Machine Vision Workshop: Gephi Tutorial

NORDMEDIA, Malmö, 22 August 2019

Jill Walker Rettberg, Ragnhild Solberg and Linn Heidi Stokkedal
University of Bergen

This short Gephi tutorial is part of the workshop Using Network Analysis to Understand Cultural Representations of Machine Vision held at NORDMEDIA 2019. It uses an early dataset from the database we are building to map representations of machine vision in art, games and narratives.

Install Gephi

Before the workshop, please install Gephi. You can download it from https://gephi.org. On some Windows computers, there might be an error that stops you from opening Gephi. If this is the case, try executing "gephi.exe" instead of "gephi64.exe"
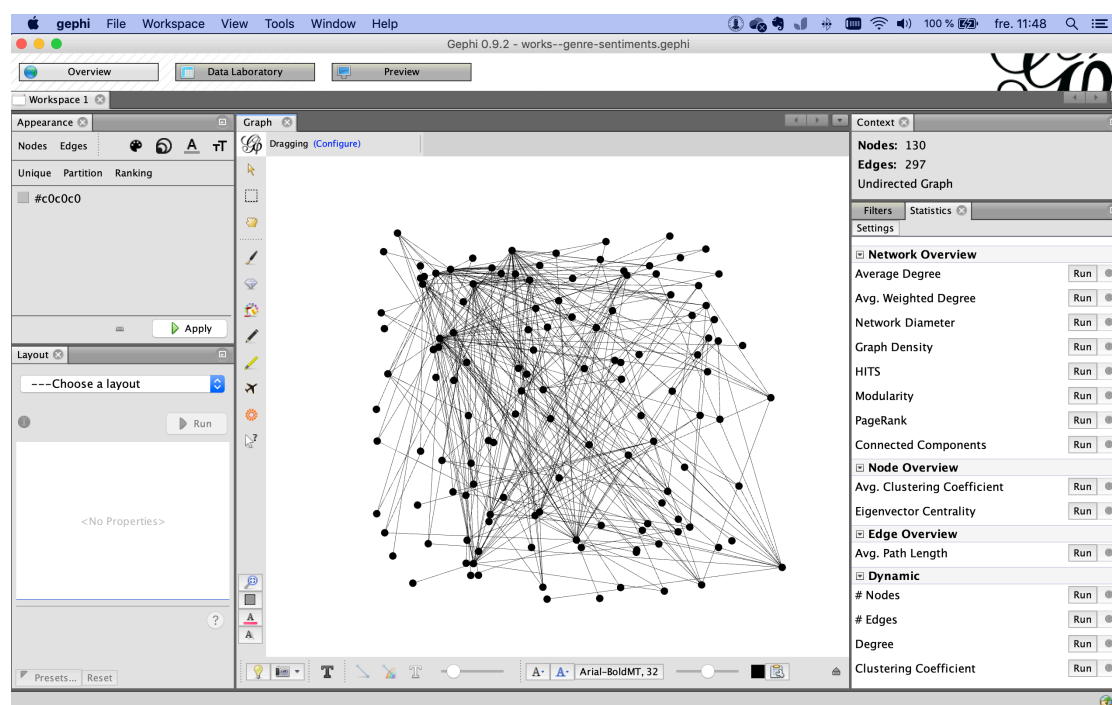
Opening the gephi file

Download the dataset from
https://figshare.com/articles/Gephi-tutorial-NORDMEDIA2019-Machine-Vision_pdf/9638654
(or a simpler alias: http://link.uib.no/nordmedia-gephi
Open the file in Gephi. It should look something like this.

There are two kinds of **node** in this graph, *creative works* (artworks, games, movies etc.) and *sentiments* with which we have tagged each work. That makes it a **bi-partite network**, or if you prefer, a two-mode network.

