



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

A Network Tour of Data Sciences: Flight Routes Project

Team 47

Frank Schmutz
Harold Sussmilch
Josef Weber
Léo Capiod

Introduction

Through the semester, we have worked on Flight Routes. During the project time, one of the main questions that had to be answered was to find a proper way to label our data as no labels were given in the initial data set. We have chosen to choose a geographical labelling based on the continent location of the airports. The implementation of the solution is presented below and in the code. A interactive visualization of the airports allowing the user to see the airports on a map is also presented. Finally, based on the continent labeling of the airports, we have been able to present some analysis on the connections between continents.

Visualization

Labeling Continents

Using another dataset we managed to be able to label each node based on their underlying country-value. Now each continent can be shown in a different color. The labeling can be turned on and off clicking the top checkbox in the left section.

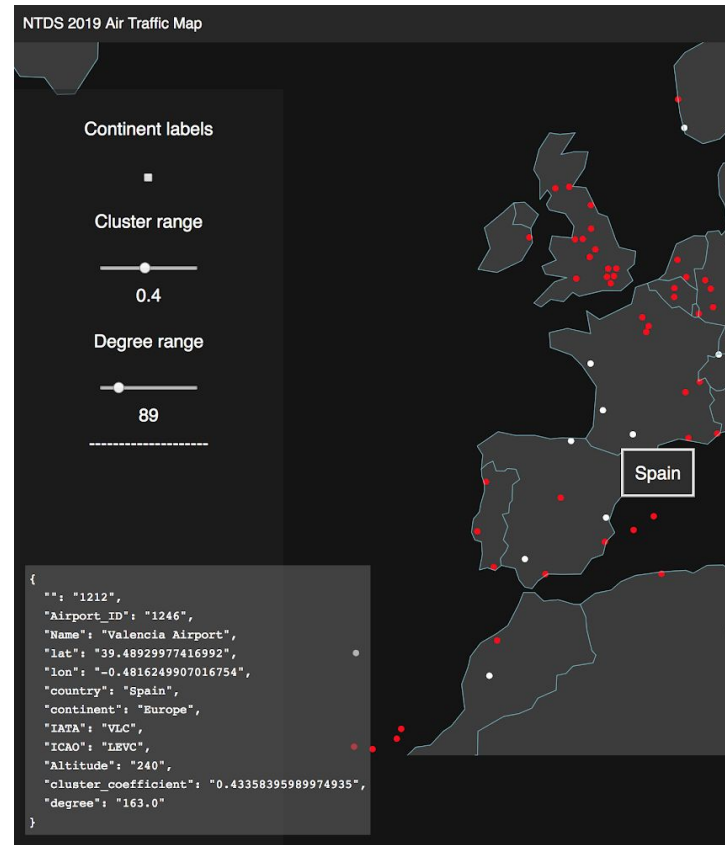
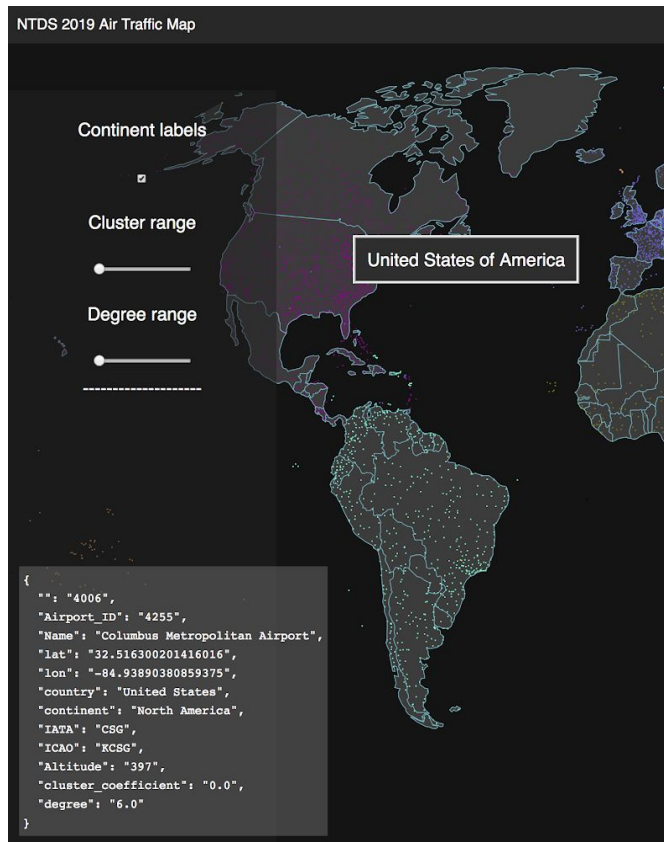
Clustering Property

To be able to filter nodes we introduced the user multiple functionalities to do so. Putting the slider to the left position will deactivate the selected filter.

