Creating a network from a table of entities and their attributes

Clément Levallois

2017-02-22

Table of Contents

last modified: 2017-02-23

Presentation of the plugin

This plugin is created by Clement Levallois.

It converts a spreadsheet or a csv file into a network.

This plugin enables you to:

- Start from a data table in Excel or csv format
- In the data table, nodes are the entities listed in column A
- Nodes' attributes must be listed in columns B, C, D, etc.
- Connections will be created between nodes, when they have identical attributes.
- Attributes can have values, stored in columns right next to the attribute.

1. The input

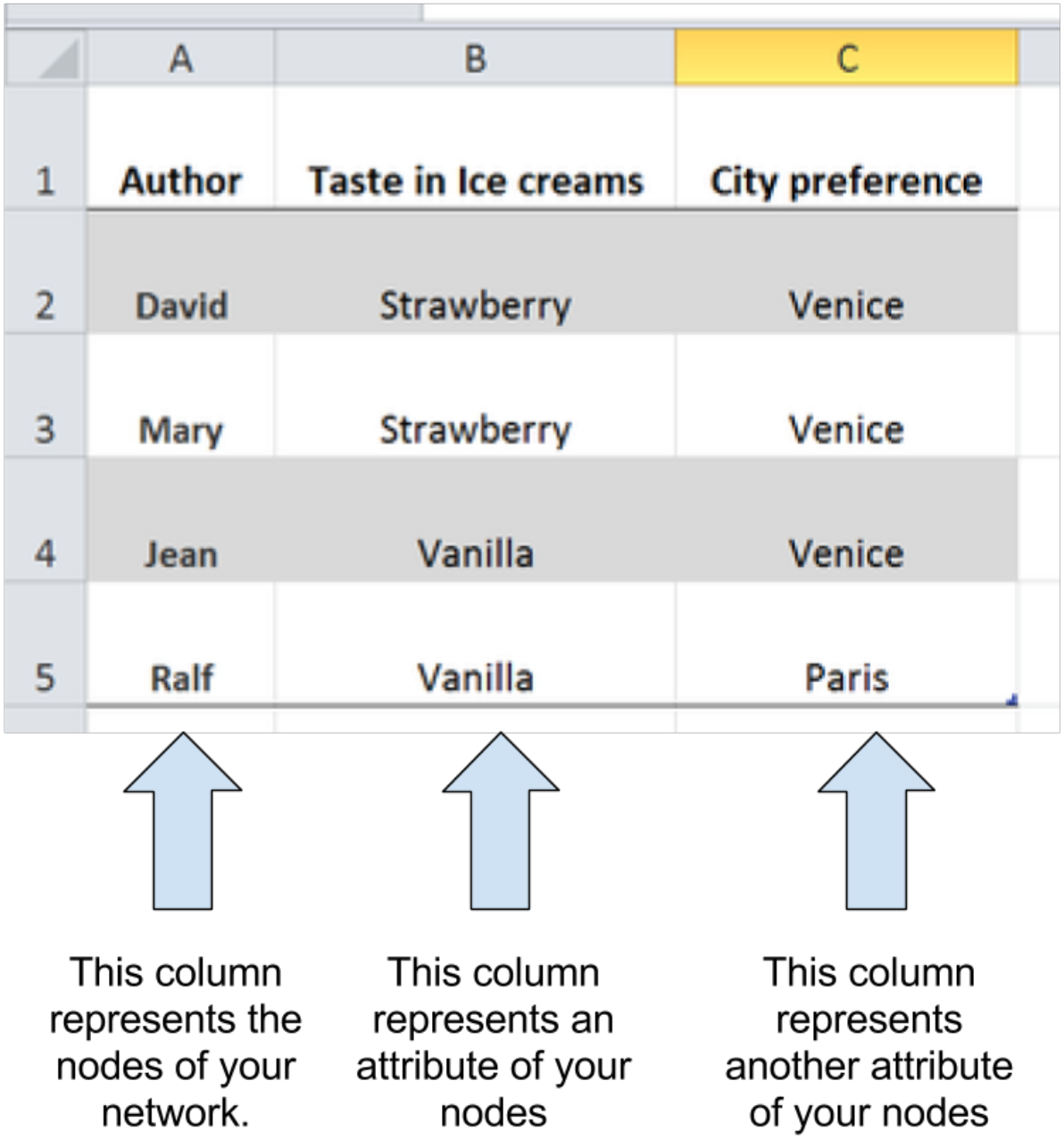

Figure 1. An Excel file

2. The output

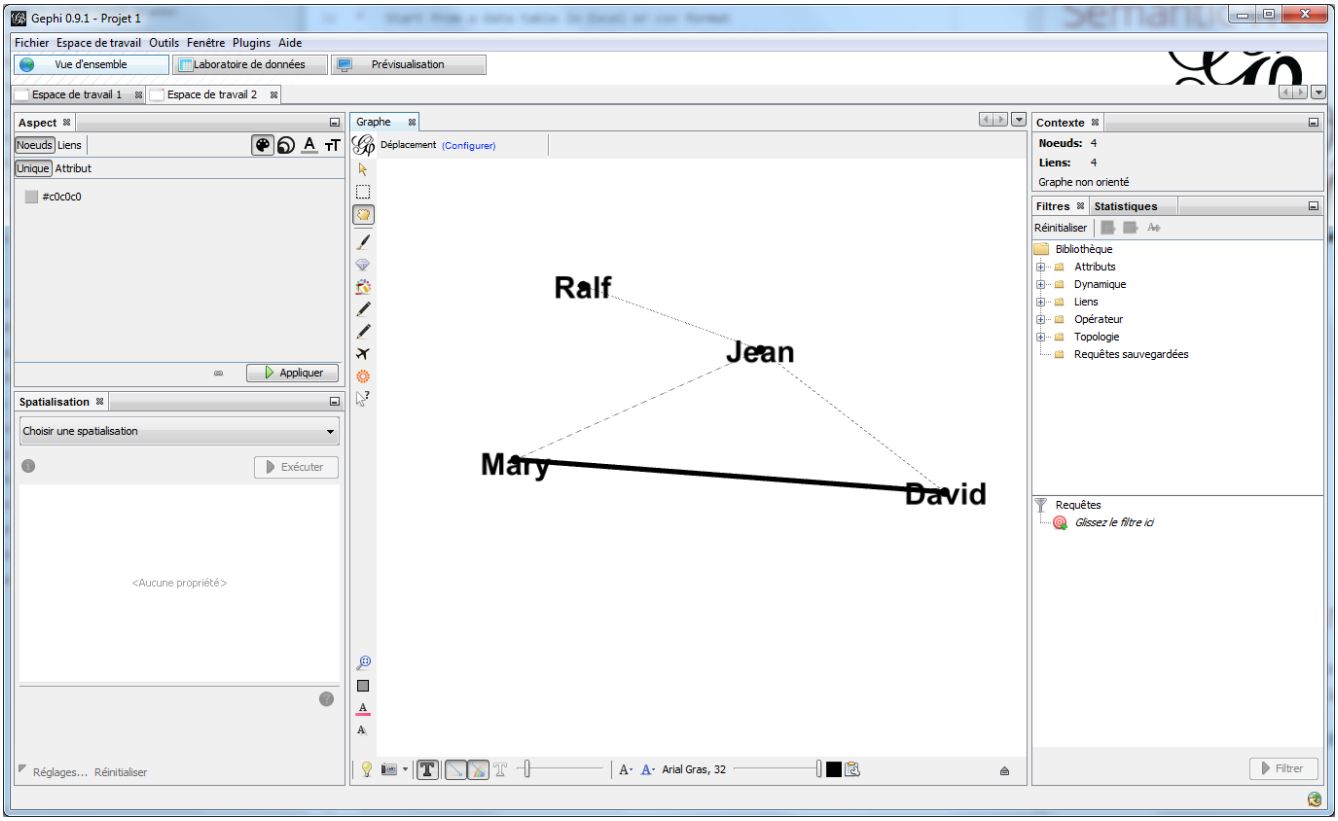

Figure 2. Resulting network

Installing the plugin

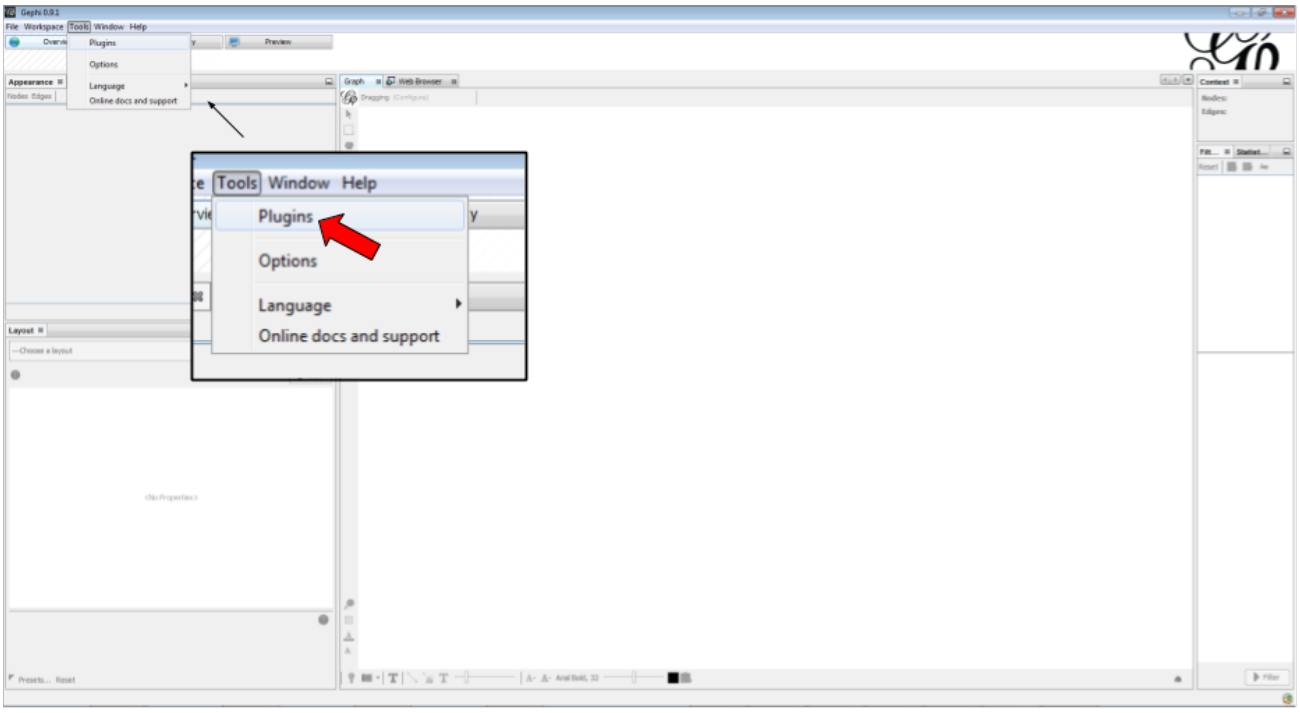

Figure 3. Choose the menu Tools then Plugins

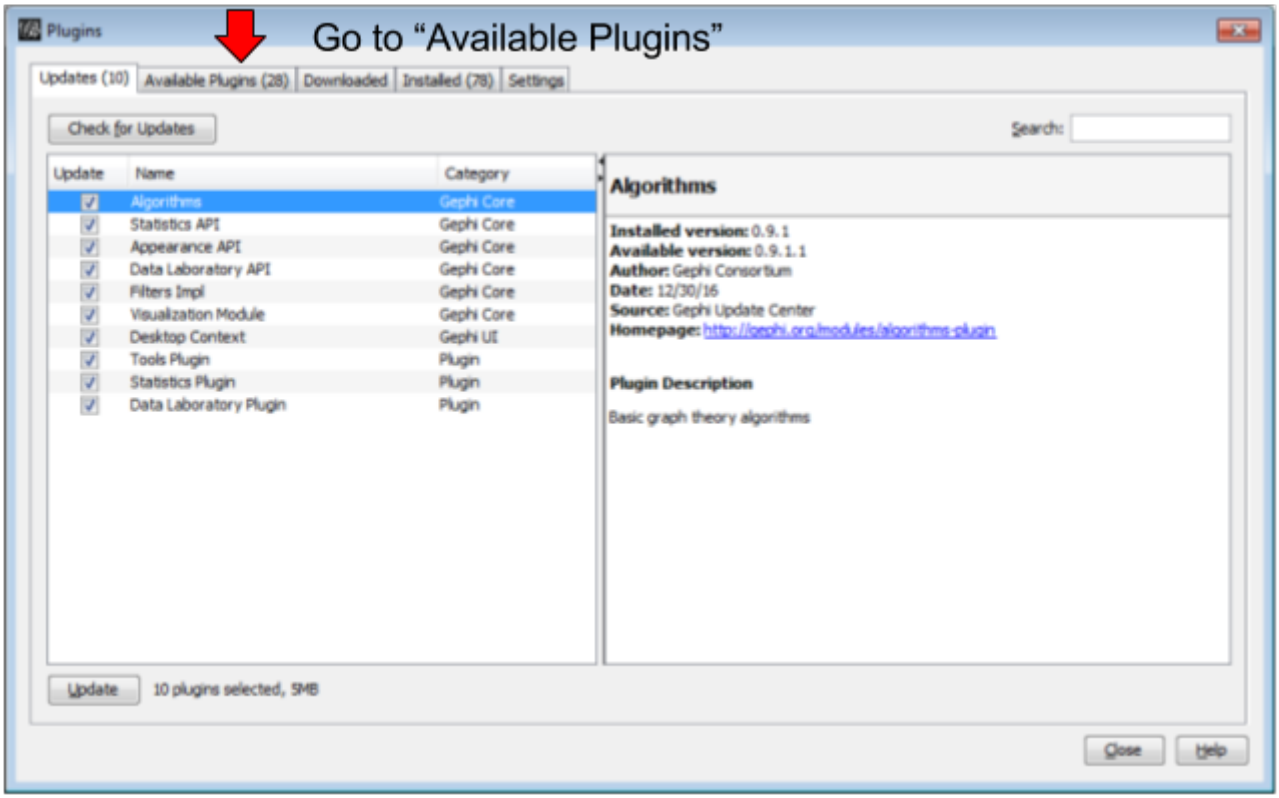

