Practice Session 01: Cytoscape Basics

Materials for this session

:warning: Do not simply right-click on the file names below, or you will download an HTML file that will not be readable by Cytoscape. See how to download in the README of the data/ directory.

- File "karate.gml"
- File "starwars.graphml"
- File "us_companies_ownership.csv"

Contents of this session

0. About Cytoscape

Cytoscape is an open source software platform for visualizing complex networks and integrating these with any type of attribute data.

1. Importing a network

1.1. Import Zachary's karate club

Let's start with a simple case: Zachary's Karate Club. This was a Karate Club with a sensei (#1) and a club president (#34) that split into two: some people remained with the sensei, and the others created a new club with the club president.

```
Linia de codi 1


Linia de codi 2
```

Linia de codi 3

Linia de taula 1

Linia de taula 2

rgdfg	dfgdfg	dfg
	sss	1
	ereer ddd	2

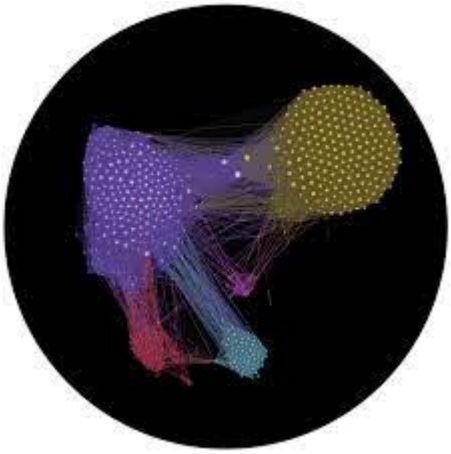
- `File > Import > Network from File ...`
- Select `karate.gml`
- `Layout > Compound Spring Embedder`
- Look at the graph and try to figure out if there is anything special about nodes 1 and 34.
- [**REPORT**] Include in your report this graph plus and a brief paragraph indicating

whether nodes 1 and 34 have visually anything special.
- [**REPORT**] The Compound Spring Embedder is an algorithm derived from force-directed graph layour algorithms. Read the Wikipedia page on force-directed graph drawing and explain in one paragraph, in your own words, how this works.

Do not use screenshots; use `File > Export as image`

1.2. Import the Star Wars characters network

Open a graph showing characters that appear or are mentioned in the same scene of a Star Wars movie.



- `File > Import > Network from File ...`
  - Select `starwars.graphml`
  - If asked, select *shared name* for the node identifier column. This will transform node identifiers into a column named "shared name" internally.
- `Layout > Prefuse Force Directed Layout > All nodes > scenes`
  - Find a node with degree larger than *D=20*
  - [**REPORT**] Include in your report this graph. Indicate which character has degree larger than *D* (right click on blank space -> add text, then right click on the text -> add arrow).
  - [**REPORT**] Include a brief commentary of what kinds of characters are represented by nodes with degree larger than *D*

1.3. Import US companies co-ownership

Open a graph representing company co-ownership in the US:

- `File > Import > Network from File ...`
  - Select `us_companies_ownership.csv`
  - Click OK (accept default import)
  - It might take a couple of minutes to open
  - Layout > Edge Weighted Spring Embedder (might take ~10 minutes in some PCs)
  - [**REPORT**] Include in your report this graph.
  - [**REPORT**] Do you see more than one connected component? What do connected components represent in this graph?
  - [**REPORT**] Include a brief commentary on large-degree nodes in this graph, which are they? What do those nodes represent?

Note: you can zoom in and zoom out with the mouse scroll wheel, you can also use the panel on the bottom-right of the screen to navigate the graph.

2. How to edit node and edge styles

You do not need to include anything from this part (part 2) in your report.

Reload the Karate Club dataset (`karate.gml`).

Now you can play with the "Style" panel (top-left, between "Network" and "Filter"). Here are some ideas.

2.1. How to change the style of the entire network

To style the entire networks in different ways:

- Play with predefined styles, e.g. "Minimal", "Curved", or others.

2.2. How to name nodes

To include in each node its name:

- In the label property,
- create a "Passthrough mapping"
- for attribute "name"

To remove these names:

- Remove the mapping (trash can icon)

2.3. How to change the shape of node

To change the shape of nodes:

- Click on the drawing to the left of "Shape" and choose another shape

2.4. How to change edge width

To change the width of edges:

- Click on "Edge" on bottom left (between "Node" and "Network")
- Click on "Width" property
- Mapping Type = Continuous Mapping
- Column = scenes (this will work in the *Star Wars* graph, which has a column with the number of scenes in common)
- Change the "Current Mapping" by double clicking. You should see a window "Continuous Mapping Editor ..."
- Create a mapping that gives a clear visual separation between thin and thick edges, by editing the mapping so that it has a broader range of values

2.5. How to add arrows

To add arrows, you need a directed graph such as the company ownership dataset.