First of all we can filter for cluster coefficients. Since we calculated the cluster coefficient for each node we used that value as a filter variable. The more the slider gets to the right the more nodes are shown in the color red. Note: the filtered nodes won't disappear - they will just turn red.

Showing Stats

The next filter is the degree filter. Based on the node's degree an airport has, we can either select nodes that have more neighbours than a specific amount of degree. This filter actually makes unwanted nodes disappear. This filter can be combined with the filter above. Below are presented views of the filters:

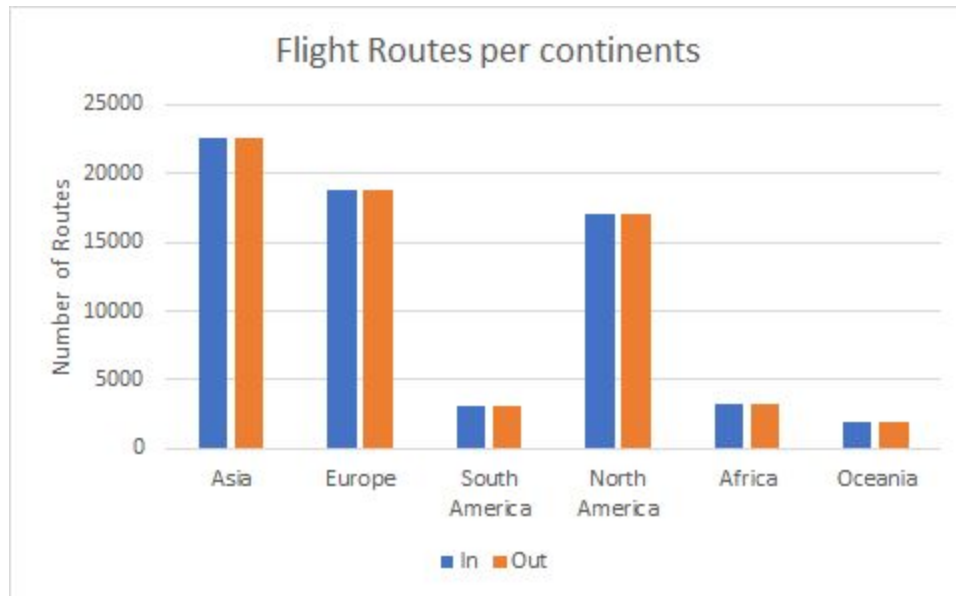


Hovering over an airport displays all node information that we can provide in the bottom-left corner section. On leave will remove the content.

Intercontinental flight routes and airport classification:

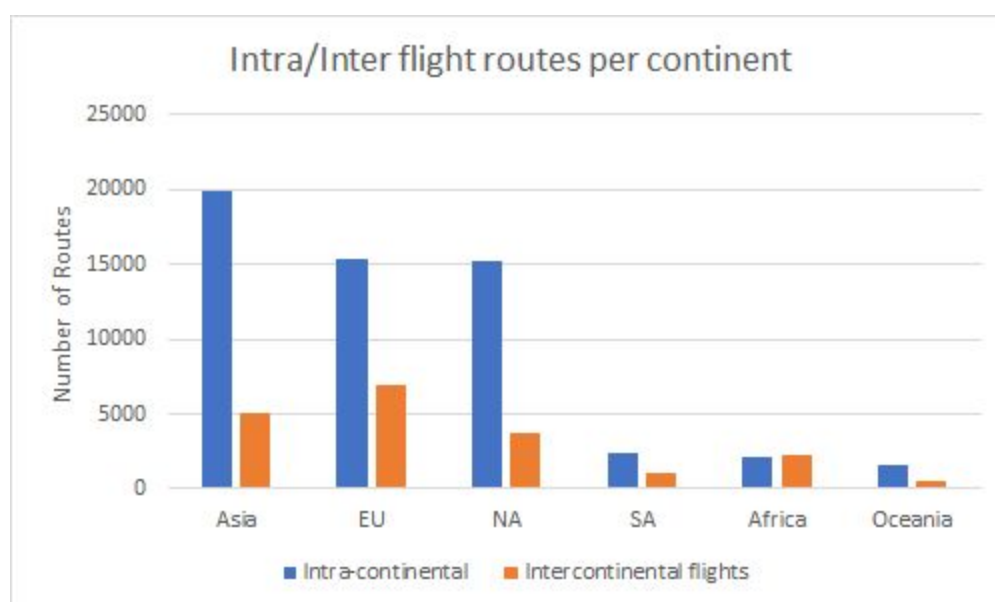
From the data labelling presented above, depending on the location continent of the airports, we wanted to know how continents were connected between each other, how many routes could be considered as intercontinental, how many airports were intercontinental hubs... and to analyze these results.

