Data Analytics and Visualisation

Assignment

Quentin FEVER – S383387

The complete code of this study is available on my github at this address :

https://github.com/quent1fvr/Airline_network_DAV

Department: Applied Artificial Intelligence

Date : 09/01/2022

Information categorisation:
Open

Table of Contents

Objectives

The purpose of this study is to perform a multi-scale analysis of domestic air traffic for selected countries: United States, China, United Kingdom and Australia. For this, we will mainly use the NetworkX and Cartopy libraries under python

Airlines networks

To create a graph that represents the air traffic network, we consider airports as the nodes of the graph and flights as the links between these nodes Each node is assigned a unique identifier that represents the airport (e.g., the airport's IATA code). We take into account the size of the airports by using different sized circles to represent each node: larger airports have a larger circle, as they are linked to a larger number of flights. We use the NetworkX and Cartopy libraries to implement these graphs.

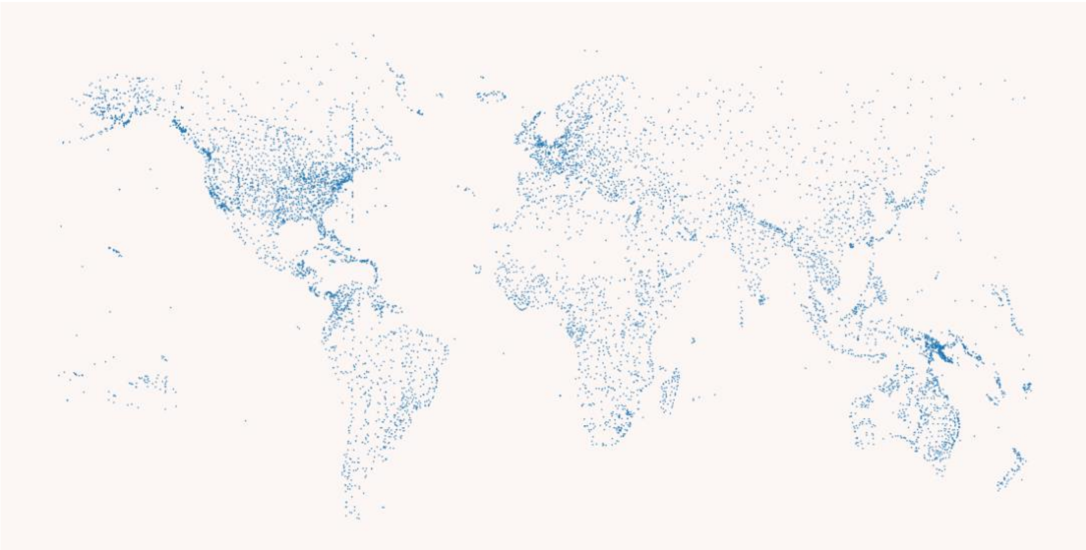

Figure 1: Distribution of world airports on a map

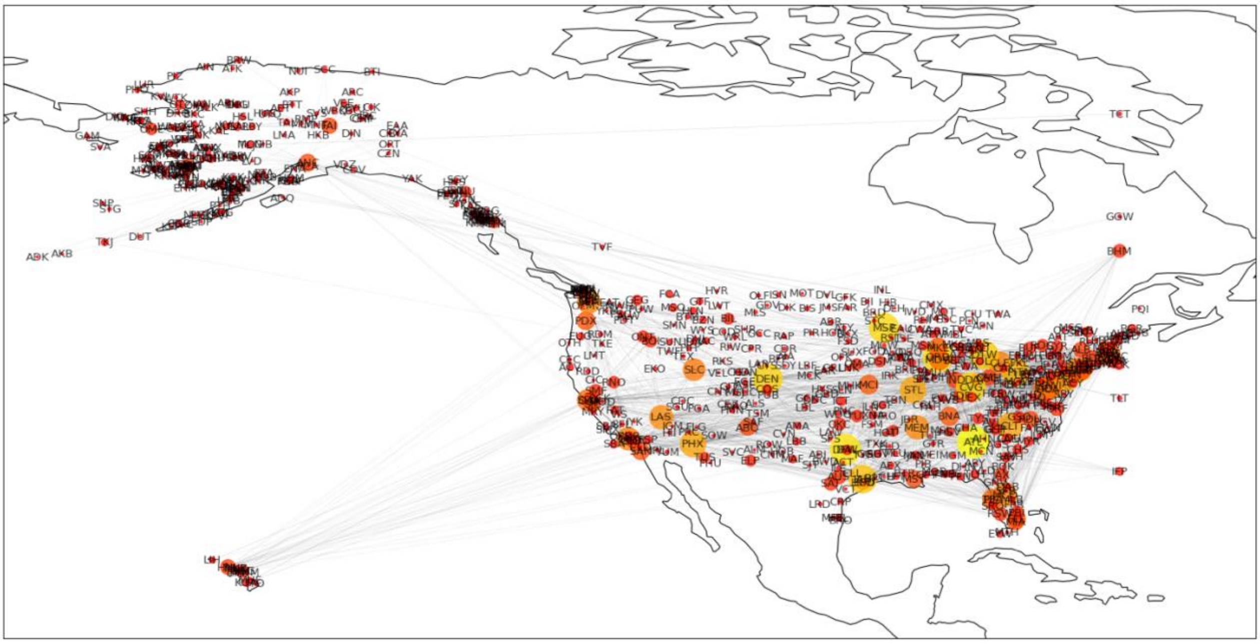

Figure 2: Airline network of the USA

We can wonder what are the flights made on the island off the west coast of the United States, it is the only American state of the USA, the archipelago of Hawaii. We can therefore consider as a domestic flight of the United States.

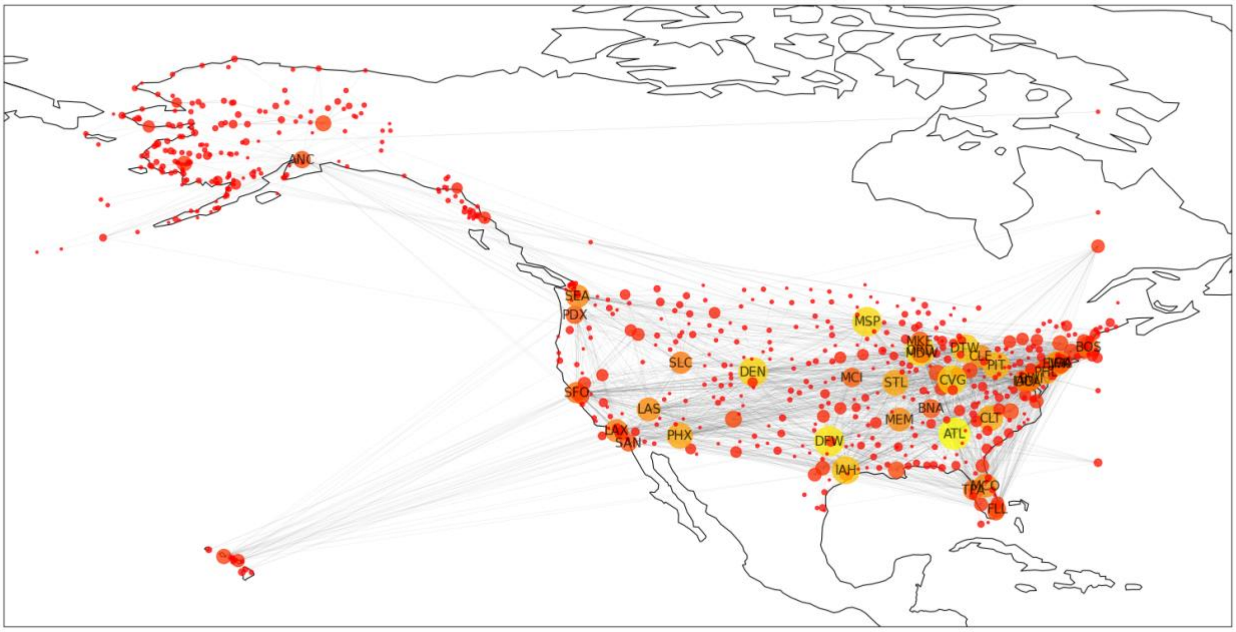

Figure 3: Airline network of the USA where only airports with the highest number of connections are identified

Some airports were missing in the initial dataset and have been added directly to the file.
Moreover, after projection, we notice that some airports in the United States are misplaced, a correction has been made manually, the figures below represent a more consistent network.