- Change the setting of "Target Arrow Shape".

2.6. How to change the entire layout

Try some layouts ("Layout" menu)

- Degree Sorted Circle Layout > All nodes
- Edge Weighted Spring Embedded Layout
- Try this with the Karate Club, look for nodes 1 and 34.
- Prefuse Force Directed Layout

You do not need to include anything from this part (part 2) in your report.

3. Basic network analysis

3.1. Analyze network

Perform basic network analysis. `Tools > Analyze network`. Consider the network is not directed.

- Load the Karate Club network
- The analysis adds some node attributes
- Look at these node attributes (e.g., find the node with the largest betweenness centrality)
- [**REPORT**] Indicate which are the two nodes with largest betweenness centrality in the Karate Club
- Change the fill color of nodes to be a *continuous mapping* of column *Betweenness Centrality*, choose the colors so that higher betweenness centrality is associated with a darker color.
- [**REPORT**] Include this graph in your report

3.2. Plot different distributions

Look at the results from the network analysis (you will need to go to `View > Show results panel` -- if it does not show up, try hiding and showing the results panel)

- [**REPORT**] Include two plots with degree distributions in Karate Club and Star Wars
- [**REPORT**] Include two plots with the distribution of shortest path lengths in Karate Club and Star Wars

3.3. Style the network

- Load the Star Wars network
- Make the size of the node larger either for nodes with high degree or nodes with high betweenness
- Change the width and color of edges so it depends o the "scenes" attribute of the network (number of scenes in common). More scenes should mean thicker and darker edges.
- [**REPORT**] Include an image of the network from *Star Wars*, styled as indicated above

4. Use a Cytoscape App (ClusterMaker2)

Cytoscape has "apps" that can be installed and used.

4.1. Install ClusterMaker2

Install ClusterMaker2 (`Apps > App Manager`). You may need to download a jar file from the releases directory of clustermaker2, and then `Install from file ...` in the App Manager.

4.2. Use ClusterMaker2

Run the affinity propagation clustering algorithm in ClusterMaker2 (`Apps > ClusterMaker2 > Affinity Propagation ...`) on the *Star Wars* network.

- Select any temporary folder if prompted
- ClusterMaker2 requires an attribute for the weight: use `Array source = scenes` in *Star Wars*
- Once you run it, the network will have a new attribute in the nodes (in the node table you will see an attribute named `_APCluster`)
- Use the new attribute in the nodes for "Fill color" using a "Discrete mapping" on `_APCluster`. You might have to pick the color for each group, just pick a color for the three largest groups.

[**REPORT**] Include in your report an image of the Star Wars network with the three largest clusters in three different colors (the rest of the nodes can be white).

[**REPORT**] Include a brief commentary on what do you see in these clusters, what do you think they represent and why.

4.3. Apply to Karate Club

Use ClusterMaker2 on the Karate Club

- Here you MUST run the network analyzer first so you can have "Edge betweenness" as an attribute in edges
- Use "Edge betweenness" as the attribute for the weight (`Array source`)
- Run the module, you should get two groups, led by #1 and #34 . Are they close to the actual way in which this club splitted?

[**REPORT**] Include in your report an image of the Karate Club network with nodes painted according to clusters.

[**REPORT**] Include a brief commentary on what do you see in these clusters, and whether they have some relationship with the way in which the Karate Club actually splitted

DELIVER (INDIVIDUALLY)

:warning: First of all, read "delivering your report" on the evaluation guidelines, and check your report against those guidelines before submitting.

Deliver a brief report of at most 4 pages (it can be less!), in PDF format. Organize your report as follows:

- The first section should briefly describe the three networks, including the number of nodes and edges in each one; you can make a table with this.
- Then, you should have one section about the *Karate Club*, one section about *Star Wars*, and one brief section about the *US Companies* network; include in each section the elements marked [**REPORT**] above.

Please be brief, you do not need to write too much, specially if you are not going to say anything substantive: your report can be less than four pages. A "brief commentary" means one or two paragraphs.

Your report should end with the following text:

I hereby declare that all of the text, tables, and figures in this report were produced by myself.