

Creating a network from a table of entities and their attributes

Table of Contents

Presentation of the plugin	1
Installing the plugin.....	2
Opening the plugin.....	3
Using the plugin	3
How is the similarity computed, exactly?.....	4
FAQ / special notes on the plugin	4
The end	4

last modified: 2019-07-13

[gephi logo 2010 transparent] | *gephi-logo-2010-transparent.png*

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

[An Excel file] | *An-Excel-file.png*

Figure 1. An Excel file

2. The output

[result] | *en/similarity-computer/result.png*

Figure 2. Resulting network

Installing the plugin

[Choose the menu Tools then Plugins] | *Choose-the-menu-Tools-then-Plugins.png*

Figure 3. Choose the menu Tools then Plugins

[Click on the tab Available Plugins] | *Click-on-the-tab-Available-Plugins.png*

Figure 4. Click on the tab Available Plugins

[Install the plugin Similarity Computer then restart Gephi] | *Install-the-plugin-Similarity-Computer-*

then-restart-Gephi.png

Figure 5. Install the plugin Similarity Computer then restart Gephi

Opening the plugin

[Open the plugin via the menu File Import] | *Open-the-plugin-via-the-menu-File---Import.png*

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

Using the plugin

First panel

[Select a file] | *Select-a-file.png*

Figure 7. Select a file

[file without header en] | *en/similarity-computer/file-without-header-en.png*

Figure 8. A file without headers

[file with header en] | *en/similarity-computer/file-with-header-en.png*

Figure 9. A file with headers

Second panel

[plugin 4 en] | *en/similarity-computer/plugin-4-en.png*

Figure 10. Parameter for weight

Third panel

[plugin 5 en] | *en/similarity-computer/plugin-5-en.png*

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.

[plugin 6 en] | [en/similarity-computer/plugin-6-en.png](#)

Figure 12. When a csv file is selected

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

[numerical attributes en] | [en/similarity-computer/numerical-attributes-en.png](#)

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

[plugin 7 en] | [en/similarity-computer/plugin-7-en.png](#)

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](#)