

Tutorial 6: Network Visualization with Gephi

1. Getting started

- (a) Download Gephi (<https://gephi.org>), a software for graph visualization and analysis
- (b) Using Menu **File > Open**, open the Game of Thrones network (`GOT.graphml`) downloaded from the page of the class.
- (c) Using panel **Layout/Spatialisation** (bottom left), change the layout. Try in particular *Fruchterman Reingold*, *Yifan Hu*, *expansion*, *noverlap*, *ForceAtlas 2*. Try to play with the parameters of ForceAtlas 2 (*prevent overlap*, *LinLog mode*, etc.).
- (d) You can move nodes by dragging them. Right-clicking on them provides additional functionalities.
- (e) Zoom in/out using the wheel of your mouse (or of your trackpad. NOT the zoom gesture). The position of the cursor is the center of the zoom
- (f) By clicking with the right button of your mouse on the background and dragging; you can move the window/graph around.
- (g) Using the **Appearance** panel (top left), assign the **size** of nodes to be proportional to their degree.
- (h) Using the **Layout** panel (bottom left), change the layout to adapt to these new sizes. If you use *ForceAtlas 2* layout, you can check to activate **Prevent overlap/Eviter le recouvrement** property.
- (i) Use the button **T** at the bottom to display the name of nodes. Using another option at the bottom, make **node names proportional to node size**.
- (j) Go to **Preview/Previsualisation** tab. Using the **refresh/rafraichir** button to show a finer version of the visualization. You can export it in PDF or PNG.

2. Computing network properties.

- (a) Using Menu **File > Open**, open the airport network (`airportsAndCoord.graphml`) from the page of the class.
- (b) Take a look at the **Data Laboratory/Laboratoire de données** window, accessible by clicking on the tab of the same name at the top of your window. Check the data for both Nodes and Edges (panels on the top left)
- (c) Go back to **Overview** window, and, using the **Statistics** panel (right), compute the average degree. Interpret the degree distribution.
- (d) Go back to the **Data Laboratory** window and observe that new columns in the Nodes table have been created when you computed statistics.
- (e) On the **Appearance** panel, check that you can now change the color and size of nodes based on those statistics. Make the size correspond to the degree and color of countries (this information is present in the dataset).
- (f) Compute the clustering coefficient, the average path length, and the graph density.
- (g) Still using the Statistics panel, computer how many connected components there are in the graph. Would you say that there is a giant component in the graph?

3. Node centralities

- (a) In Gephi, some node centralities need to be computed explicitly, like *PageRank*, but many others are computed when it is convenient. In particular, when you compute the average path length, it computes the *Betweenness*, *Closeness*, and a few other centralities. Compute these centralities, and check that you can attribute node sizes and colors to nodes according to those centralities.

- (b) From the **Data Laboratory** , you can now check the values of centralities of the different nodes in the Nodes table. Try to sort nodes by descending order of some centralities.
- (c) How do you explain the centrality values? For some centralities, nodes of small degrees have very high centrality values. Is it expected or normal? Look for the details of how each centrality is calculated to understand that.
- (d) We would like to see more clearly the difference between some centralities. In the **Appearance** panel, use the **spline** option in the color/size selector and palette choices. Click on Apply button to update the setting on the visualization.

4. Community detection

- (a) Community corresponds to graph clustering, i.e., searching for groups of nodes strongly connected together and more weakly connected to the rest of the graph.
- (b) Compute the modularity statistics.
- (c) Visualize communities using node colors. Note that you should use the **Partition** tab, not **Ranking** .
- (d) Compare visually the communities found with the **country** attribute. Remark similarities and differences
- (e) Using the **Data Laboratory** , you can duplicate the column (in the bottom) to copy the values into a new column.
- (f) Recompute the modularity with a different Resolution parameter, compare the community assignments and search for larger/smaller communities.

5. Spatial network

- (a) From Menu **Tools > Plugins** , install the plugin called **geolayout** . You need to restart Gephi after the installation to use the plugin. (Remember to save the Gephi project if you want to keep your work!)
- (b) You can now find **geolayout** among layouts. Use **Mercator** projection to position nodes according to their **longitude** and **latitude**.
- (c) Make the graph readable (e.g., avoid nodes being too large/small, etc.)

6. Filtering nodes and edges

- (a) Using the **Filter** option on the right, filter nodes according to their closeness centrality and keep only those with a large value (**Library > Attribute > Range**).
- (b) Apply a non-spatial layout again. How does it behave regarding the filtered graph?
- (c) Try different options to filter the network. Filtering what seems to make sense to create a readable and interpretable graph.
- (d) In the **Preview** tab, you can export the network visualization into PNG or PDF format.

Acknowledgement: This tutorial is based on the *Learning how to use Gephi* tutorial, prepared by Rémy Cazabet from Université Claude Bernard Lyon 1.