Gephi: Making your relational data pretty

Jesse Fagan
LINKS Center or Social Network Analysis
aftersox@gmail.com

Data and other links
Attribute data: https://goo.gl/vFVzg5
Edgelist data: https://goo.gl/zGvFl6

Lots of network data!: https://icon.colorado.edu/#!/
My network visualization slides: https://goo.gl/0biHfK
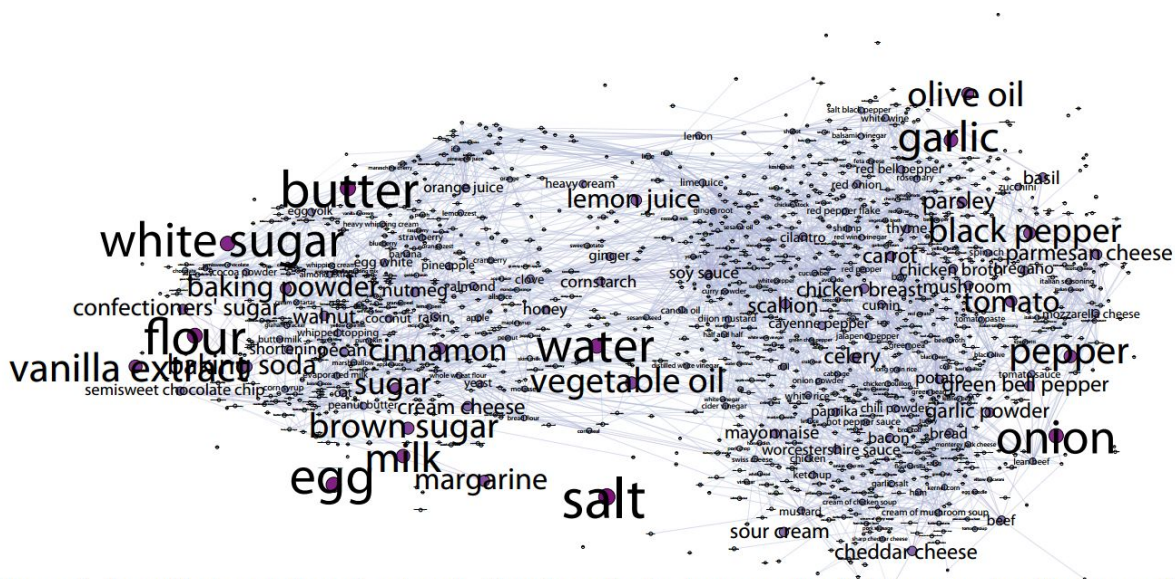Gephi Book: https://goo.gl/XbgvKo

Figure 2: Ingredient complement network. Two ingredients share an edge if they occur together more than would be expected by chance and if their pointwise mutual information exceeds a threshold.

Teng, C.-Y., Lin, Y.-R., & Adamic, L. A. 2011, November 16. Recipe recommendation using ingredient networks. arXiv [cs.SI]. http://arxiv.org/abs/1111.3919.

Gephi is open source software. You can find the source here: https://github.com/gephi/gephi

Gephi is designed for visualizing very large, complex networks. It's best if used on a computer with a good graphics card. The network rendering uses OpenGL and the GPU. The layout algorithms also take advantage of multi-core CPUs. This is an introduction to Gephi, we won't be covering some advanced features (such as directly connecting to databases, longitudinal

Navigating Gephi

There are three sections to Gephi. The Overview, Data Laboratory, and Preview. There are also lots of little windows and tabs and it can all be moved, minimized, and closed. This can result in some situations in which you mess everything up and can't find the windows you need any longer. If that happens under the *Window* menu there is an option for *Reset Windows* which will set all the tabs and such back to the defaults.

