Experimenting with Gephi

The objective of those exercises is to become familiar with the functionalities of Gephi. Do not hesitate to ask questions if there is something you do not understand

1. Visualizing small networks with Gephi
   (a) Using Menu `File>Open`, open a small network, downloaded from the page of the class.
   (b) Using the bottom left panel, 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.).
   (c) You can move nodes by dragging them. Right clicking on them provide additional functionalities.
   (d) Zoom in/out using the wheel of your mouse. The position of the cursor is the center of the zoom
   (e) By clicking with the right button of your mouse on the background and dragging, you can move the window/graph around.
   (f) Using the *Appearance* panel(top left), assign the **size** of nodes to be proportional to their degree.
   (g) Using the *Layout* panel (bottom left), change the layout to adapt to these new sizes.
   (h) 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**.
   (i) Using the `Statistics` tab of the right panel, compute PageRank
   (j) Using the top left panel, you can now assign a color scale to nodes corresponding to their PageRank score.
   (k) Have a look at the `Data Laboratory` window, accessible by clicking on the button of the same name at the top of your window. Check the data for both Nodes and Edges (panels on the top left)
   (l) Go back to `Overview` window, and, using the right panel, compute the different statistics. Observe the generated plots.
   (m) Go back to the `Data Laboratory` window, and observe that new columns have been created when you computed statistics.
   (n) Check that you can now change the color and size of nodes (overview, top-left) based on those statistics.
   (o) Would you say (informally, without comparing with a null model at this point) that the network is a *small world* network ?
2. Visualizing larger networks with Gephi, and spatial networks
   (a) Using Menu `File>Open`, open the network called airports from the page of the class.
   (b) Manipulate it as the previous one
   (c) From Menu `Tools>Plugins`, install the plugin called `geolayout`.
   (d) In the layout panel, you now have a new layout called geolayout. Use the Mercator projection to position nodes according to a latitude and longitude positions.
   (e) Check that you can change the color of nodes according to an attribute, for instance the country of the airport in this dataset.

(f) Use the `Filters` tab in the right panel to filter some nodes and/or edges in your graph. For instance, remove the nodes of lower degree, edges of lower weight, edges of higher betweenness, etc.

3. Application
   (a) Using all what you have learned, create a nice looking visualization of your favorite network.
   (b) Share your visualization with others

4. Network Description
   (a) On the airport dataset, compute the degree distribution using the button *Average Degree* from the Statistics panel. Does it looks more like a normal distribution or a power law one ?
   (b) Using the corresponding button in the Statistics Panel, compute the Average Path length, the average clustering coefficient and the Graph density. Would you say (without computation) that this network is a *small world* network ?
   (c) 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 ?

5. Network Centralities
   (a) In Gephi, some 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 Betweeness, Closeness, and a few other centralities. Compute those 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. Try to sort nodes by descending order of some centralities.
   (c) How do you explain that, for some centralities, small nodes have very high values ? Is it a bug ?
   (d) If you ignore those cases (you can even remove them from the data), are the ordering of nodes according to the different scores very similar, completely different, or somewhat correlated ?
   (e) Let's say that we would like to check that nodes of higher closeness are indeed central in the network, as seen in the lecture. Color nodes according to centrality.
   (f) All nodes seems probably to be of the same color. The problem comes from nodes in small connected components having unrelevant values. To remove them, a simple way is to use the filter tools to filter them out (based on closeness of on component ID), and then to export the obtained graph in a new workspace using the corresponding option (Filters panel, button with an arrow at the top)
   (g) The result is still not convincing. Change the color palette so that we have one color at one hand, another color at the other hand, and white in the middel, to clearly see the differences.
   (h) Finally, we would like to see more clearly the difference between the nodes of highest values. Using the *spline* option of the color chooser, try several possibilities and check the differences.
   (i) Now that you can clearly see the "center/periphery" effect of closeness, compare it between the forceAtlas 2 visualization and the GeoLayout visualization.