The nodes are connected by **edges** – the lines between them. In this dataset, edges are all between works and sentiments (so not from one work to another work), and a connecting edge means that a specific work has been tagged with a specific sentiment.
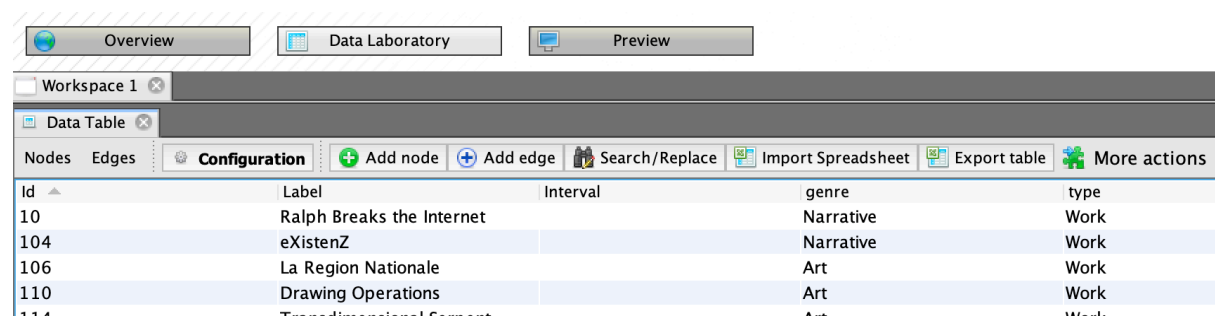
Zooming and dragging

On a mac:
- to zoom, pull two fingers up or down on your mouse or trackpad.
- To drag the graph, Control-click the graph and drag.
- Doubleclicking the magnifying glass to the left of the graph resets.
- If you lose a window altogether, use the menu to select Windows → Reset windows

On windows:
- To zoom, use the scroller on the computer mouse (the trackpad changes the size of the brush).
- To drag the graph, right-click the graph and drag.
- You can personalize other functions in the top menu: "tools"/"options"/"keymap"

The Data Laboratory

To see the actual data that is being visualized as a network, click on Data Laboratory at the top of the window. This shows the **node table**. The top nodes are movies and artworks. Each has a unique ID number, a label (the title of the work), a genre that we have assigned it, and a type.

Id ▲	Label	Interval	genre	type
10	Ralph Breaks the Internet		Narrative	Work
104	eXistenZ		Narrative	Work
106	La Region Nationale		Art	Work
110	Drawing Operations		Art	Work
114	Transdimensional Serpent		Art	Work

If you scroll down, you'll also find nodes where the type is Sentiment:

Id ▲	Label	Interval	genre	type
2000107	Oppressive			Sentiment
2000129	Dangerous			Sentiment
2000188	Overwhelming			Sentiment

If you click on the word Edges, in the top left, you'll see the **edge table**, which is simply a list of the connections. The top line in the screenshot below means that work 239 is tagged with the sentiment 200096

| Source	Target	Type	Id
239	200096	Undirected	308
204	200096	Undirected	309
233	200096	Undirected	310
203	200096	Undirected	311
167	200096	Undirected	312
179	200096	Undirected	313

Colour the nodes based on what they represent

Now let's play with the network! Return to the Overview.
In the Appearance window, choose Partition, then the attribute Genre.

The sentiments don't have a genre, so they should up as "null", which is fine.

Click Apply, and the Creative works and the Tags will have two different colours. You can change the colours by clicking Palette – red and green aren't great choices for colour blind people, for instance. Now your network will look something like this:

To see what each node represents, you can **Show node labels**. Click the black T near the bottom left of the Graph window. Use the slider on the bottom right to **adjust the size**.

Try clicking on an individual node to see what it is connected to.