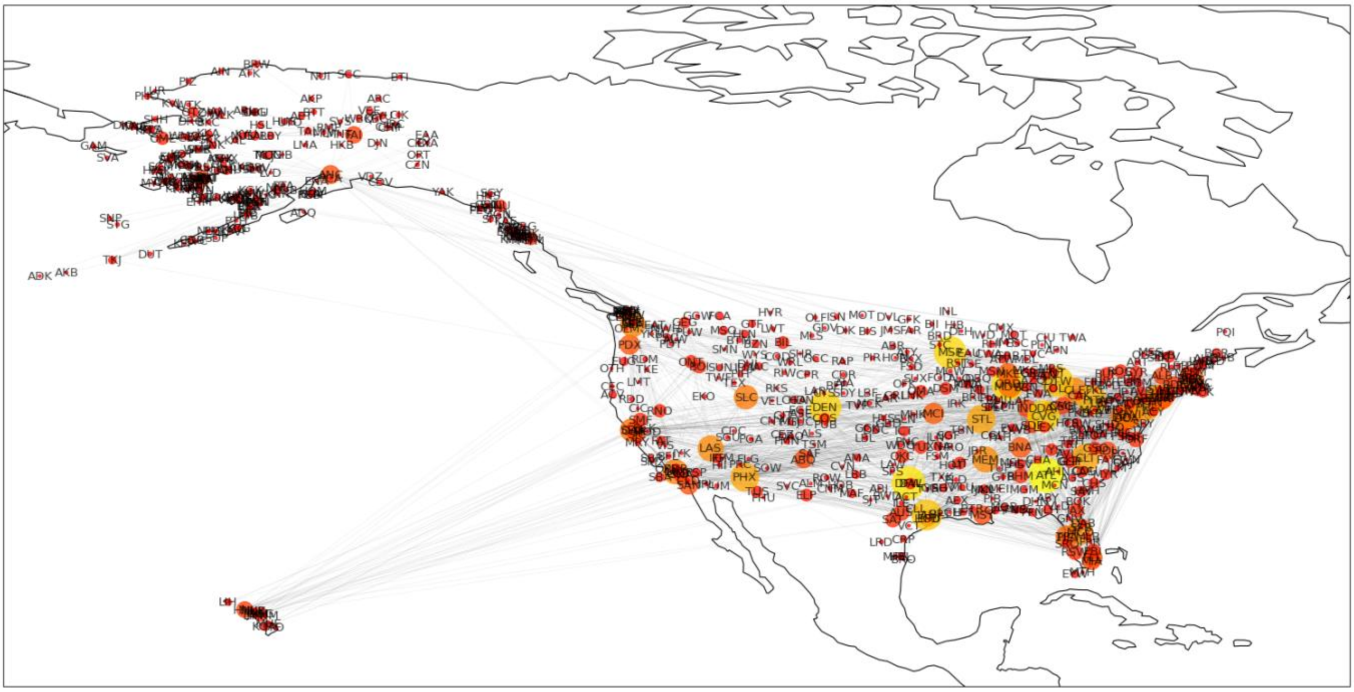

Figure 4: US Airlines network corrected

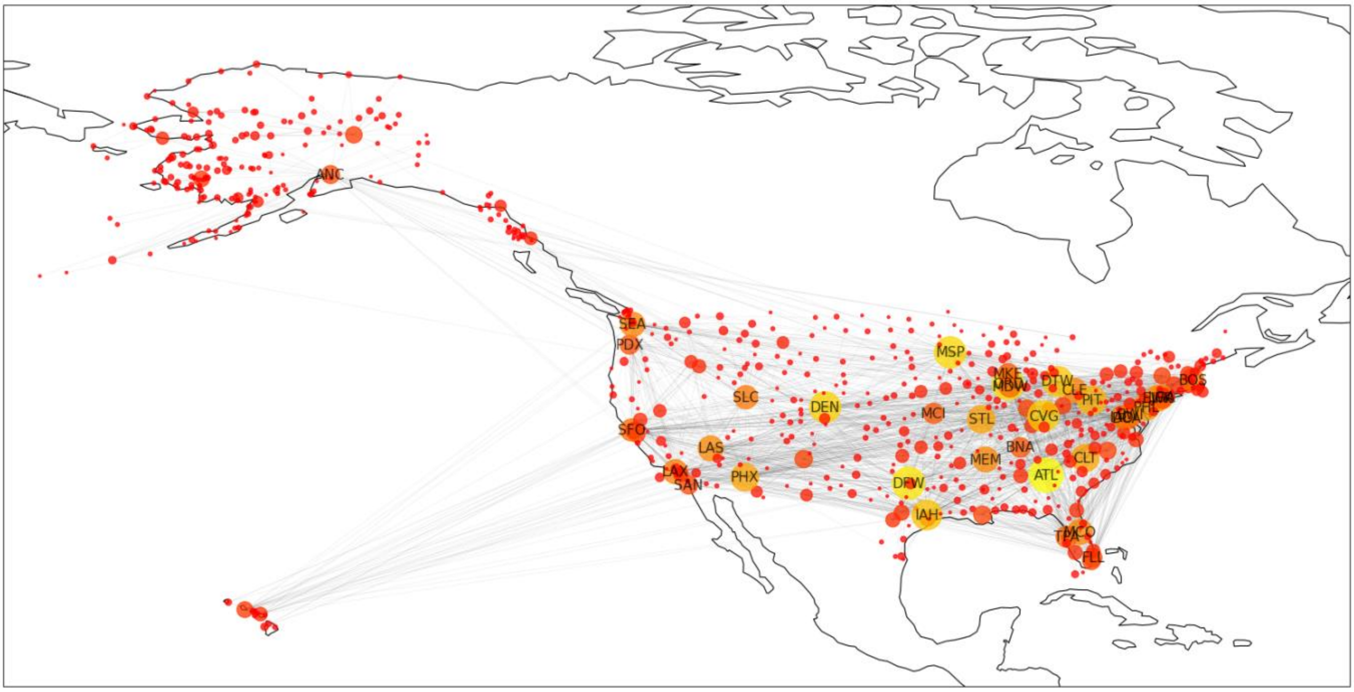

Figure 5: Airline network of the USA where only airports with the highest number of connections are identified

