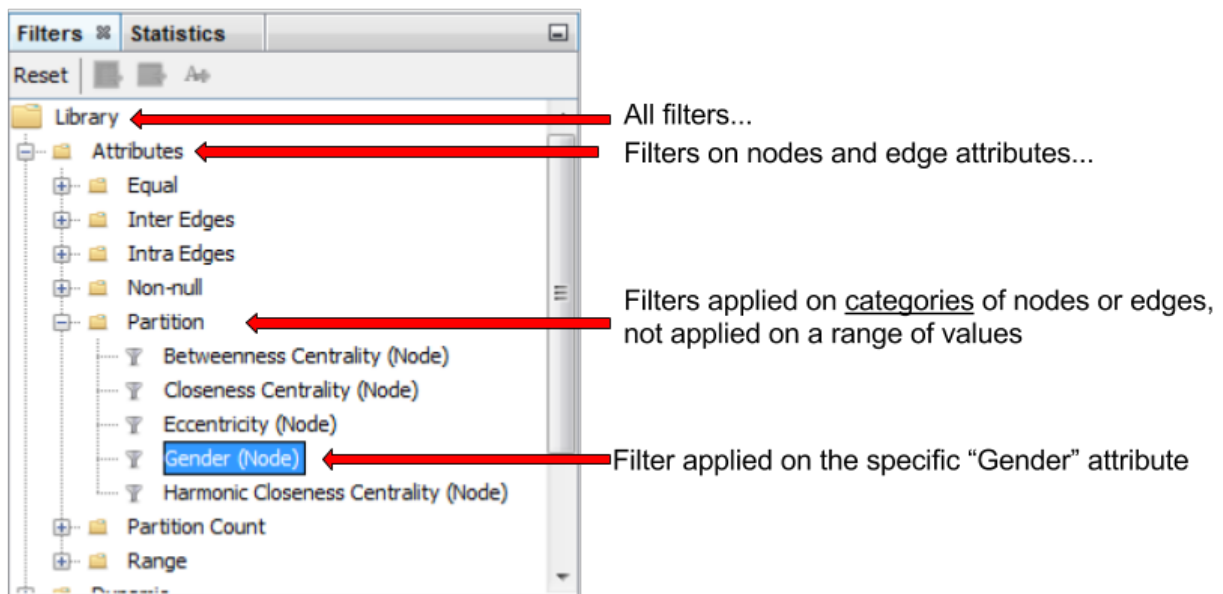


Figure 10. Filter on the Gender attribute

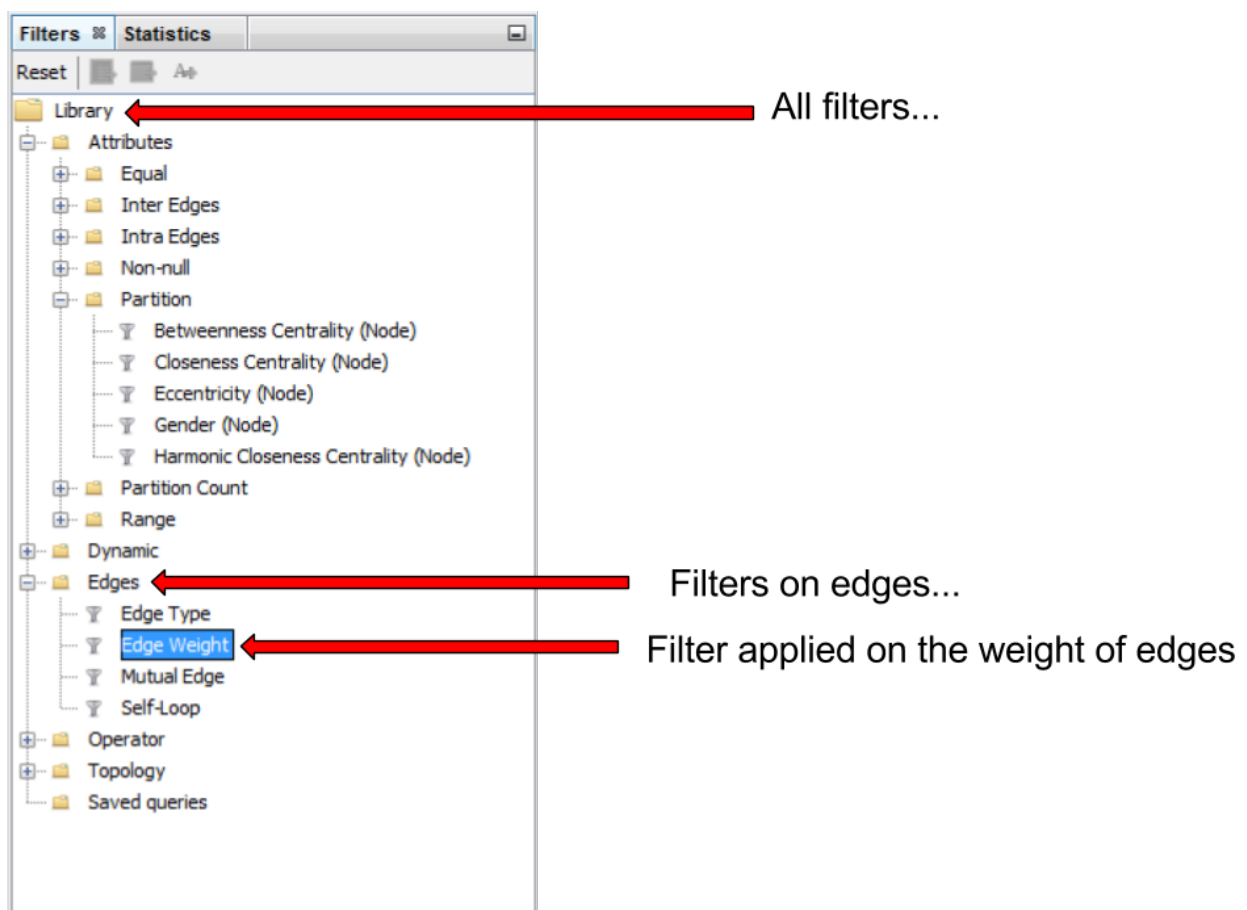


Figure 11. Filter on edge weight

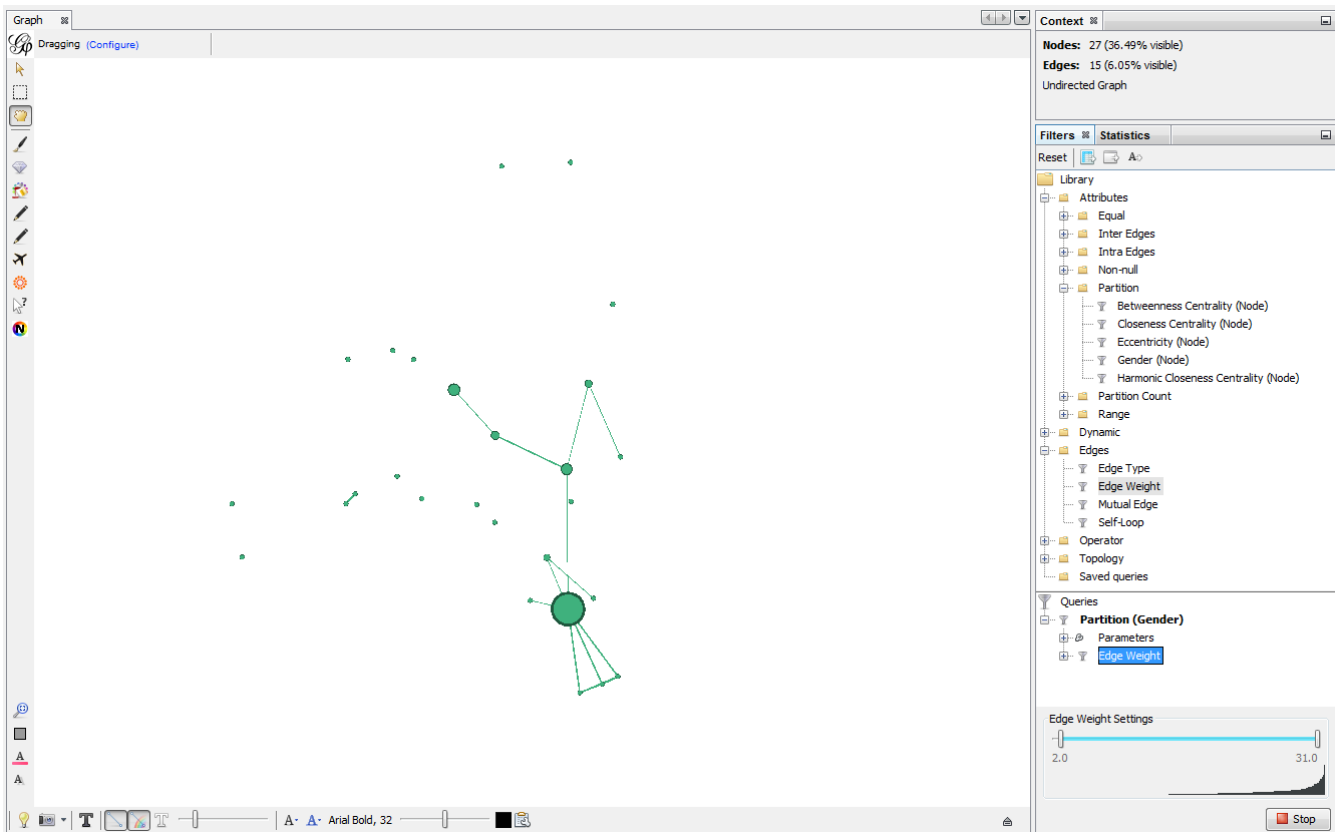


Figure 12. Keeping only female characters with at least 2 ties

view online animation - link: <https://tinyurl.com/gephi-tuto-1>

In this case, it was equivalent to:

- nest the "Gender" filter inside the "Edge weight" filter or
- nest the "Edge weight" filter inside the "Gender" Filter

→ The result was the same (the network on screen is identical in both cases)