Overview This is where you will be spending most of your time. This is where all the work is done translating the data into a visualization.
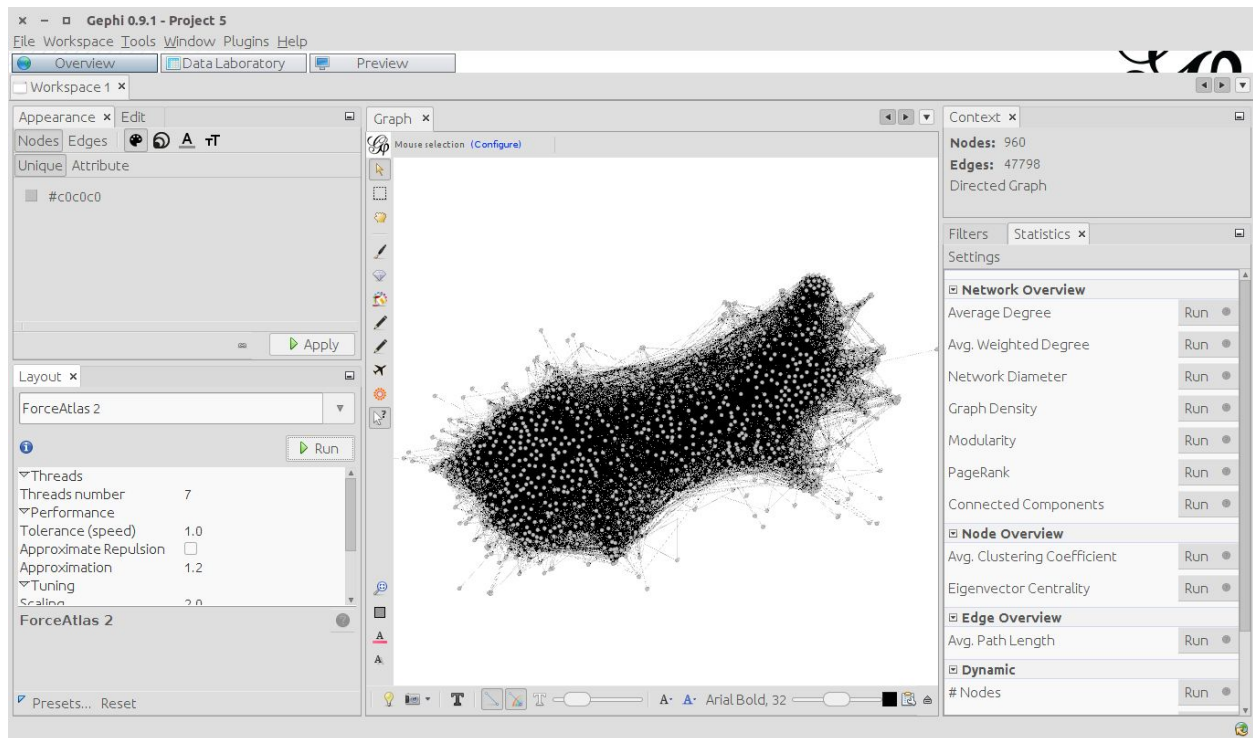
The *Appearance* window controls the color, size of the nodes, labels, and edges. The *Edit* tab will let you see the attributes of the nodes, and change them. The *Layout* controls the position of the nodes. The graph is a visualization / exploration of the network. *Context* gives you a big current overview of what you're working on. The *Statistics* tab runs some basic network analyses. The *Filters* tab lets you filter out nodes and edges based on a range of different criteria.

The **Data Laboratory** is where you can see a spreadsheet of all the data in your project. It is clumsy and limited. There is a limit on the number of columns it will display. Click the *light bulb* icon on the top right part of the window to choose which columns you want to see.

This tab can be helpful if you notice a potential error in your visualization. For example, you may notice that one edge in the visualization is enormous because you coded 999 as a missing weight while everything else is 1-5. You can make edits in this tab.

You can also select a node and have it highlighted in the Overview tab. And vice versa, you can select a node in the Overview and have it highlighted in the Data Laboratory.

Gephi 0.9.1 - Project 5

File Workspace Tools Window Plugins Help

Overview | Data Laboratory | Preview

Workspace 1

Data Table

Nodes Edges | Configuration | Add node Add edge Search/Replace Import Spreadsheet Export table More actions Filter: [Id]