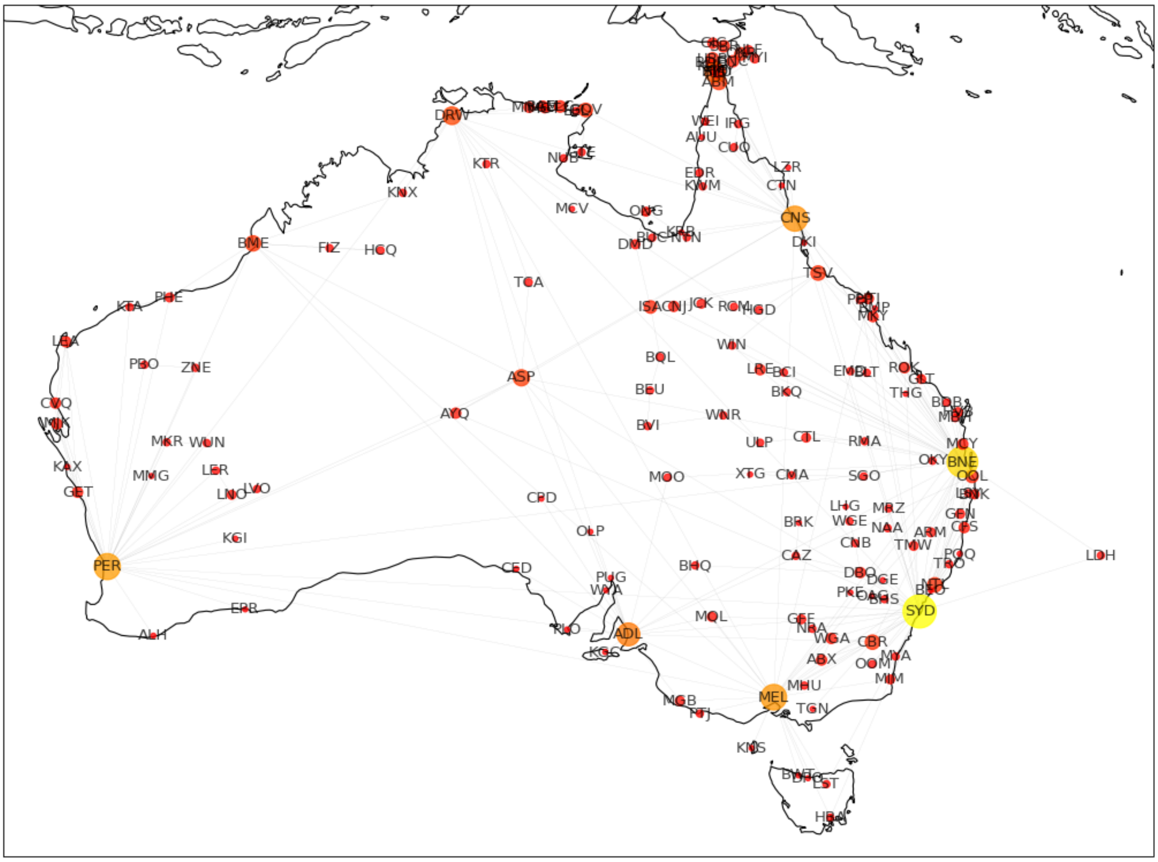

Figure 6: Airline network of Australia

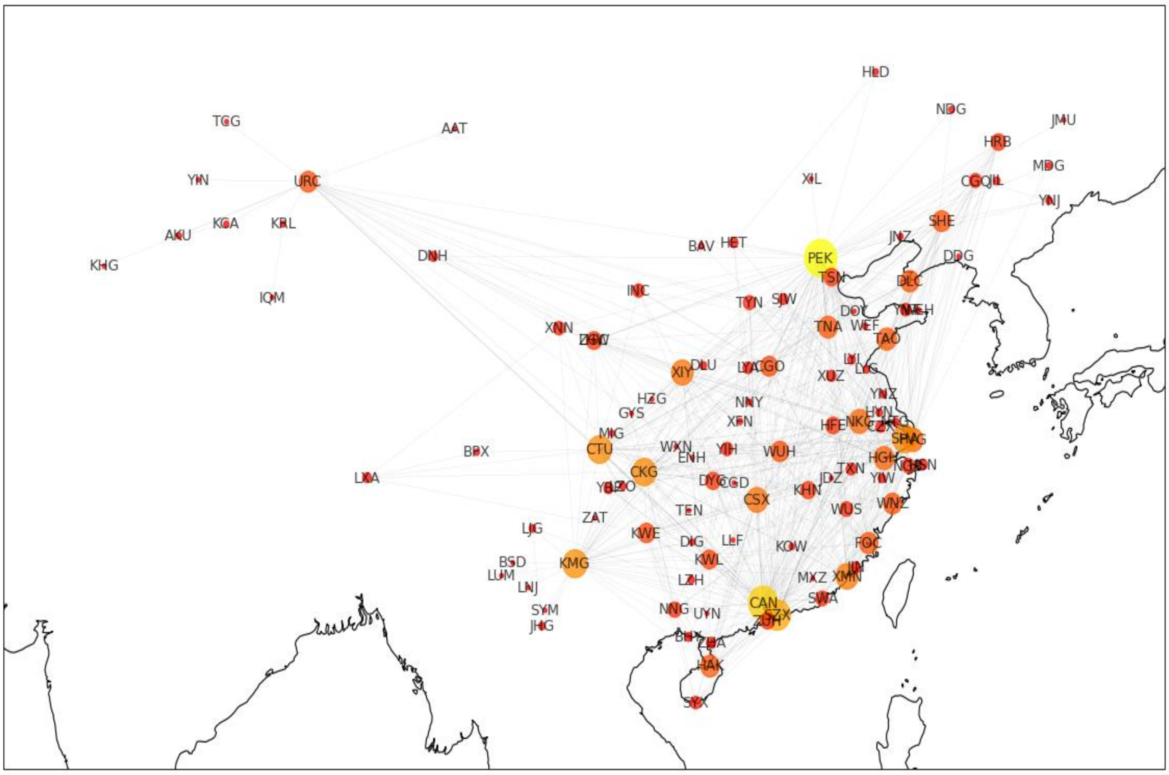

Figure 7: Airline Network of China