## 2. Case when the placement of filters makes a difference

Here, we want to visualize:

- only the nodes which have **less than** 10 relations <1>
- and among these, only those which form the "main island" of the network (we want to hide small detached groups of nodes) <2>

① in technical terms, nodes with a **degree** of less than 10.

② in technical terms, we are looking for the **giant component**

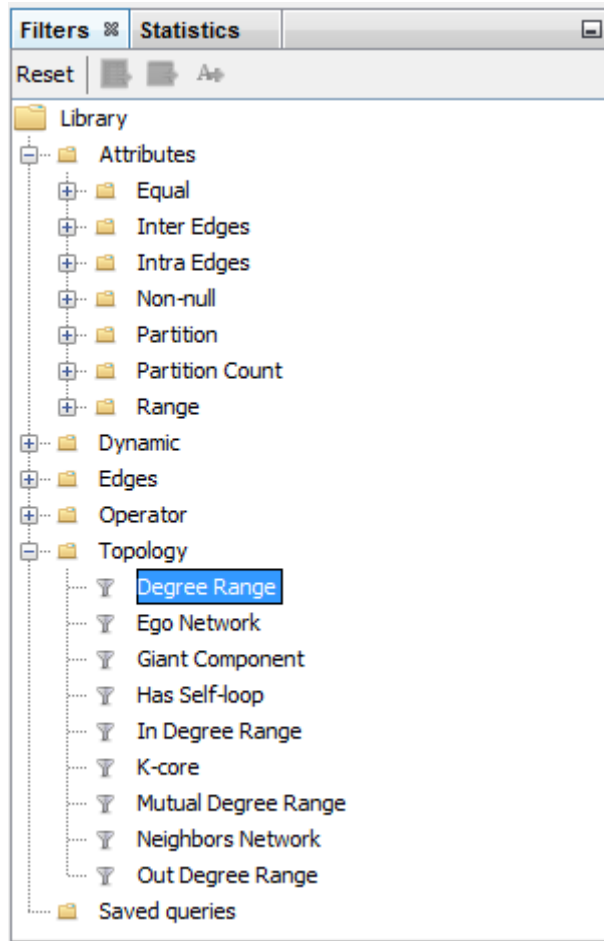


Figure 13. Filter on degree

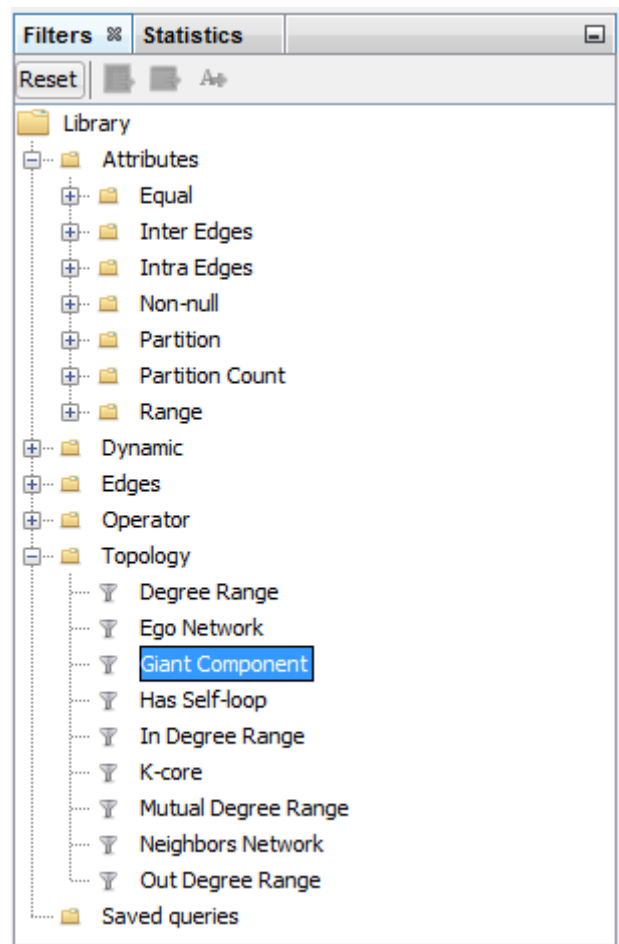


Figure 14. Filter on giant component

We will see that the placement on the filters in the zone will make a difference.