Figure 4. Click on the tab Available Plugins

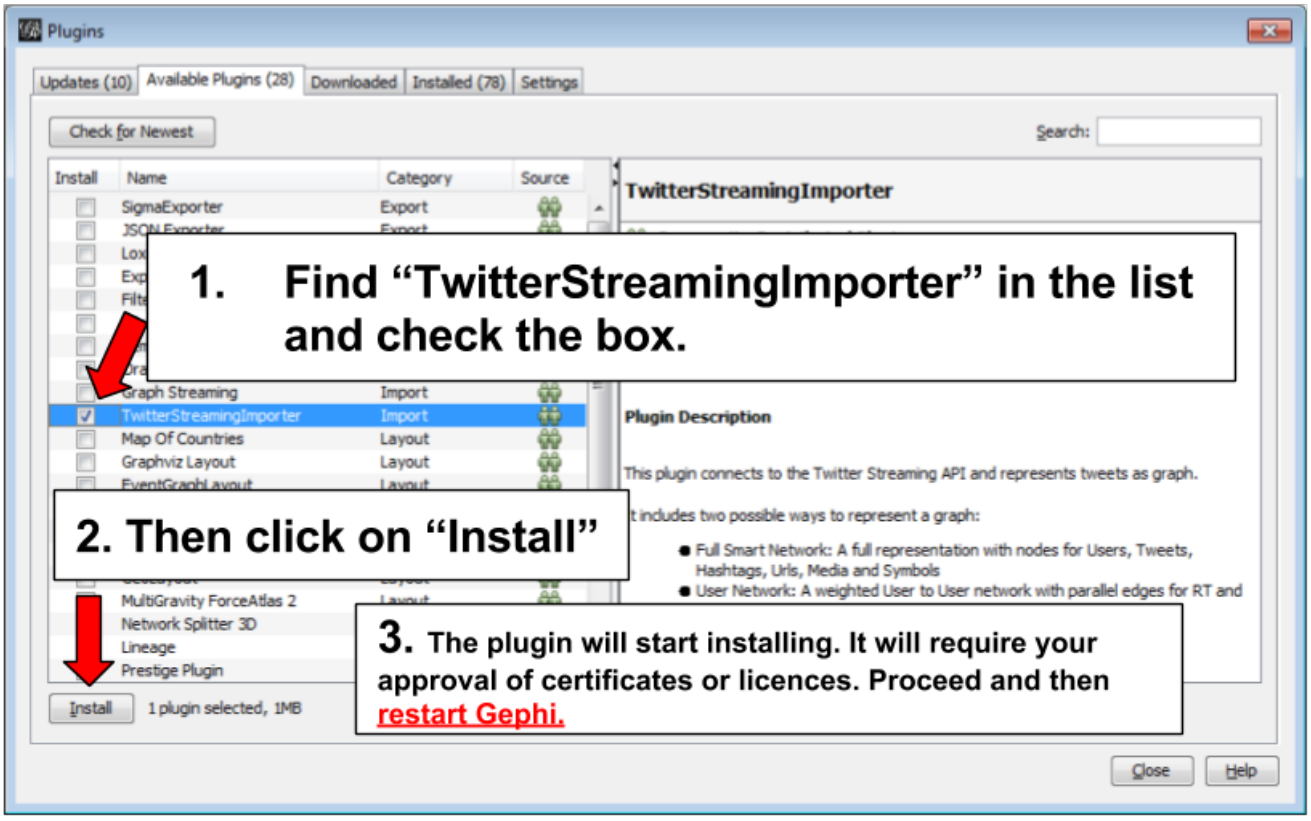

Figure 5. Install the plugin then restart Gephi

Opening the plugin

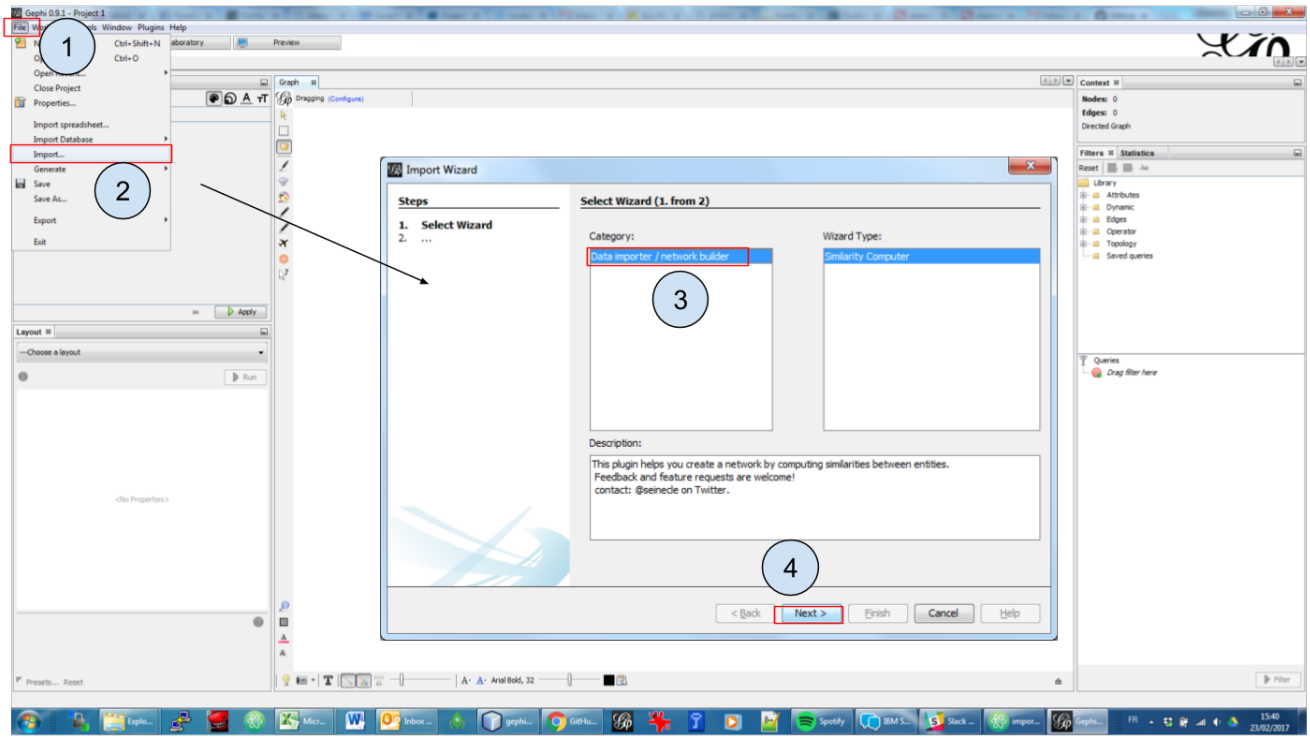

Figure 6. Open the plugin via the menu File - Import

Using the plugin

First panel

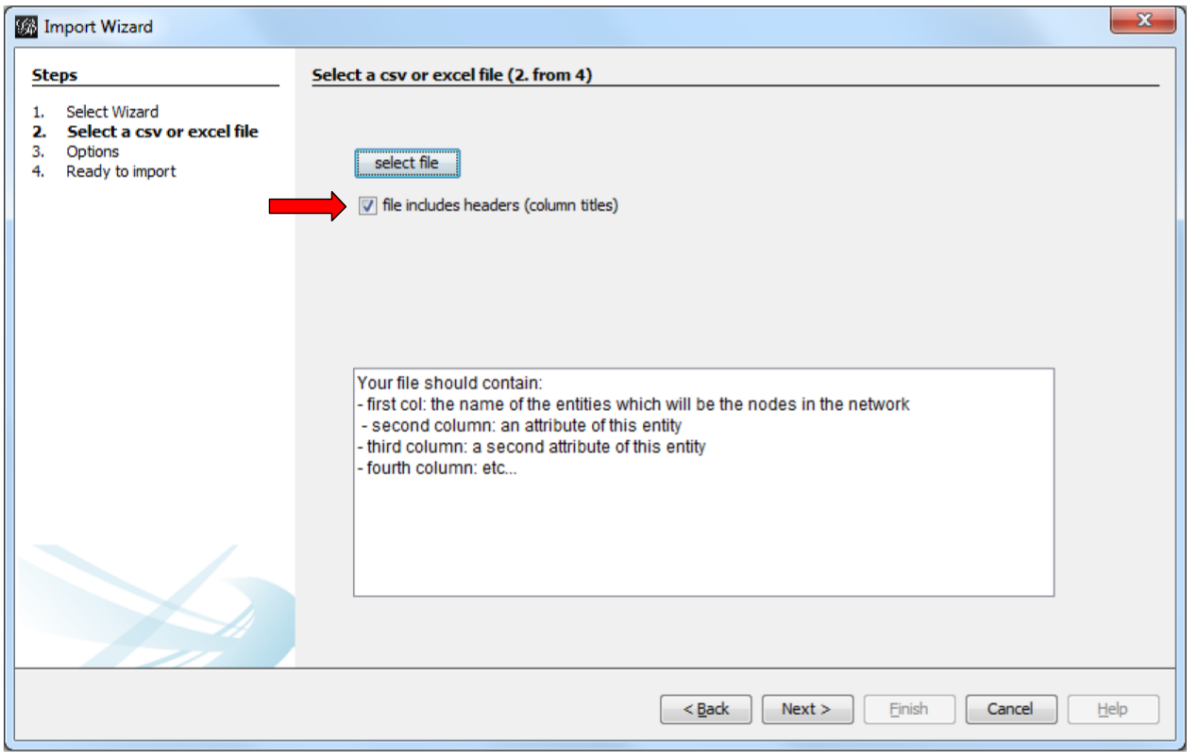

Figure 7. Select a file

	A	B	C
1	David	Strawberry	Venice
2	Mary	Strawberry	Venice
3	Jean	Vanilla	Venice
4	Ralf	Vanilla	Paris

Figure 8. A file without headers

	A	B	C
1	Author	Taste in Ice creams	City preference
2	David	Strawberry	Venice
3	Mary	Strawberry	Venice
4	Jean	Vanilla	Venice
5	Ralf	Vanilla	Paris