Id	Label	Interval	Years	Sex	Site	Building	Floor	Mil	Dom.	AC	NSRD.	PAF.	Crim.	S.T.	Edu.	Health.
A0004			0.0	0	0	0	35.796939...	7.0	1.0	97.800003	330.0	28422.0	0.0	999999968...	0.0	23.0
A1600			0.004823	100	720	35309	0	3.0	0.0	1.0	999999968...	0.0	0.0	0.01095	100.0	1.0
B0046			3.0	2	48.046092...	5	2	0.0	750.0	31616.0	0.0	0.0	0.0	31.0	1.0	0.06
B4822			31726.0	0	1	1	0	1.0	0.00122	15.5	1.0	0.0	770.0	17741.0	0.0	2.0
C0097			8.0	2	49.509552...	999999968...	1	0.0	24.0	32006.0	0.0	0.0	0.0	31.0	1.0	0.0
C0327			0.0	6	46	0	35966	0.047025	94.699997	2.0	0.0	7.0	11.0	19.1	30.0	3.128259
C3021			0.478195	80.900001...	770	25041	100	2.0	0.3	1.0	1.0	0.0	0.0	0.014632	0.0	1.0
C4831			36039.0	0.1000000...	1	1	0	0.0	0.231098	100.0	999999968...	0.0	770.0	36094.0	0.0	3.0
D5901			45.951126	3	710	35142	0	0.0	0.0	14.0	1.0	0.0	0.0	34.434471	999999968...	1.0
E0002			4.0	0	0.200000003	1	0	10.0	34151.0	0.0	2.0	0.0	1.0	1.0	21.200001	0.0
E0119			0.0	0	12.5	0	44.729476...	1.0	1.0	0.0	31.0	35235.0	0.0	0.0	0.0	0.0
F0221			6.0	2.9197080...	0	999999968...	8.5859203...	100.0	999999968...	0.0	1.0	3.0	0.0	2.0	7.09072	3.0
F1991			1.0	1	0.2985334...	0	1	0.0	20.0	36180.0	0.0	2.0	0.0	1.0	1.0	100.0
H0234			0.0	3	0	33.344924...	3	1.0	0.0	770.0	30894.0	0.0	1.0	0.0	5.0	1.0
H8355			0.0	0	0	0	42.490032...	3.0	1.0	100.0	11.0	33308.0	1.0	1.0	0.0	1.0
I0008			999999968...	3.8132836...	0	7	35	50.599998	1.0	6.360793	2.0	46.0	0.0	999999968...	2.26648	100.0
I0031			1.0	2	48.361068...	8	1	0.0	24.0	36164.0	0.9	0.0	0.0	11.0	2.0	0.0
I0251			0.0	0	2	0	710	36774.0	0.0	999999968...	0.0	6.0	1.0	0.0	0.0	38.436874
I2801			0.016624	0	630	34344	0	2.0	0.0	1.0	1.0	23.700001	0.0	0.196454	91.800003	4.0
J0441			0.0	999999968...	1	0	0	42.565468	3.0	1.0	0.0	720.0	36617.0	0.0	0.0	0.0
J5982			0.933917	100	6	33	0	1.0	3.858186	999999968...	46.0	0.0	999999968...	0.533815	0.0	1.0
K0224			0.0	3	46	0	999999968...	6.384098	100.0	2.0	0.0	7.0	12.0	100.0	7.0	1.147028
L0048			1.0	0	0	1	10.844630...	999999968...	46.0	0.0	999999968...	0.835487	99.800003	2.0	0.0	4.0
L0225			710.0	36606	0	1	1	0.0	1.0	0.011081	100.0	999999968...	0.0	250.0	34246.0	0.0
L0274			1.0	1	0.0224657...	26	4	0.0	740.0	35206.0	0.0	2.0	0.0	1.0	1.0	0.0
L0319			8.0	999999968...	36248	0	0	0.0	19.0	1.0	0.0	0.0	46.621292	3.0	1.0	100.0
M6271			0.149225	100	770	35737	81.699996...	2.0	0.0	1.0	1.0	0.0	1.0	0.126721	72.5	1.0
N0111			1.0	0.115139775	24.200000...	250	35844	0.0	3.0	0.0	1.0	999999968...	0.0	1.0	0.151992	0.0
N1514			0.0	999999968...	1	0	0	36.687069	6.0	2.0	0.0	770.0	26567.0	0.0	0.0	0.0
N6278			0.0	240	1	0	1	1.0	40.2	0.0	0.000158	0.0	1.0	0.0	750.0	28497.0

Add column | Merge columns | Delete column | Clear column | Copy data to other column | Fill column with a value | Duplicate column | Create a boolean column from regex match | Create column with list of regex matching groups | Negate boolean values | Convert column to dynamic