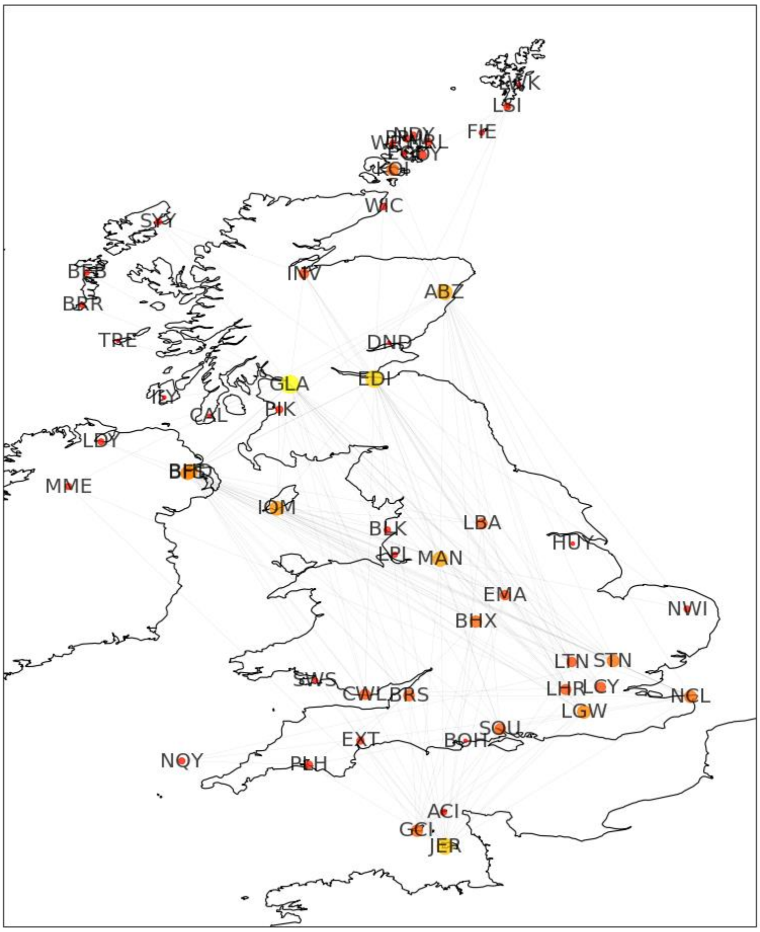

Figure 8: Airline network of the UK

Graphic results

Degree distribution

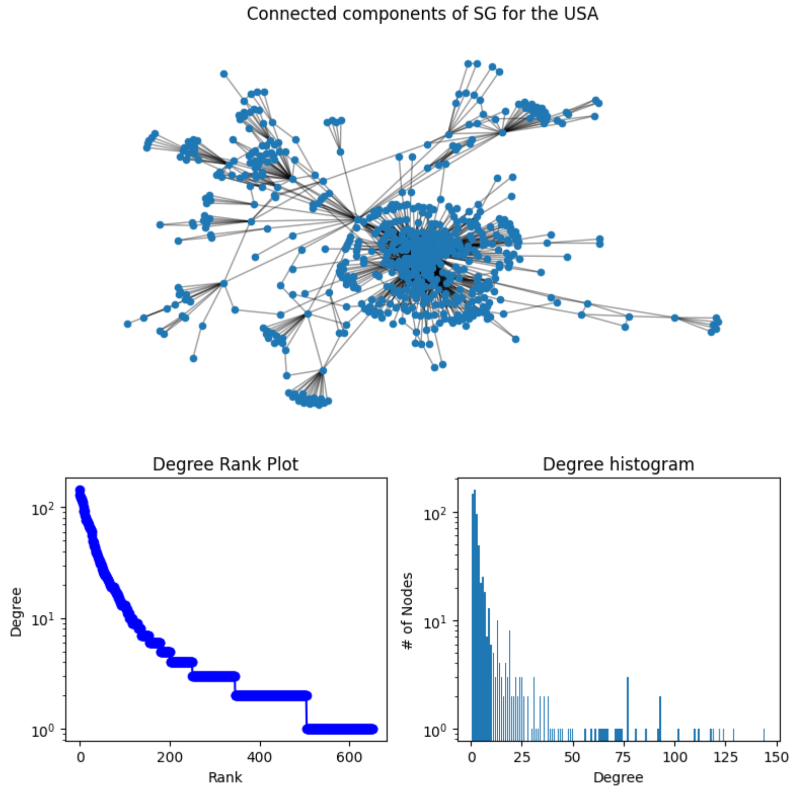