First of all, we used the data set to observe whether we could see any trends in the number and type of routes depending on geographic location.



Clearly, Asia, Europe and North America appear to have the more flight route connections. It is logical regarding the wealth of these geographical regions and the huge population in Asia. As expected, the number of incoming flights is really close from the number of outgoing flights for each continent.

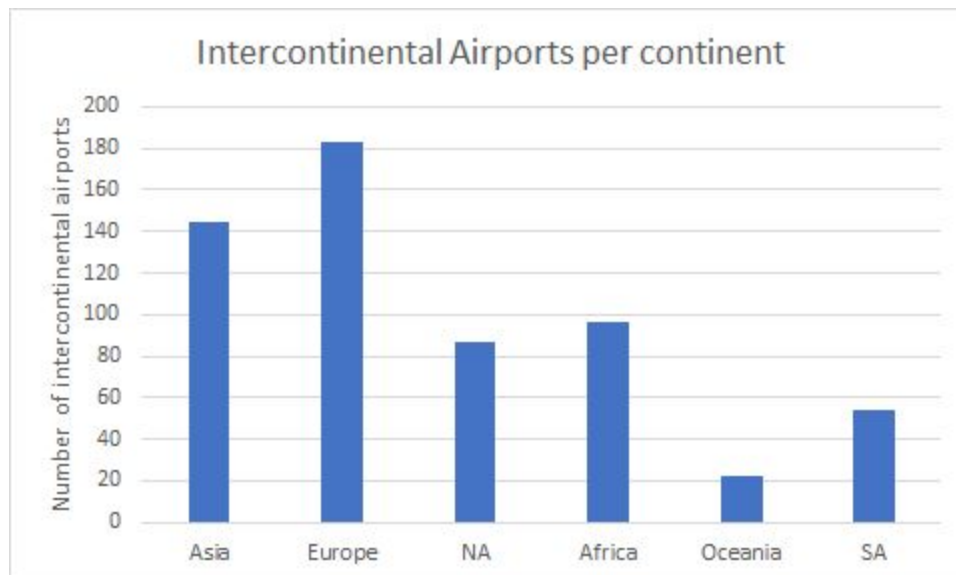
Then, we decided to classify airports and routes in intra-continental and intercontinental categories to see which continents had more continental connections or intercontinental connections.



From the above graph, it is possible to draw some conclusions from simple observations and to analyze them:

- North America has a number of intra continental connections similar to the one of Europe, however Europe is a clearly bigger hub for intercontinental flights. This can be explained with a simple fact: North America is a bigger continent than Europe, the plane is a lot more efficient than the train in countries like Canada, the USA and Mexico.
- Africa has more intercontinental routes than intracontinental. This can be explained by the relatively poor network of airports in Africa, where many countries are still developing countries, and by the strong connections between Southern Europe and North Africa which are counted as international routes.

While Asia has the biggest number of airports, it is however Europe that has the biggest number of intercontinental airports as shown by the graph below. It confirms the last graph.



It is interesting to remark that Africa has more intercontinental airports than North America. As before, the connections between many destinations of Northern Africa with Europe explain the relatively high number of intercontinental airports in Africa. The relatively low number of intercontinental airports in the US can be explained because of the presence of a few huge international hubs in North America that gather most of the intercontinental routes.

Finally, we were able to plot the connected graph of intercontinental routes between continents. Antarctica has been voluntarily ignored because of the very few number of airports on this

continent. The weights have been plotted according to the percentage of intercontinental routes between two continents on the total intercontinental routes worldwide.

The very weak connections between continents are represented but the associated weight is not given as they can be considered as negligible.

The graph is not complete: indeed, one edge between Oceania and Europe is missing. In the data set, islands in Oceania belonging to European countries like New Caledonia have been considered as Oceanian island and no direct routes between Oceania and continental Europe exist because the trip would be too long.

The graph is shown below. For a better representation, we have chosen to draw the edges so that the more connections exist between continents, the bigger the width of the corresponding edge is.