Figure 9. A file with headers

Second panel

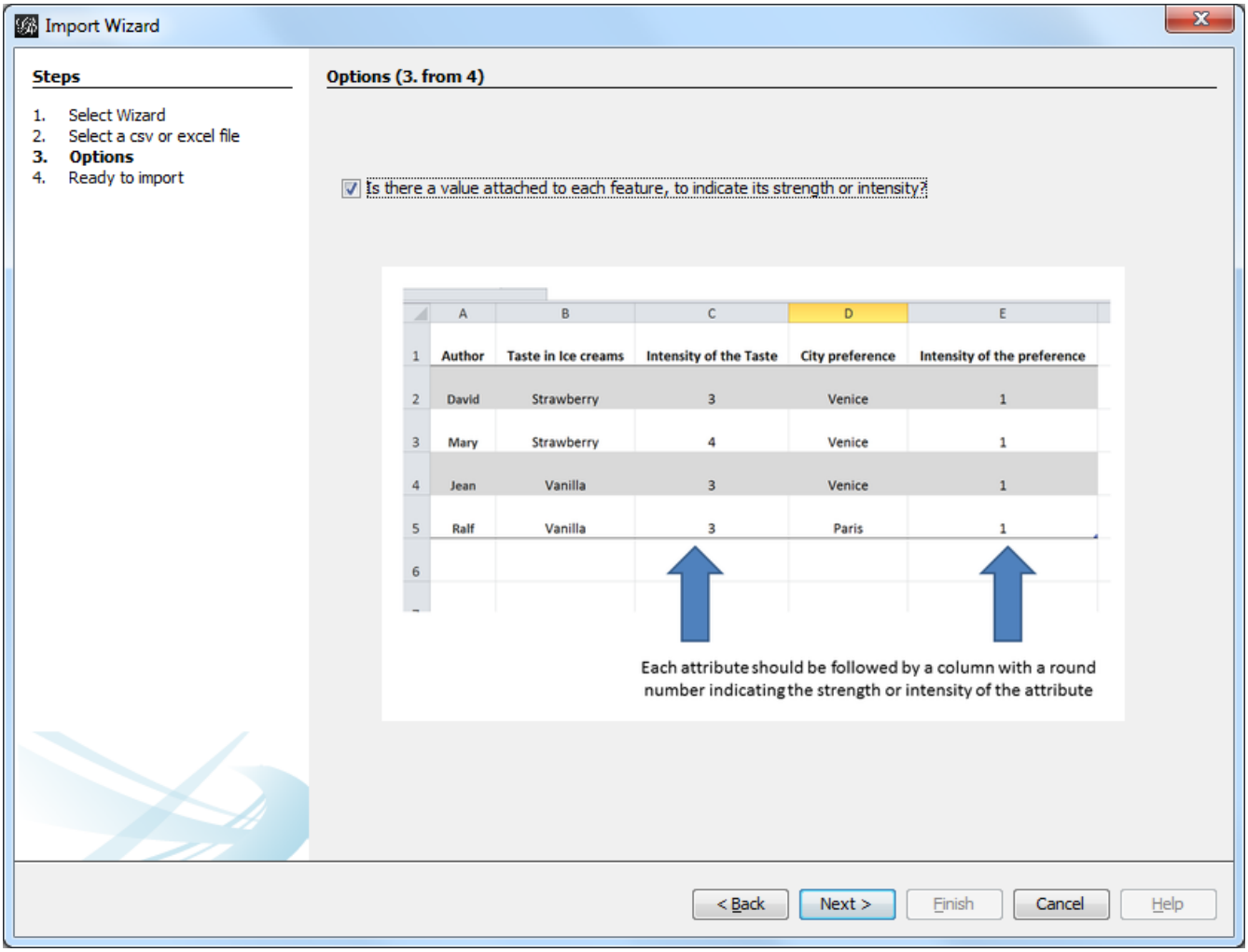

Figure 10. Parameter for weight

Third panel

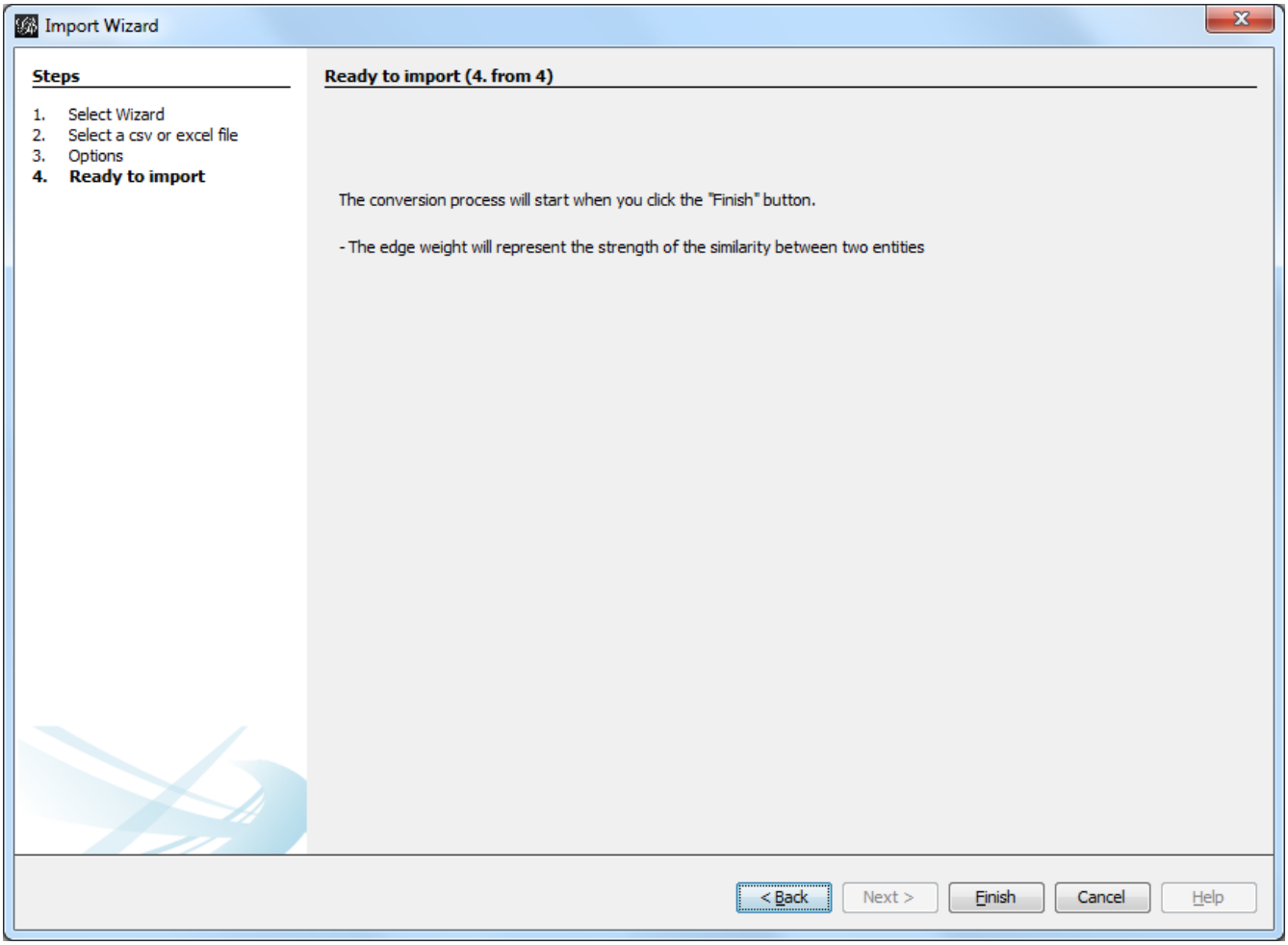

Figure 11. Confirmation panel

How is the similarity computed, exactly?

We use the cosine similarity. Sounds complicated, but it is not. Check here.

The source code for the cosine calculation is in this file, at this place.

FAQ / special notes on the plugin

1. Excel files should be .xlsx, not .xls

Because they represent two slightly different files formats, and the plugin supports only .xlsx

2. csv files are ok.

If you select a csv file, you will be asked to indicate the field delimiter and optionally the text delimiter.