Figure 9: Degree Distribution of the United States

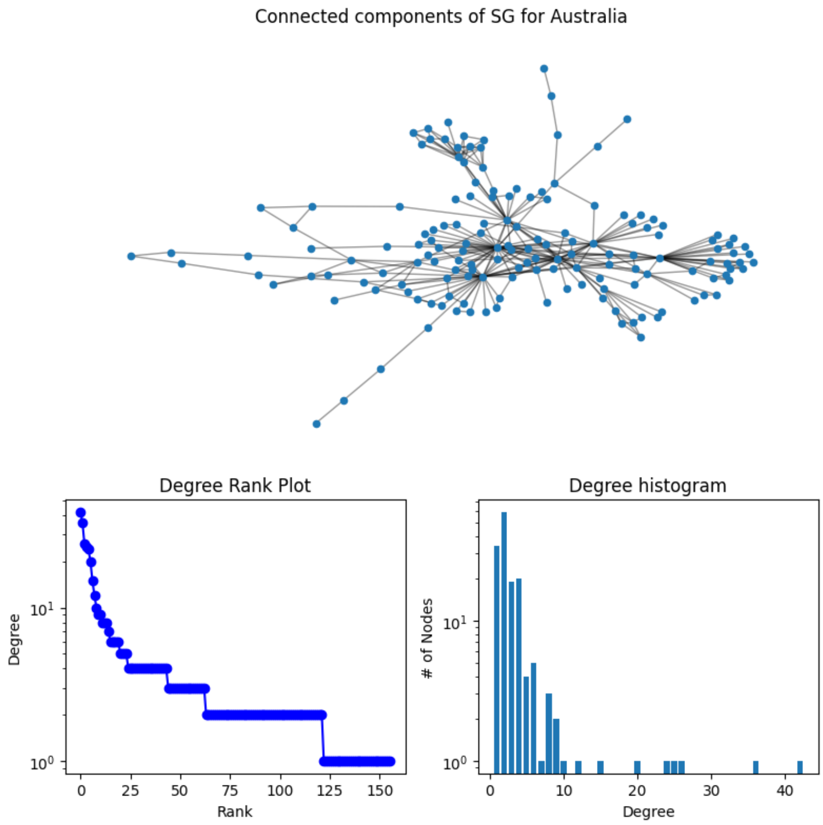

Figure 10: Degree Distribution of the Australia

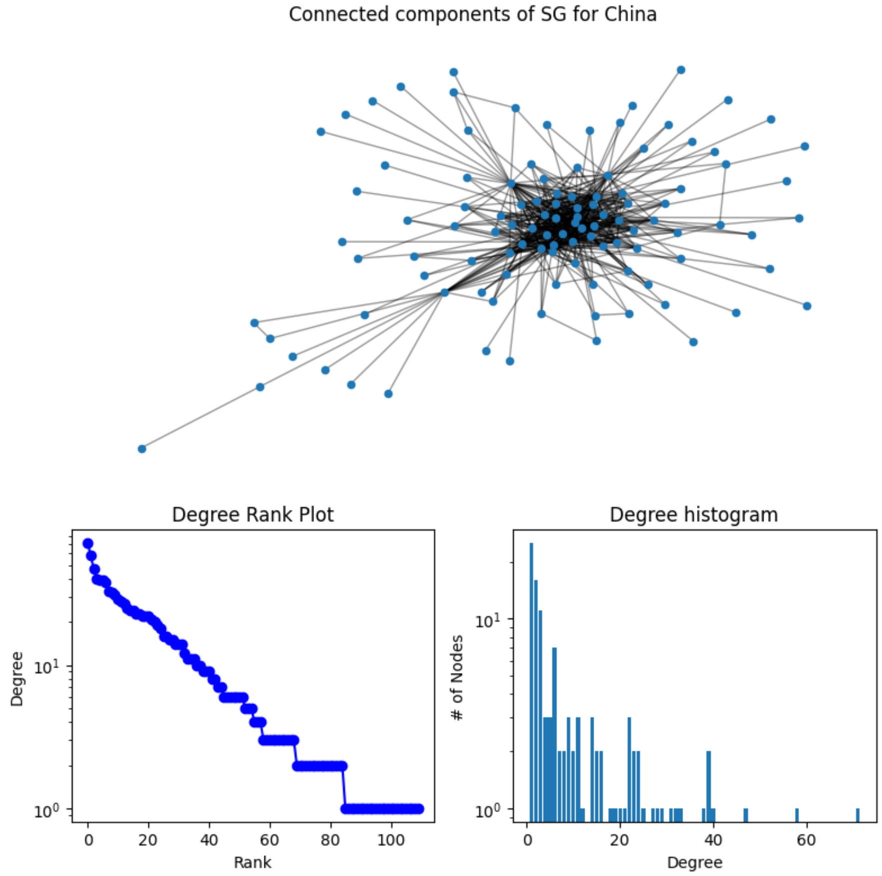