The **Preview** tab is where you prepare your final image for export. It takes information from the Overview tab then produces a high quality render of the results.
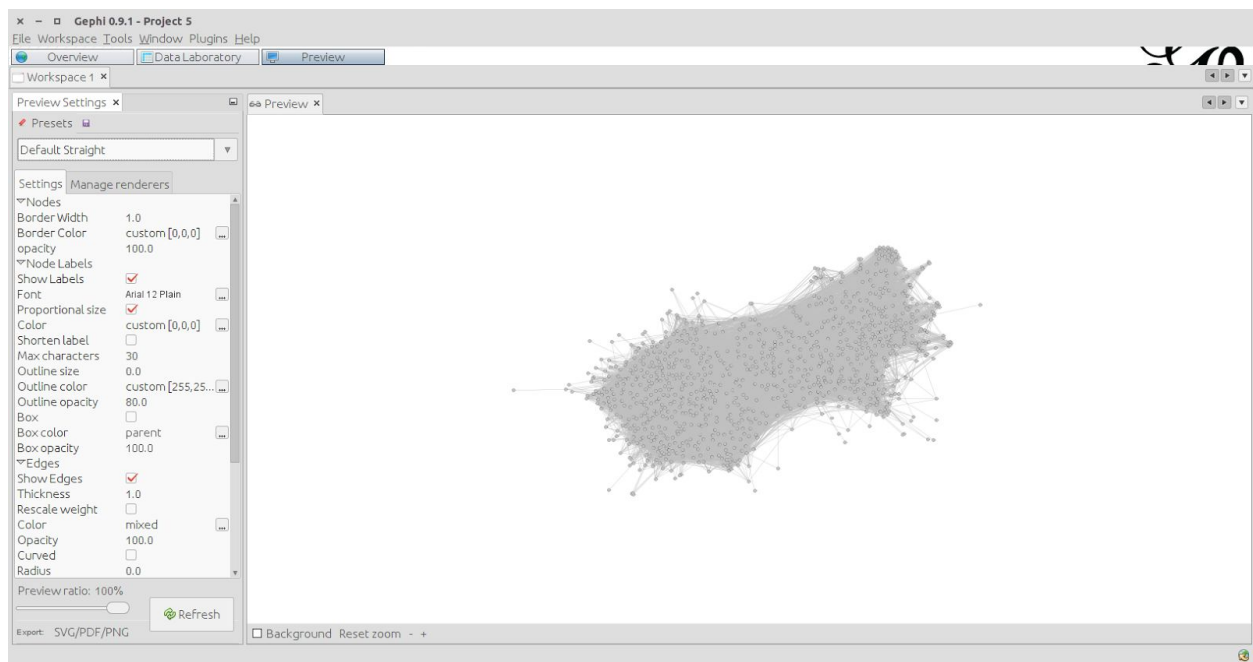
Gephi also has Workspaces. If you are working on a project with multiple relationships, say a co-citation network and a co-authorship network, you can keep them in the same project but separate them to different workspaces.

Exploring a Network

- Zoom with mouse wheel, right click to move around
- Selecting nodes
  - Using ctrl + mouse wheel
  - Rectangle selection
  - Moving nodes around
- Center on the graph
- Gephi screenshots
- Edit / view node attributes using the ? tool
- Manually sizing and coloring nodes
- Manually adding new nodes and ties

Filtering the network

- Filtering by attribute (subnetworks)
- Filtering edges (edge weight)
- Sub-filters (removing isolates)
- By topology (degree range, k-core, ego-networks)
- Sending the filtered network to a new workspace

Sizes and Colors

- Nodes
  - Color (overall, by attribute, nominal data, interval data)
    - Different palettes, editing colors and gradients
    - Splines
  - Size
    - Minimum and maximum sizes
    - Using splines
- Edges
  - Size controlled at the bottom.
  - Color by attribute
  - Color by the node colors

Statistics and Analyses

Gephi is made for visualization, not for analyses. So things are bit sparse here. There are some plugins that add some analyses too.

- All forms of degree
- Centrality: betweenness, eigenvector, PageRank
- Clustering coefficient
- Components and modularity

Layouts

There are a wide variety of layouts available for network visualizations. My favorite is Force Atlas 2, which is only available in Gephi (with a rudimentary version in R). Some of these layout options are helper layouts, not made to be used on their own.

- Using Force Atlas 2
  - Most of my tweaking is based on *scaling* and *gravity*
  - Prevent overlap if nodes are too crowded
  - LinLog mode for very dense, large networks, then re-tune the scaling and gravity

- Once you get a setting you like, save it as a preset
- Tweaking with expansion, contraction, rotation, and label adjust

Preview and Export

In this window you create a high quality render of the network and export it for sharing. After every change, remember to hit **Refresh**.
- Zoom, background, preview ratio
- Presets
  - Using a default setting
  - Saving your own settings
- Nodes
  - Border width, color, opacity
- Node labels
  - Size, font
  - Proportional size
  - Outlines, boxes
- Edges
  - Thickness, rescale weight
  - Opacity
  - Curved
  - Arrows and labels

Investigating and Changing Data

- Highlight in the Overview / Data Laboratory
- Copying, filling, merging columns
- Showing and hiding columns
- Filtering tables
- Export data

Reading in Data

Data is read using the Data Laboratory tab. Click "Import Speadsheet". Gephi only support CSV file format.

The network data must be structured as an edge list. The titles of the first two columns need to be "source" and "target". These are special column names that Gephi looks for and they will be used to construct the network ties. The word "weight" is also a keyword that Gephi looks for. You can have all sorts of other edge attributes, but make sure you assign them the correct data type.

For nominal data use *String*, and for interval data use *Double*. For the most part, you can

entirely ignore the long list of other data types.

If you have node attribute data you should import the edgelist first. The node attribute data needs to have a column called "Id". The values of the ID's need to match the names of the nodes in the edgelist. Make sure you have the box "Force nodes to be created as new ones" *unchecked*, otherwise it will import all your node data as brand new nodes.

☐ Force nodes to be created as new ones