First, let us place the filter on giant component **inside** the filter on degree:

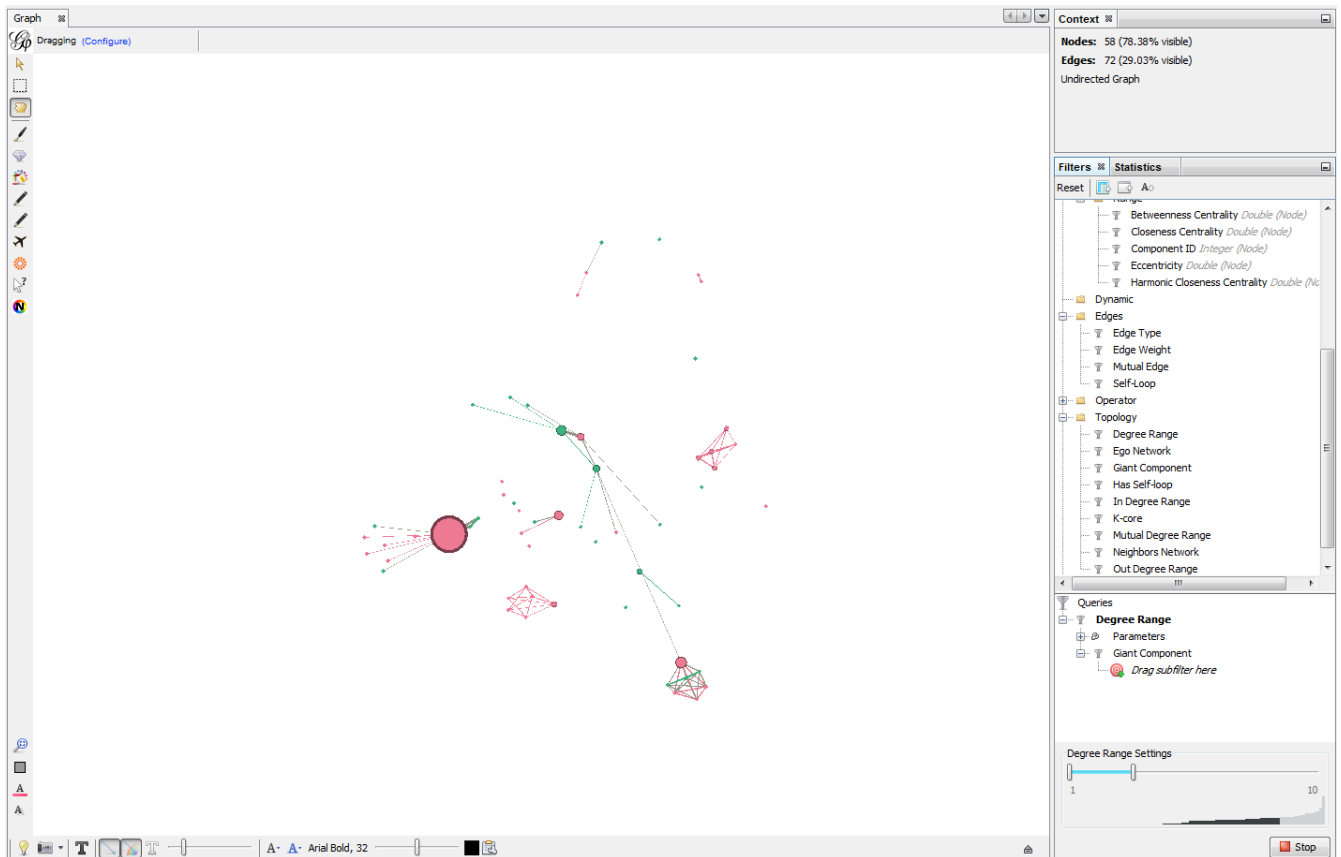


Figure 15. Filters in one configuration

In this first case,

- only the giant component of the network was made visible.
- Since the network was just one big connected "island" to start with, it did not change a thing.
- then, all characters with more than 10 relations were hidden
- this hides nodes which were connecting with many others, so that we end up with many groups, disconnected from each others.

Now instead, placing the filter degree **inside** the filter on giant component:

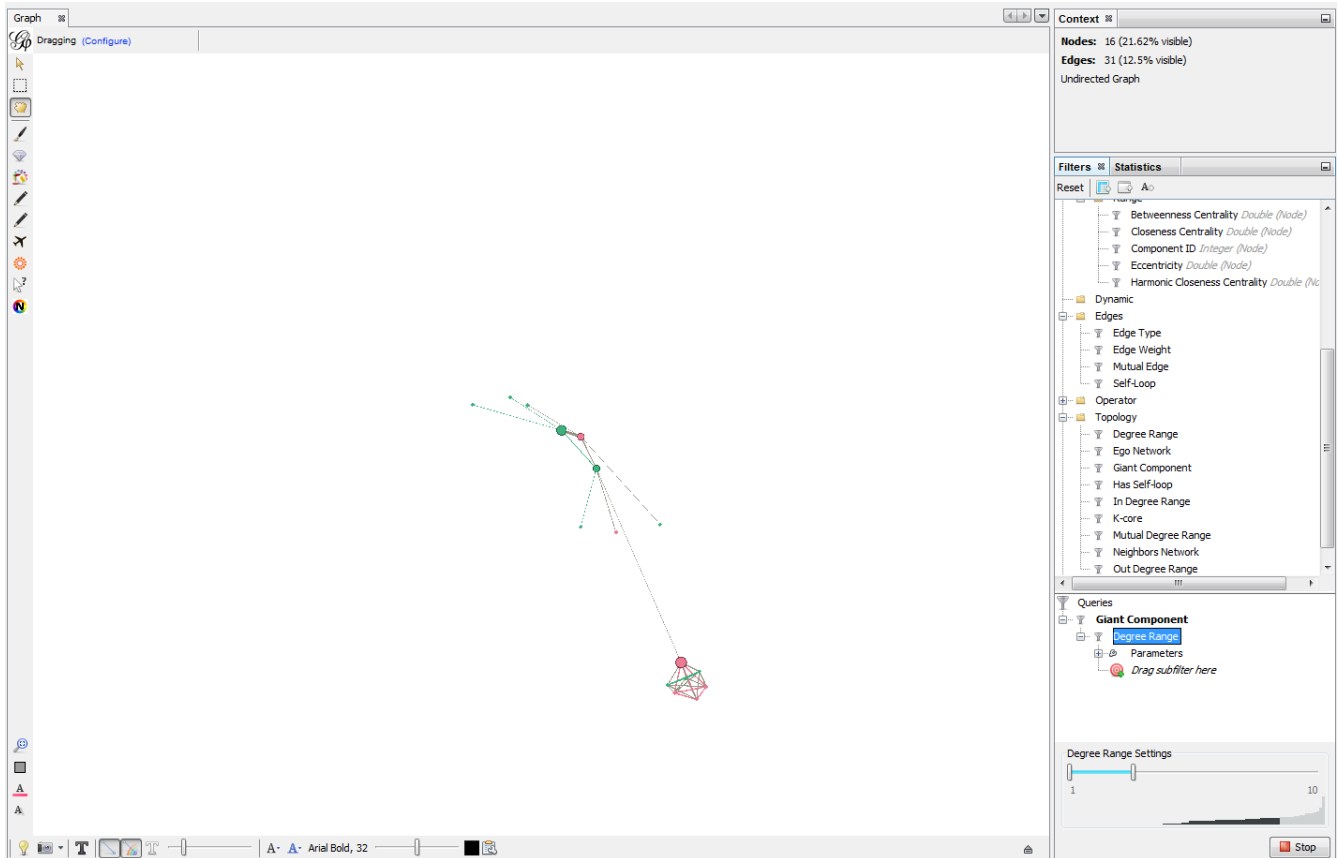


Figure 16. Same filters in another configuration

In this second case,

- starting from the complete network, all characters with more than 10 relations where deleted.
- this created a network made of many disconnected groups of nodes
- then the giant component filter is applied,
- which had for effect to hide small groups, to keep in view only the biggest group of connected nodes.



In summary: be careful how you apply several filters at once, this might have an effect on the logic of filtering.

# filter operators

## 1. The MASK operator

Imagine you are interested in the female characters of the novel "Les Miserables".

- you are interested in female characters and the relations among them
- you are interested in the relations between female characters and male characters
- you are **not** interested in the relations between male characters

How to display this?

