Creating a network from a table of entities and their attributes

Table of Contents

Presentation of the plugin ................................................................. 1
  1. The input .............................................................................. 2
  2. The output .......................................................................... 3
Installing the plugin ...................................................................... 3
Opening the plugin ....................................................................... 5
Using the plugin .......................................................................... 5
  First panel ............................................................................... 5
  Second panel .......................................................................... 7
  Third panel .............................................................................. 7
How is the similarity computed, exactly? .................................... 8
FAQ / special notes on the plugin ................................................. 8
  1. Excel files should be .xlsx, not .xls .................................. 8
  2. csv files are ok. ................................................................. 8
  3. You can’t use numerical values in the attributes .............. 9
  4. Each entity should appear only on one line ................... 9
The end .................................................................................... 10

last modified: 2019-07-13

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

![Excel Table](image)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Author</td>
<td>Taste in Ice creams</td>
<td>City preference</td>
</tr>
<tr>
<td>2</td>
<td>David</td>
<td>Strawberry</td>
<td>Venice</td>
</tr>
<tr>
<td>3</td>
<td>Mary</td>
<td>Strawberry</td>
<td>Venice</td>
</tr>
<tr>
<td>4</td>
<td>Jean</td>
<td>Vanilla</td>
<td>Venice</td>
</tr>
<tr>
<td>5</td>
<td>Ralf</td>
<td>Vanilla</td>
<td>Paris</td>
</tr>
</tbody>
</table>

This column represents the nodes of your network.  
This column represents an attribute of your nodes.  
This column represents another attribute of your nodes.

*Figure 1. An Excel file*
2. The output

![Figure 2. Resulting network](image)

Installing the plugin

![Figure 3. Choose the menu Tools then Plugins](image)
Go to “Available Plugins”

Figure 4. Click on the tab Available Plugins

1. Find “Similarity Computer” in the list and check the box.

2. Then click on “Install”

3. The plugin will start installing. It will require your approval of certificates or licences. Proceed and then restart Gephi.

Figure 5. Install the plugin Similarity Computer 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

Figure 8. A file without headers

Figure 9. A file with headers
Second panel

Figure 10. Parameter for weight

Third 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.
3. You can’t use numerical values in the attributes

![Figure 13. Age is a numerical attribute](image)

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
Figure 14. An entity appearing twice

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
<td>Product purchased</td>
<td>Recommender</td>
<td>Age</td>
</tr>
<tr>
<td>David</td>
<td>sofa</td>
<td>Janet</td>
<td>28</td>
</tr>
<tr>
<td>Mary</td>
<td>motorbike</td>
<td>Vince</td>
<td>27</td>
</tr>
<tr>
<td>Jean</td>
<td>shoes</td>
<td>Ron</td>
<td>55</td>
</tr>
<tr>
<td>David</td>
<td>wallet</td>
<td>Fred</td>
<td>28</td>
</tr>
<tr>
<td>Ralf</td>
<td>diner table</td>
<td>Lou</td>
<td>54</td>
</tr>
</tbody>
</table>

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