2. Run the Force Atlas 2 algorithm to cluster the nodes by communities.

The most common layout algorithm is Force Atlas 2, which clusters nodes that have many edges between them. Run it from the bottom left window. You may want to adjust Scaling, and to run Expansion or Contraction algorithms to adjust the size. Try different things, and backup frequently as you can't undo things in Gephi.
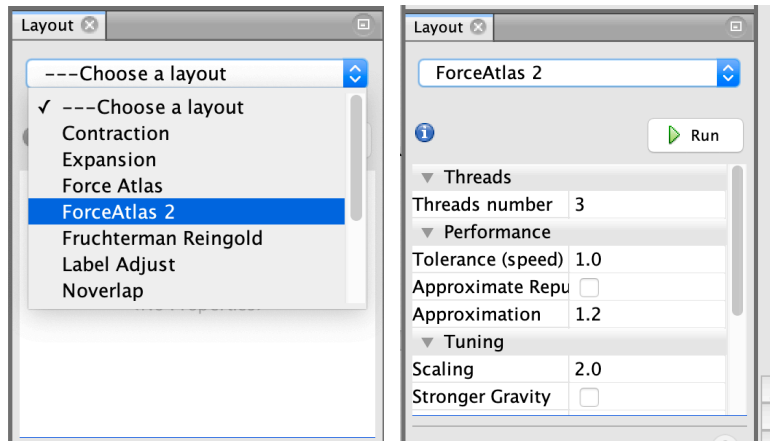
The default settings make for a very dense graph, so set scaling to 50 and gravity to 20 – you can play with these settings and see what happens.
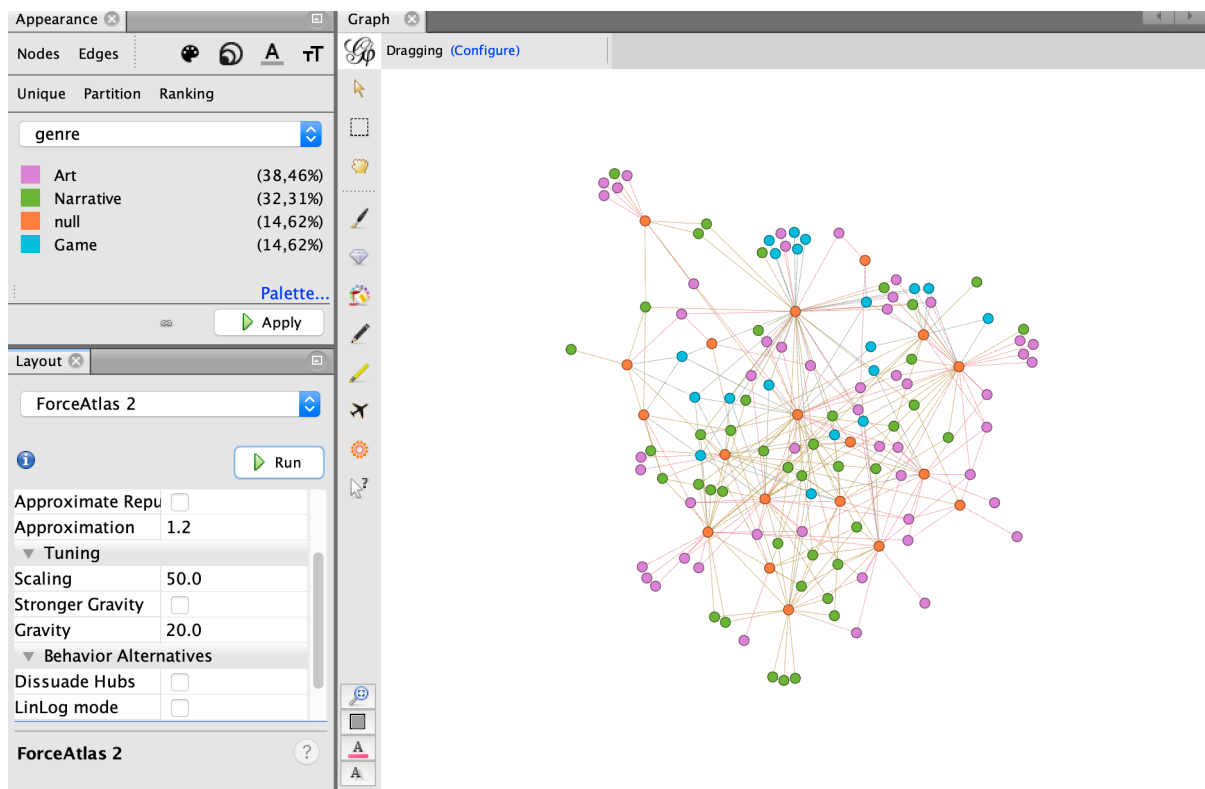

Size the nodes by degree

The visualization will become clearer if you make nodes that have a lot of connections bigger. The number of edges a node has is its **degree**.

Go back to the top left box, Appearance, and click Nodes.

To change nodes' colour, we used the palette 🎨. Now choose the size option 🔵 instead.

Select Ranking.

Choose <u>Degree</u> from the pull down menu, and choose a default minimum and maximum sizes. Click Apply.
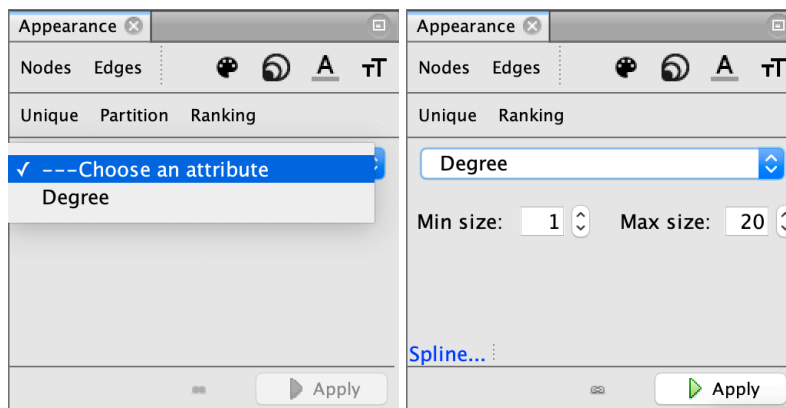
Here, I used min size 8 (so you can still see the colour coding of art/game/narrative and max size 20. You can see which sentiments are most frequent, and you might start to see patterns in what kinds of genres correlate with what sentiments. (Note that this is early-stage, incomplete data!)