The MASK operator applied on the gender partition filter enables you to:

- show all characters
- relations between female characters
- *and relations between male and female characters*
- *but masking male-male relations*

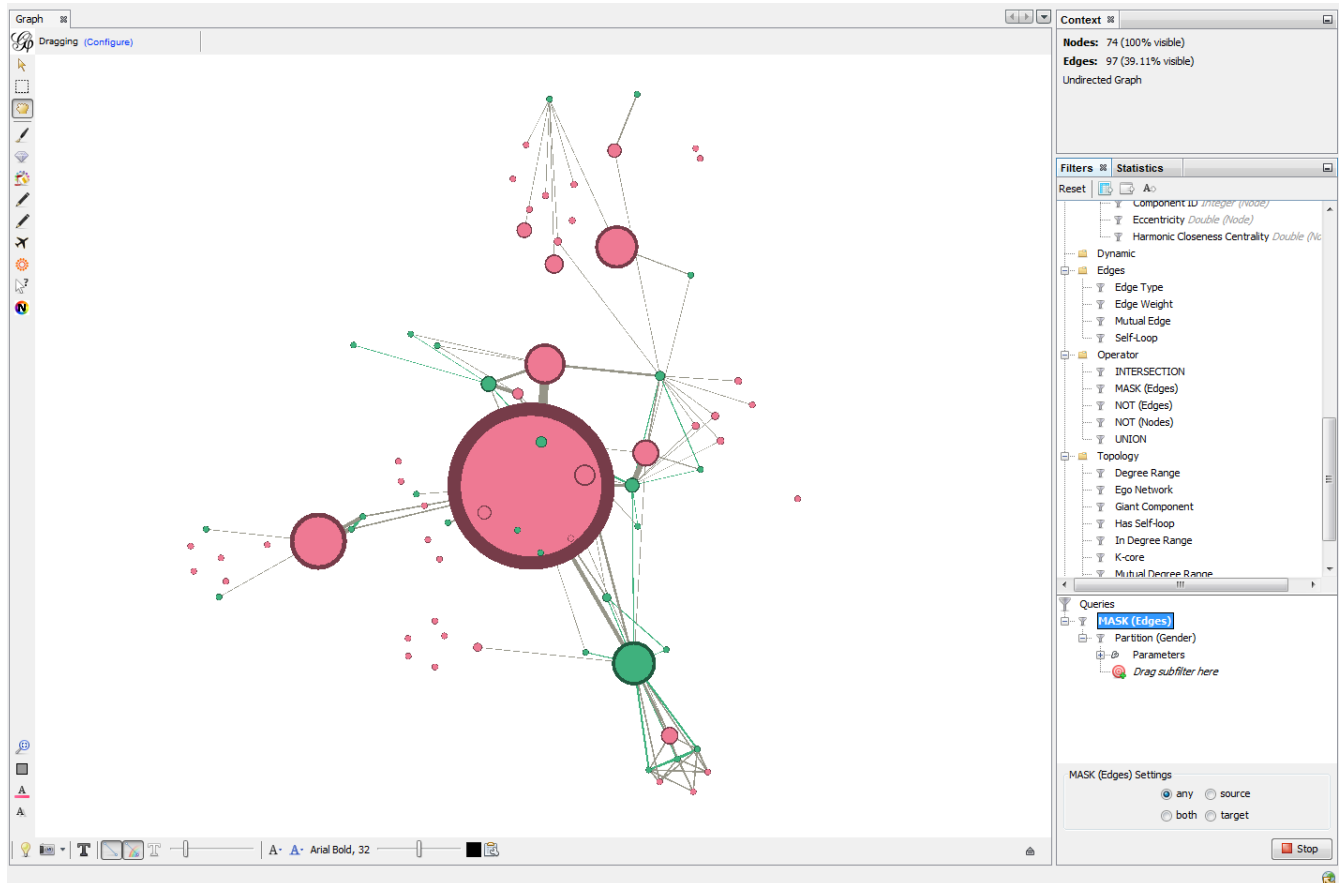


Figure 17. Using the MASK operator

It is also possible to hide / show only some of the directed relations between the visible graph and the filtered out graph:



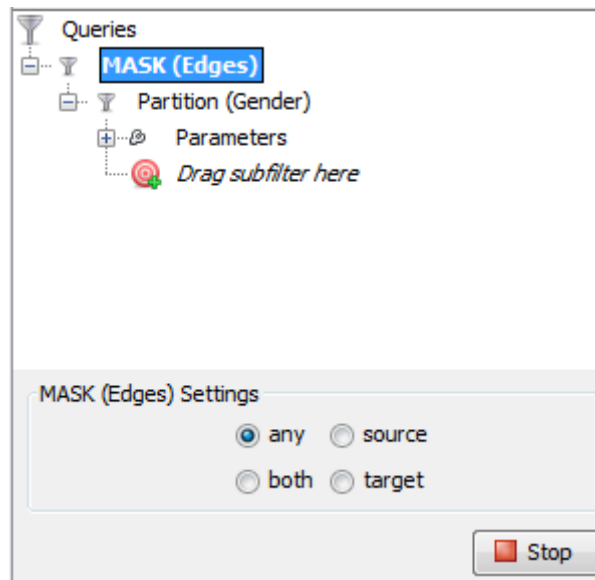


Figure 18. Parameters of the MASK operator

## 2. The UNION operator

Imagine you are interested in the characters with names starting with "L" or "J" in "Les Miserables".

How to display only these characters?

We will need to apply filters on the **Label** of the nodes, which contains the names of the characters.

However, looking at the "catalogue" of filters, we see no filter on **Label**. The reason is that **Label** is an internal property of nodes, inaccessible to filters.

So we must first copy the Labels of the nodes in a new attribute, which we will be able to apply a filter on.

Let's switch to the data laboratory and add this attribute:

