Using filters

Table of Contents

Gephi workshops	1
download a network file for practice	1
open the network in Gephi.	2
getting a sense of the attributes in the data laboratory	4
discovering the filter panel	6
combining 2 filters	9
filter operators	5
more tutorials on using filters in Gephi	7
to go further	7

last modified: 2023-04-10



Gephi workshops

I organize online workshops and personalized trainings for Gephi, for beginners and experts. To schedule one or to get more information: analysis@exploreyourdata.com.

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].

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

Welcome		×
Welcome to Gephi		Gø
Open recent	New Project	
🧐 first graph.geph	New Project	
	Open Graph File	
	Samples	
	Les Miserables.gexf	
	Java.gexf	
	Power Grid.gml	
Open at startup		

Figure 1. welcome screen

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

M Import report		×			
Source: miserables.gexf					
Issues Report					
Nodes		Issues			
GEXF version	1.3	INFO			
Graph Type: Undir	ected	More options			
# of Nodes:	74	New graph			
# of Edges:	248	Append Graph			
Dynamic Graph:	no				
Dynamic Attributes	a: no				
Multi Graph:	no				
		OK Cancel			

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:

Inde # Add nor Inde # Add nor	de 🛞 Add edge 🏙 Search Rie; I	place 📧 Import Spreadsheet 🖭 E	port table 🙀 More actions 🗸				
Table # Add not Labor Hyper Configuration & Add not Labor Hyper Hy	de 🛞 Add edge 🏙 Search,Rep 1	place 🐮 Import Spreadsheet 🐮 E	port table 🙀 More actions 🛩				
(ets) © Coefiguration © Add no Libbl Physics Pieces Pieces	de 🕣 Add edge 🎁 Search (Rep 1	place 🐮 Import Spreadsheet 🐮 E	port table 🛛 👬 More actions 🗸				
Labe Myriel Piloto Michol Count	8	Interval				Filter:	ld
Hyrie Napol Mileki Count		ATTENT THE	Gender	Eccentricity	Betweenness Centrality	Harmonic Goseness Centrality	Closeness Centrality
Napol Micto Minek Count			м.	4.0	483.0	0.496858	0.407126
Mitthe Market	eon .		N	5.0	0.0	0.328082	0.305439
Heek Court	aptistine		P	4.0	0.0	0.450913	0.41954
	Angloire		F	4.0	0.0	0.450913	0.41954
<u></u>	terralial a		· · · · · · · · · · · · · · · · · · ·	5.0		0.328082	0.305429
<u><u> </u></u>				5.0	Betweenness Centrality	0.328082	0.305439
Click on			™I Gender	5.0	betricerinebb berribbilty	0.328082	0.305409
			м	5.0	0.0	0.328082	0.305439
			м	5.0	0.0	0.328082	0.305439
"Eddes" to			м	5.0	0.0	0.328082	0.305439
			м	4.0	0.0	0.421233	0.398907
eaa tha liet	of		м	3.0	1532.151142	0.744292	0.657658
			P	4.0	0.0	0.4%347	0.41954
Education the			P	4.0	0.0	0.421233	0.398907
Eddes in tr	1e		M	4.0	0.0	0.421233	0.398907
			M	4.0	0.0	0.421233	0.398907
network			M	4.0	136-276976	0.461187	0.394995
network			M	4.0	6.0	0.402968	0.547629
(da da	- 4-		C	4.0	0.0	0.402968	0.30613
(MAU)	evee		C	10	0.0	0.702968	0.00013
Parto	La ca			4.0	0.0	0.402968	0.50519
Jacki	1		-	4.0	0.0	0.403468	0.047648
Facto				4.0	160 130276	0.540047	0.470668
Plane T	henarder			3.0	81,011655	0.527397	0.467949
Casel	te.			3.0	47.18817	0.534247	0.483444
Jacow	1		N	3.0	141.49012	0.591324	0.521429
Paul	valevent		M	4.0	72.5	0.499064	0.407821
Bana	tabois		м	4.0	22.916667	0.487443	0.434524
Perpr	due .		P	5.0	0.0	0.354338	0.323009
Simple	C.M			4.0	23.491508	0.460046	0.424429
Scauf	Maire		M	4.0	0.0	0.421233	0.398907
Work	en l		2	4.0	0.0	0.428082	0.401099
Judge			м	4.0	0.0	0.455479	0.410112
Charg	pnatkeu		м	4.0	0.0	0.455479	0.410112
Breve	et i i i i i i i i i i i i i i i i i i i		je e	4.0	0.0	0.455479	0.410112
Cheni	édeu -		м	4.0	0.0	0.455479	0.410112
Cache	epañe		м	4.0	0.0	0.455479	0.410112
Boulat	truele		м	4.0	0.0	0.365297	0.34434
Eponie	re		,	4.0	33.628408	0.477169	0.401099
Annel	ma		P	18.0	0.0	0.386986	0.354369

Figure 5. Nodes attributes.

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



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. How to use filters.