View the visualisation in the Preview window

To make the graph pretty, and to export it, go to the Preview window (top menu). The window will be empty, you have to click Refresh to show the visualization.

Try adding labels, adjusting the size of the font. To make the text easier to read try reducing the opacity of the nodes and removing the borders.

The default view has curved edges. **In network analysis, curved edges mean that the edges are directional**, like in a Twitter network: A follows B, but B doesn't necessarily follow A. The direction of the edges can be read by following the curve clockwise. Our dataset is undirected, so strictly speaking, we shouldn't use curves.
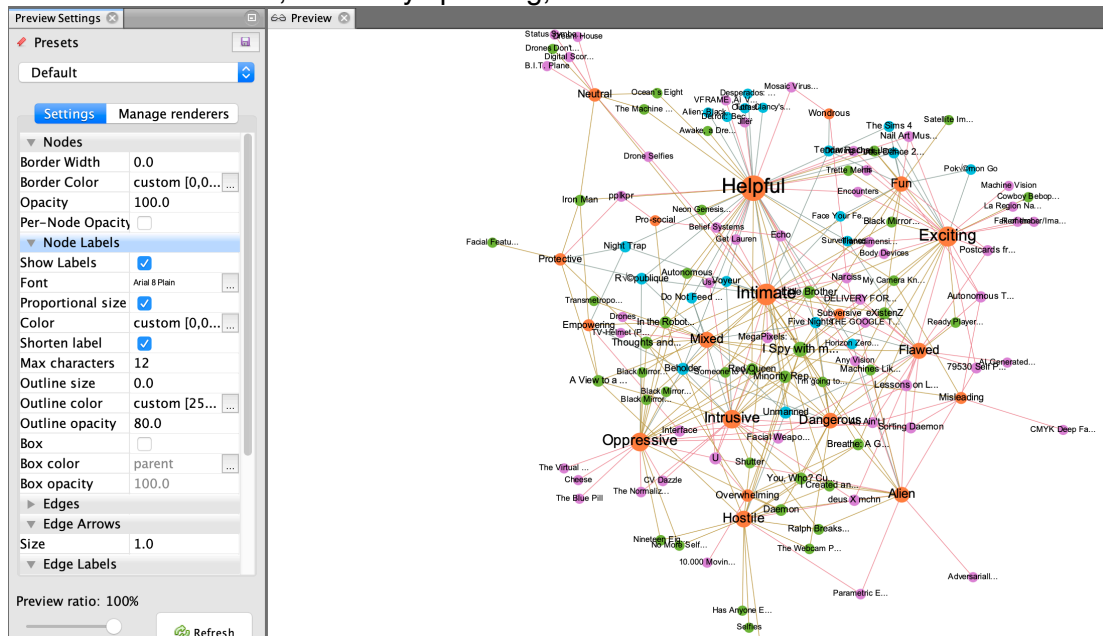
More things you can do in Gephi
There is lots more you can do, of course. If you run the statistics in the right window, you can divide the nodes by **modularity**, that is, by the community they belong to based on which other nodes they are connected to. Then you can **partition** the graph by modularity and for instance choose to colour nodes by their modularity grouping rather than by genre. Or you can **filter** out certain parts of the graph.

More about the Machine Vision project
We recently finished a conference paper for ACM HT19 describing the database and the methodology. Here is a preprint:

Rettberg, Jill W., Marianne Gunderson, Ragnhild Solberg, Linda Kronman, and Linn H. Stokkedal. 2019. "Mapping Cultural Representations of Machine Vision: Developing Methods to Analyse Games, Art and Narratives." MediArXiv. July 4. doi:10.33767/osf.io/fvwm8.

Links:
The Machine Vision Database: https://machine-vision.no/knowledgebase
Project website: https://www.uib.no/en/machinevision
Twitter: @machvisionERC