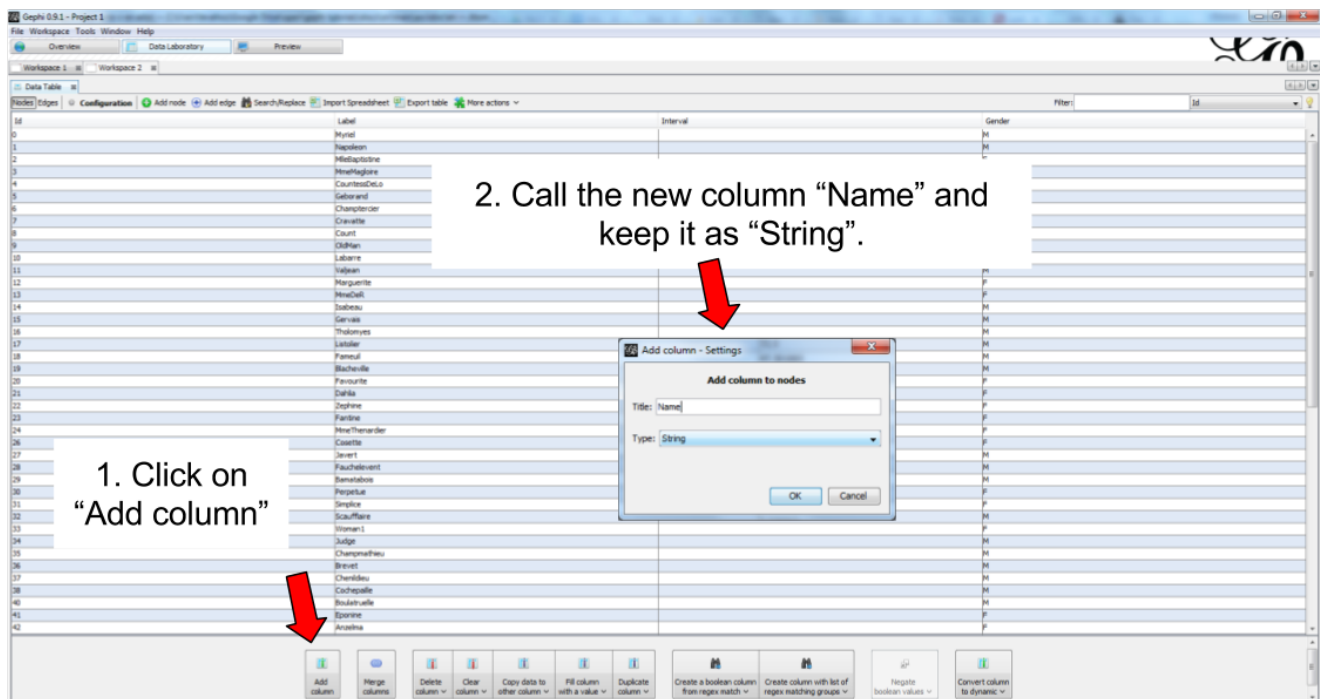


Figure 19. Adding a column for Names

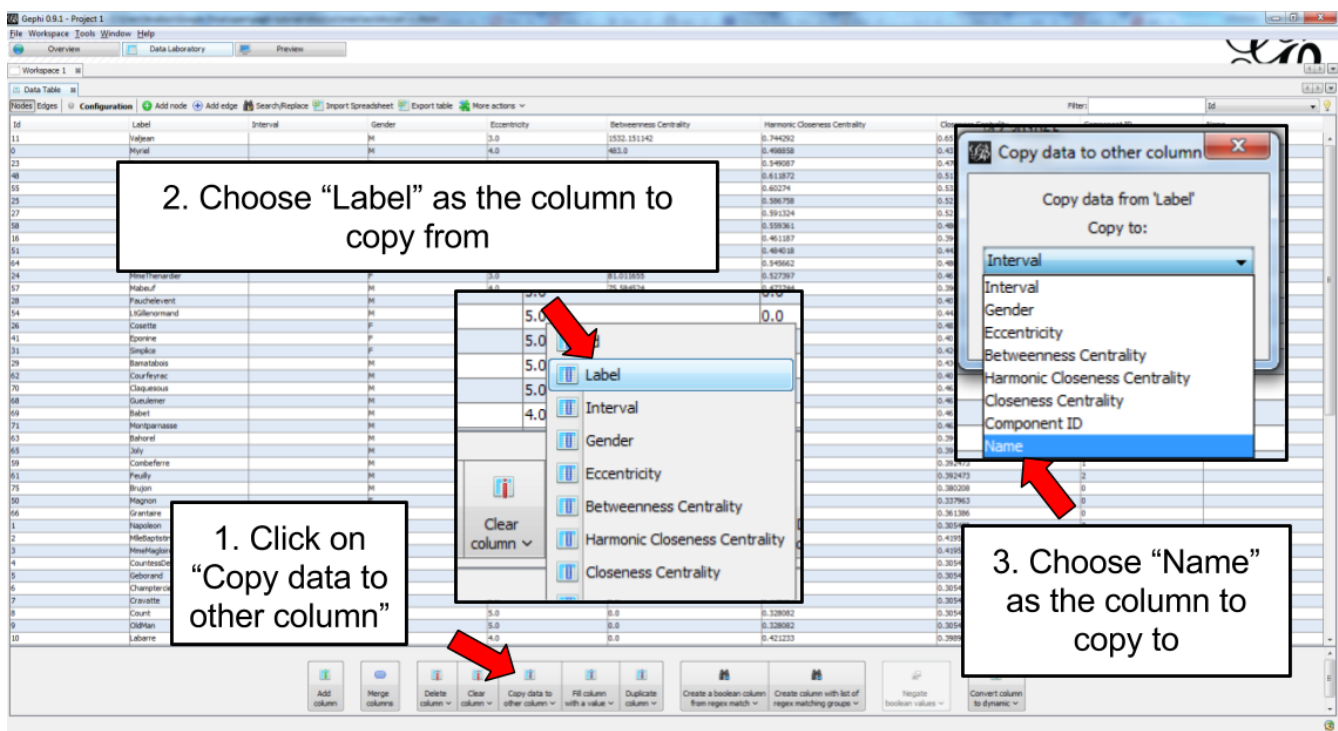


Figure 20. Copying to this new column

We now have an attribute called "Name" that we can find in the Filters:

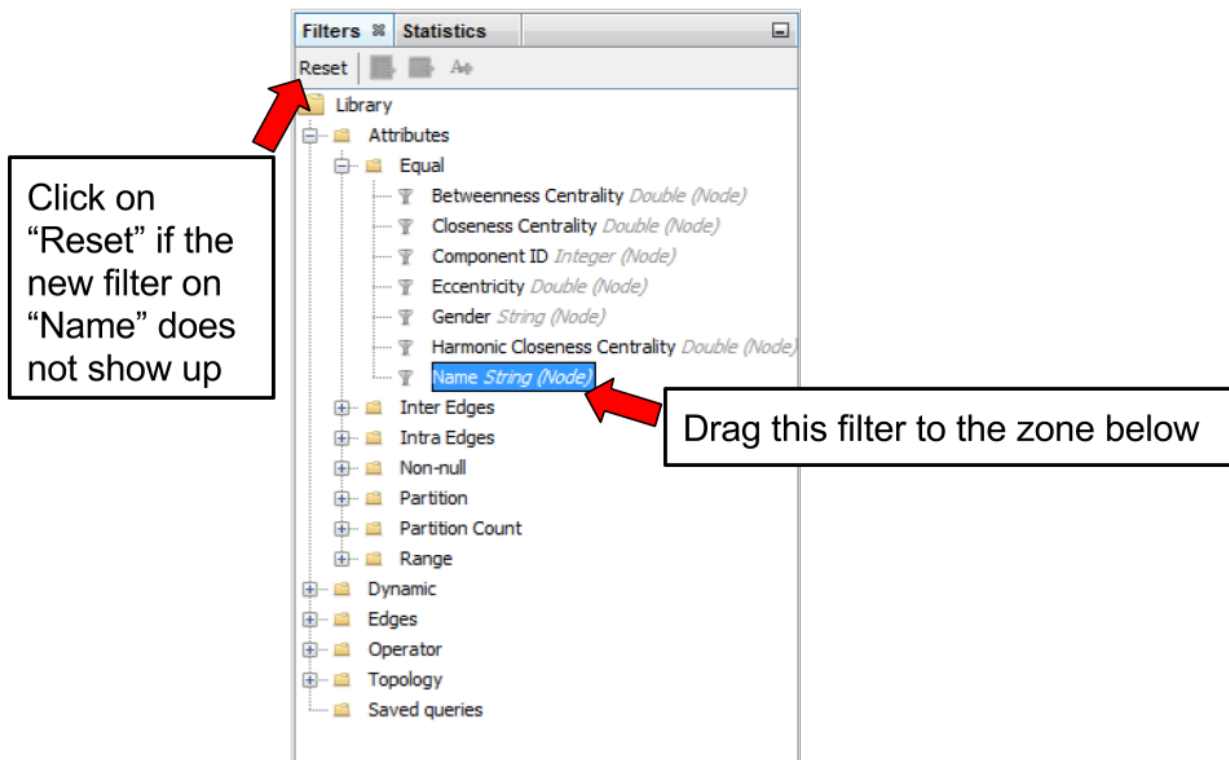


Figure 21. New filter available

This is how the filter on Name and its parameters look like in the zone:

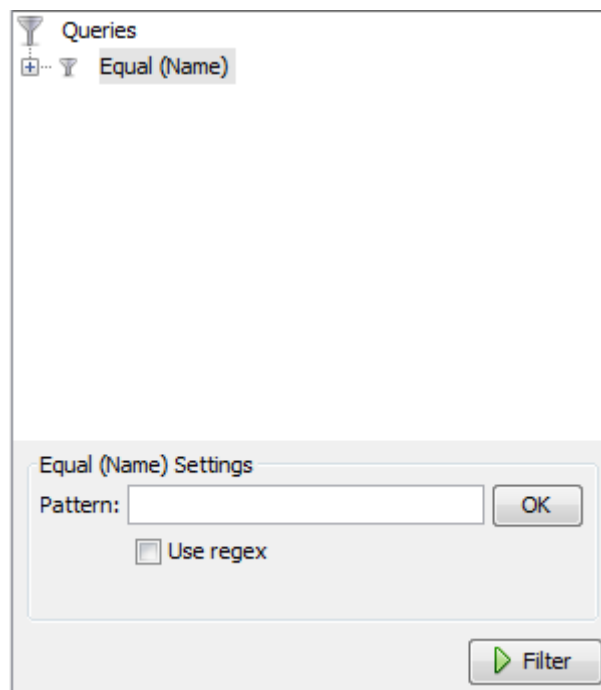


Figure 22. Name Filter

To recall, we want to show only the characters which name start with "L" or "J". Let's start with the "L" characters.

We need to find the names which match the pattern **Start with an L**. The way to describe a pattern in text is called a "regular expression".

Said differently, *a regular expressions (also called "regex") is a convenient way to express a pattern we search for in a text.*

Regular expressions can become very sophisticated. But here, we need just a simple one:

```
L.*
```

Let's examine what the L, the dot and the star mean.

- the letter "L" means we want names starting with this first letter
- . the dot means: any character
- \* the star means: the previous character, repeated any time.

So: "select nodes which have a name starting with L, followed by any character, in any number"

Please note that you need to check the box "regex":

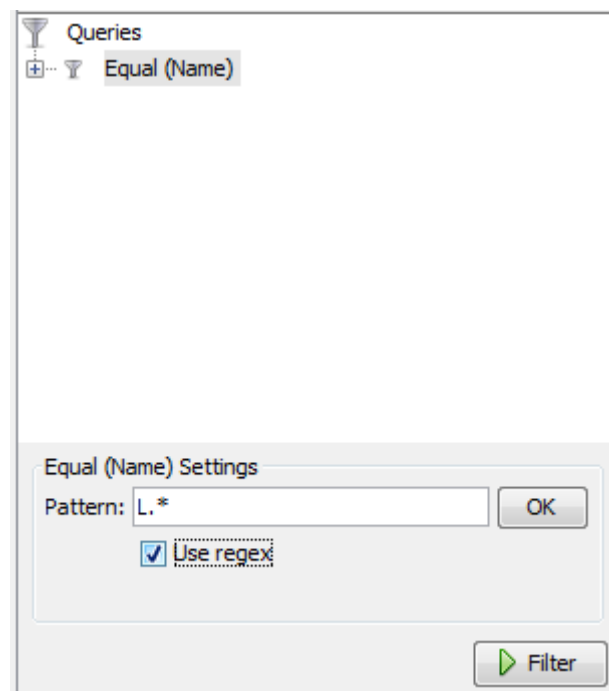


Figure 23. Using a regular expression in a filter

When the filter is applied, only the characters wit a name starting with L will be displayed:

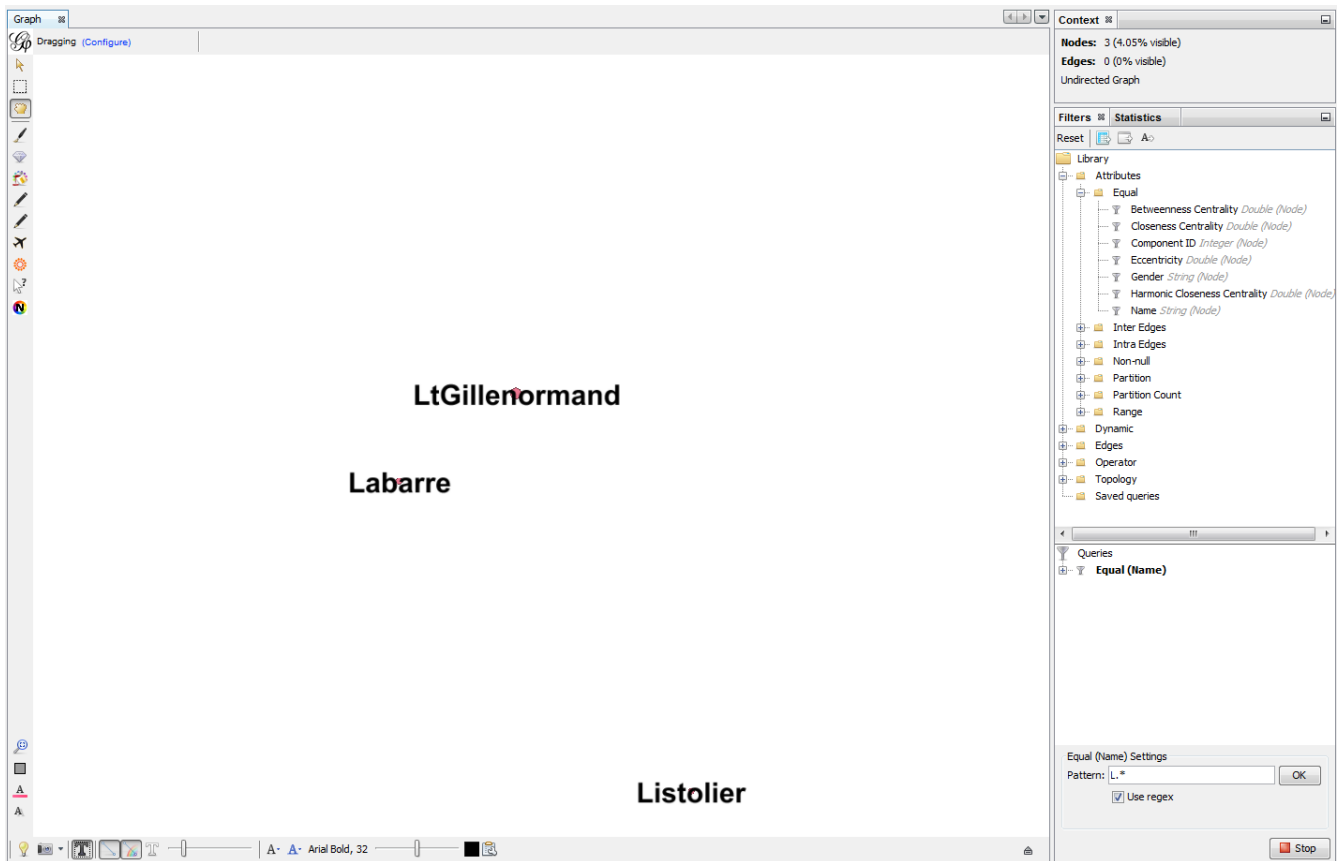


Figure 24. Using a regular expression in a filter

How to filter characters with a name starting with the letter "L" or "J"?

We could rely on a more complex regular expression to do this:

```
[LJ].*
```

Meaning: "select nodes which have a name starting with L or J, followed by any characters"

But we can also rely on 2 filters: one for L, one for J. Nesting one inside another would not work, it would mean:

"show nodes which start with an L, and among them, only those which start with a J"

→ no node can meet this condition, so they would all be invisible.