Figure 10. Filtering out edges with weight lower than 2.

view online animation - 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 11. Filter on the Gender attribute



Figure 12. Filter on edge weight



Figure 13. Keeping only female characters with at least 2 ties



Figure 14. 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
- \rightarrow 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 15. Filter on degree



Figure 16. 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 17. 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 where hidden

 \rightarrow 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 18. Same filters in another configuration

In this second case,

- starting from the complete network, all characters with more than 10 relations where deleted.
- \rightarrow this created a network made of many disconnected groups of nodes
 - then the giant component filter is applied,

 $\rightarrow\,$ 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 19. 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:

T Queries	
🗄 🖓 MASK (Edges)	
🗄 🐨 🝸 Partition (Gender)	
MASK (Edges) Settings	
any source	
🔘 both 🔘 target	
	Stop

Figure 20. 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 21. Adding a column for Names

🕼 Gephi 0.9.1 - Project 1	The second second	and the set of the set of the set of the	- Martin Color of California		a second as	the state of the s		Second and the second s	
File Workspace Tools Win	dow <u>H</u> elp								
Overview	Data Lat	oratory Preview							Y ZA
Workspace 1 #								\sim	
🗈 Data Table 📧	1.00								
Nodes Edges O Configura	etion 🔂 Add no	de 💮 Add edge 📸 Search/Replace 🖭 Import Spreadsheet 💡	Export table 🕌 More action	s ~				Filteri	•. 9
Id	Label	Interval Gender	Eccen	ricky	Betweenness Cen	raity Harmonic Closeness Cent	traity Close		
11	Valjean	м	3.0		1532-151142	0.744292	0.65	1771 A 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	-
0	Myriel	м	4.0		483.0	0.498858	0.43	Copy data to other column	
48						0.545087	0.47		
55			12	1.		0.60274	0.53		
25	2	. Choose "Labe	i as the	e coll	imn to	0.586758	0.52	Copy data from 'Label'	
27	. –					0.591324	0.52		
58		000	(from			0.559361	0.48	Copy to:	
16		COP	y nom			0.461187	0.39		
64		• •				0.545662	0.40	Interval -	
24	MmeThenardie	, , , , , , , , , , , , , , , , , , ,	3.0		81.011655	0.527397	0.46		
57	Mabeuf	м	4.0	1.1.1.1	25 584524	0.471744	0.39	Interva	
28	Fauchelevent	м				0.0	0.40	Candar	
54	LtGillenormand	M		5.0		0.0	0.44	Genuer	
41	Econice			5.0	M.		0.40	Eccentricity	
31	Simplice	4		5.0	P		0.40	Potucoppor Controlity	
29	Bamatabois	м		5.0			0.40	Detweenness Centrality	
62	Courfeyrac	м			Label	-	0.40	Harmonic Closeness Centrality	
70	Claquesous	M		5.0			0.46	Classeners Controlity	
49	Babet			40	Interval	-	0.46	Closeness Centrality	
71	Montparnasse	M		4.0			0.46	Component ID	
63	Bahorel	м		T	Gender		0.39	Name	
65	Joly	м					0.39	Name	
59	Conbeferre	M			Eccentricity		0.2924		
75	Bruten	E Contraction of the second se	ii		cecementy		0.3802		
50	Magnon			· ·	Petusenner	Controlity	0.3379	0	
66	Grantaire		ci		betweenness	Centrality	0.3613	м р	
1	Napoleon	1 Click on	Clea	r 👘			0.3054		
3	MieBaptistin	T. CIICK ON	colum	ı 🗸 📔 🛄	Harmonic Clo	seness Centrality	0.4193		
4	CountessDe	" ~ • • • •					0.3054	3 Choose "Name	"
5	Geborand	"Conv data to			Closeness Ce	ntrality	0.3054	0. 0110030 Maine	
6	Champtercia	oopy data to					0.3054	as the seluments	
7	Cravatte			1			0.3054	as the column to	
9	OldMan	other column	5.0		0.0	0.329082	0.3054		
10	Labarre		4.0		0.0	0.421233	0.3989	conv to	-
									A
			12 12			-			
			4 4	-		10 10			E
		Add Merge column columns	column v column v of	opy data to Fill her column ~ with a	column Duplicate	Create a boolean column from regex match ~ regex matching group	st of Negate boolean values v	to dynamic ¥	
		()							-
									3

Figure 22. Copying to this new column

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



Figure 23. New filter available

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

T Queries	
Equal (Name) Settings	
Pattern:	ОК
Use regex	
	Filter
	1 Inder

Figure 24. 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":

T Queries	
🗄 🖤 🍸 Equal (Name)	
Equal (Name) Settings	
Pattern: L.*	ОК
Use regex	
	Filter

Figure 25. 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 26. 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"

 \rightarrow 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 27. The UNION operator in filters

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



Figure 28. 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:

(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 29. 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 and 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:

Queries INOT (Nodes) INOT (Nodes) INION <	
Equal (Name) Settings Pattern: J.* VISE regex	ОК
	Stop

Figure 30. The NOT nodes operator - 1

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



Figure 31. 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]].*

 Queries Equal (Name) Parameters Oprag subfilter here 	
Equal (Name) Settings Pattern: [^LJ].* Vse regex	ОК
	Filter

Figure 32. Achieving a NOT effect with regex

Tutorials about regular expressions:

- https://regexone.com/
- 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

to go further

Visit the Gephi group on Facebook to get help,

or visit the website for more tutorials

Give a try to nocodefunctions.com, the web application I develop to create networks for Gephi. Click-and-point, free, no registration needed.