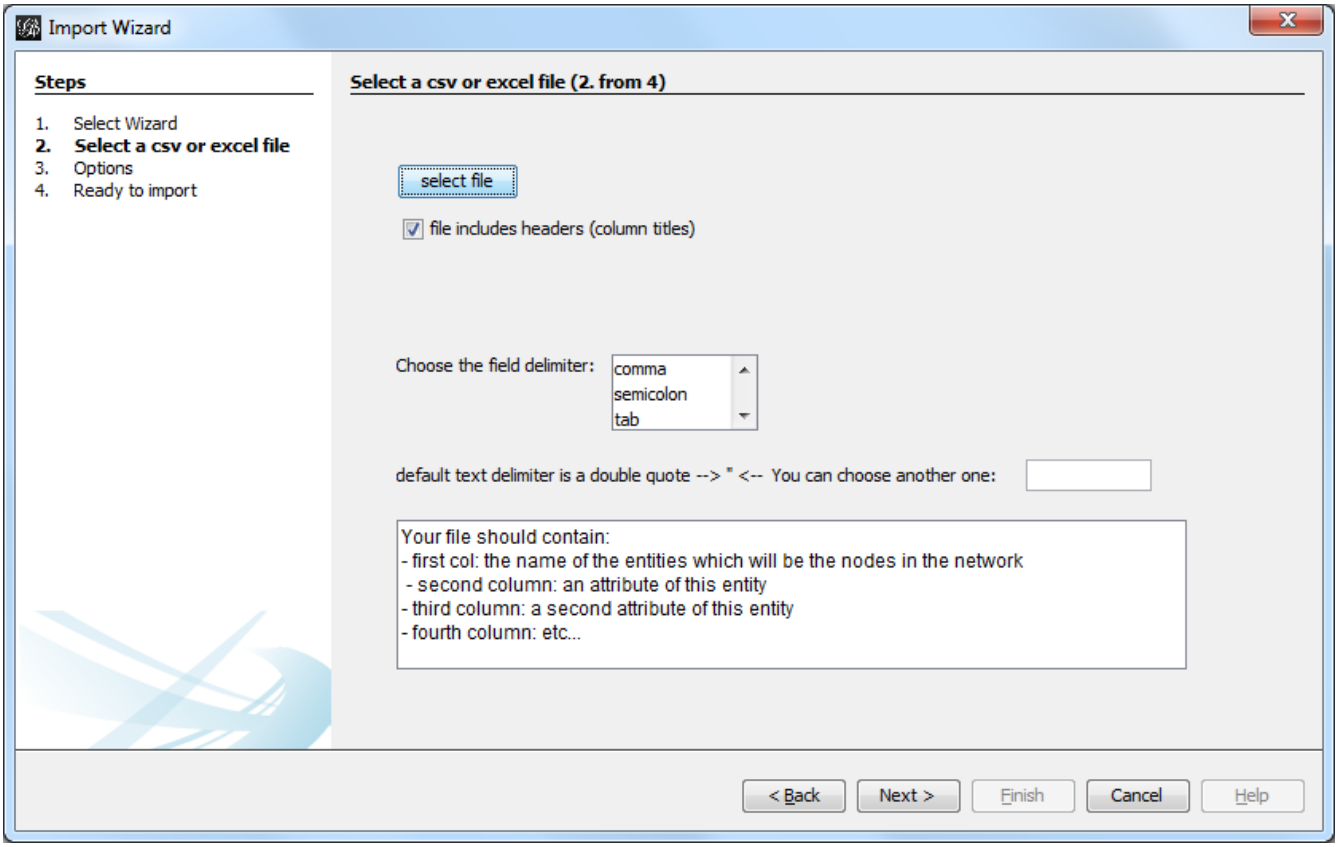

Figure 12. When a csv file is selected

3. You can't use numerical values in the attributes

	A	B	C	D
1	**Author**	**Taste in Ice creams**	**City preference**	**Age**
2	**David**	Strawberry	Venice	28
3	**Mary**	Strawberry	Venice	27
4	**Jean**	Vanilla	Venice	55
5	**Ralf**	Vanilla	Paris	54

Figure 13. Age is a numerical attribute

This is too bad. If there is enough demand for it I'll add this feature, which is not trivial.

4. Each entity should appear only on one line

	A	B	C	D
1	**Author**	**Product purchased**	**Recommender**	**Age**
2	**David**	sofa	Janet	28
3	**Mary**	motorbike	Vince	27
4	**Jean**	shoes	Ron	55
5	**David**	wallet	Fred	28
6	**Ralf**	diner table	Lou	54
7				
8				

Figure 14. An entity appearing twice

David appears on lines 2 and 5 (because he made two purchases). Only the latest line where David appears (line 5) will be taken into account.

The end

Visit the Gephi group on Facebook to get help,

or visit the website for more tutorials