Instead, we should use the **UNION** operator that can be found here:

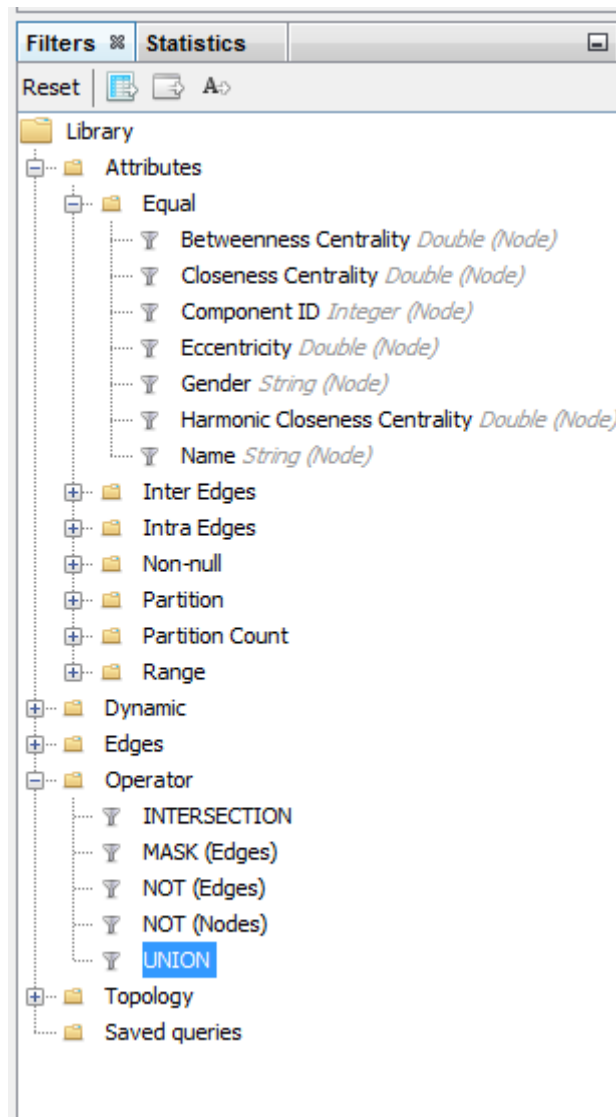


Figure 25. The UNION operator in filters

Drag it to the zone, and then drag inside it twice the **Attributes** → **Equal** → **Name** filter:

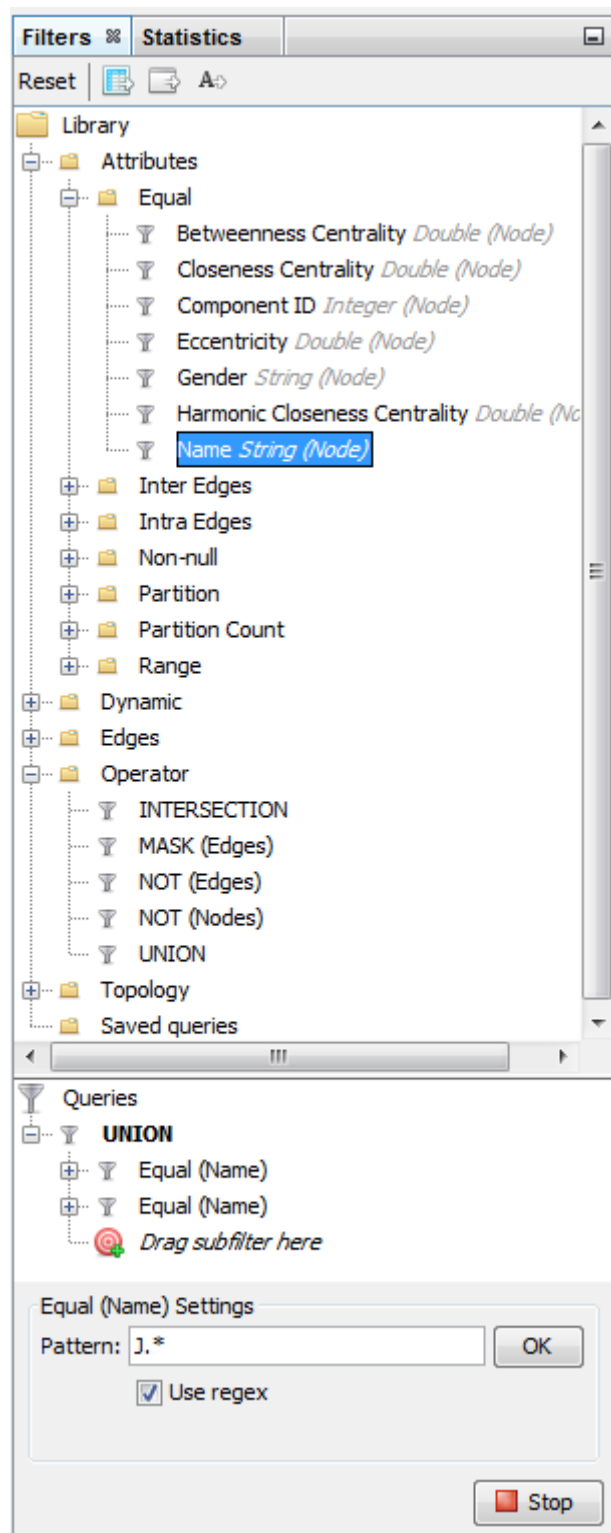


Figure 26. The UNION operator and 2 subfilters

In the settings of the first Name filter, put the regular expression:

`L.*`

In the second Name filter, put:

J.\*

(make sure the "regex" box is checked in both cases)

As a result, the nodes selected by both filters are added up in the display:

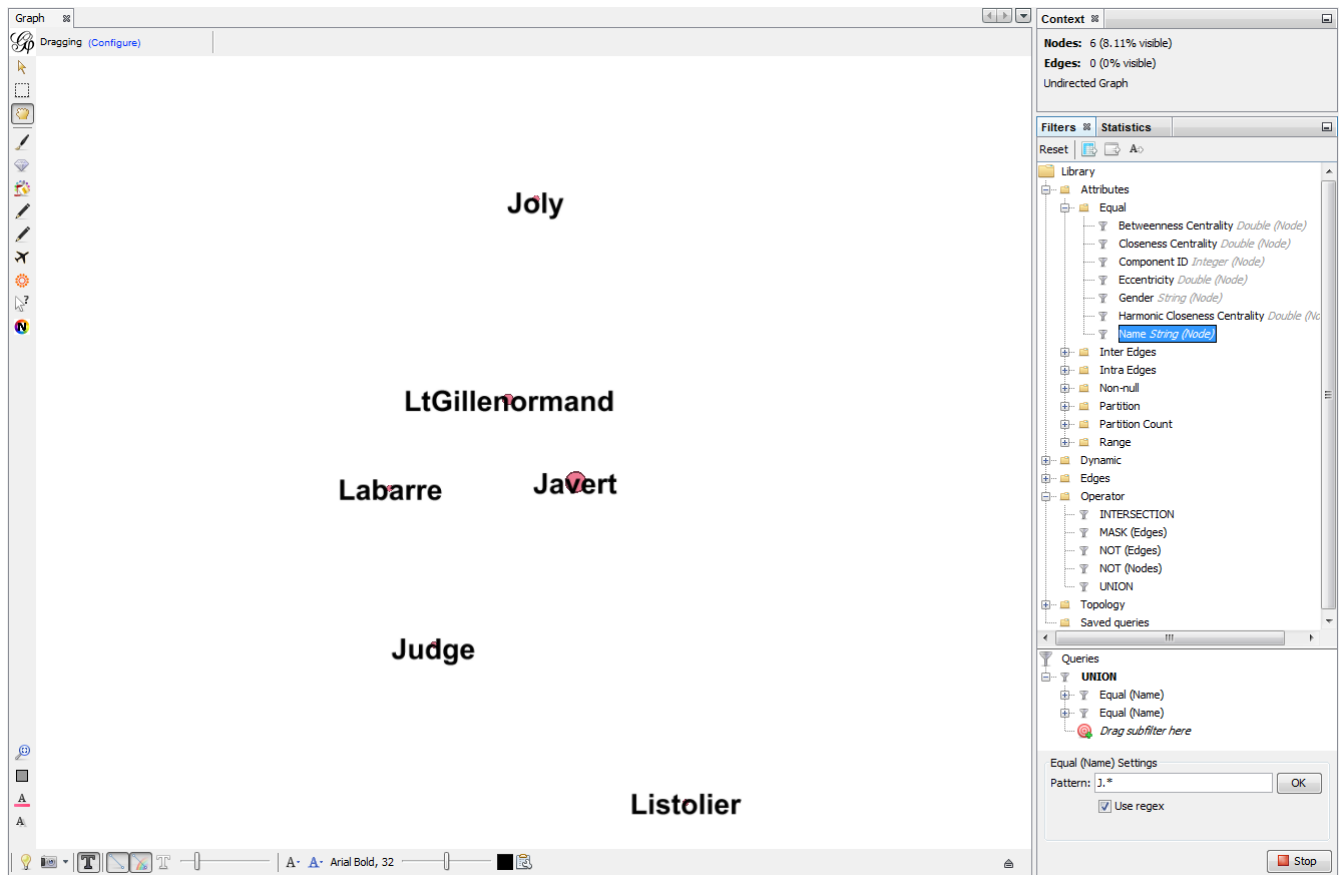


Figure 27. The UNION operator and 2 subfilters

### 3. The NOT operator

The NOT operator flips the result of a filter: what was hidden becomes visible and vice versa.

Example: if we want to display all characters except for those returned by a UNION on 2 Name filters on L and J initials:



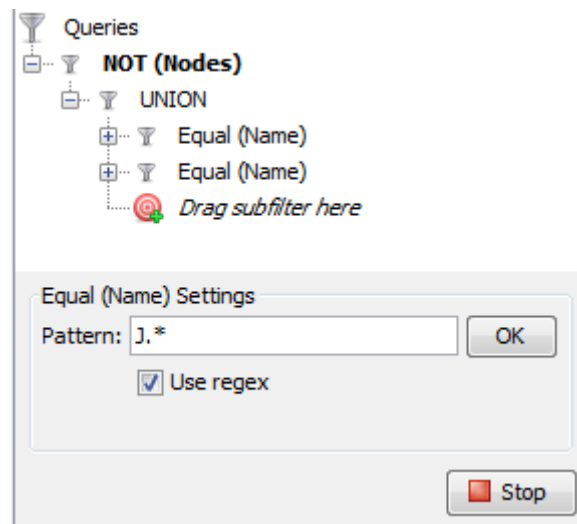


Figure 28. The NOT nodes operator - 1

Same effect, but applying the NOT operator on single filter using a regex on L or J:

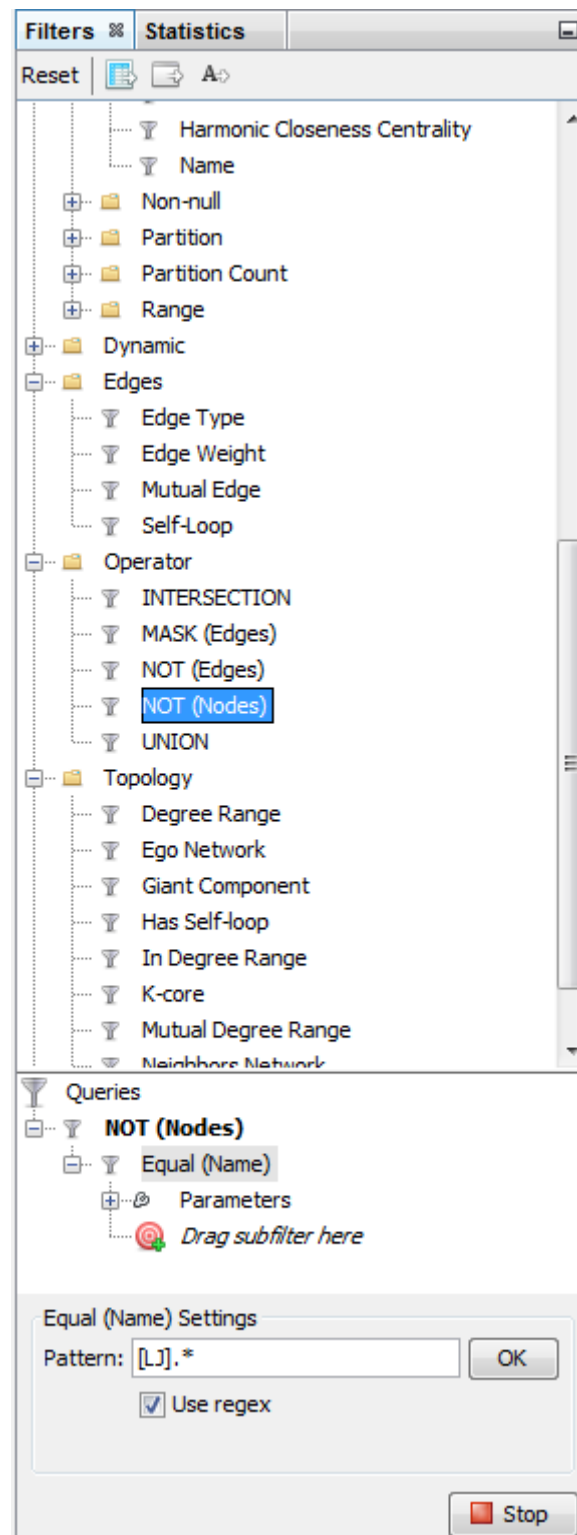


Figure 29. The NOT nodes operator - 2

Same effect again, achieved without using the NOT operator. In regular expressions the ^ sign inside square brackets means "NOT":

`[^L].*`

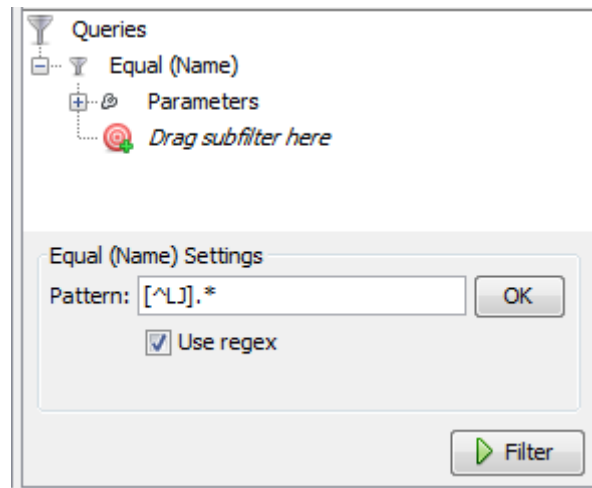


Figure 30. Achieving a NOT effect with regex

Tutorials about regular expressions:

- <https://regexone.com/>
- [http://www.themacroscope.org/?page\\_id=643](http://www.themacroscope.org/?page_id=643)

And a web page where you can test your regular expressions: <http://regexpal.com>

## more tutorials on using filters in Gephi

- [Video on using filters by Jen Golbeck](#)

## the end

Visit [the Gephi group on Facebook](#) to get help,

or visit [the website for more tutorials](#)