Figure 11: Degree Distribution of China

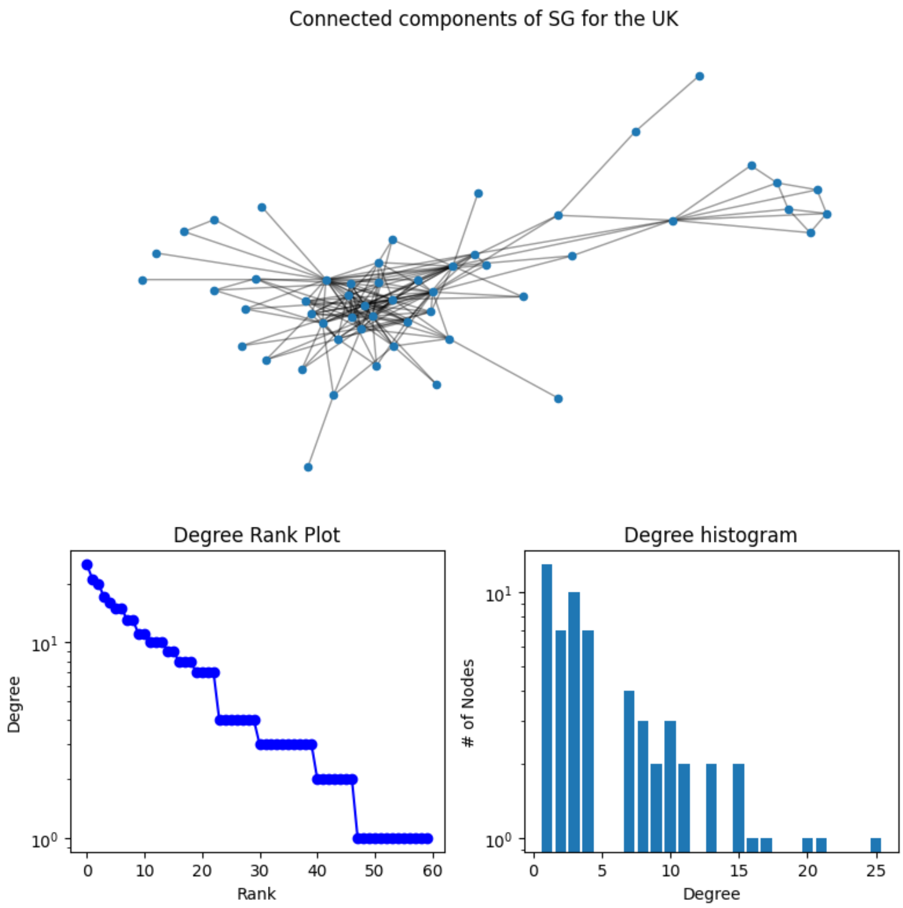

Figure 12: Degree Distribution of the UK

Degree vs. betweenness distribution

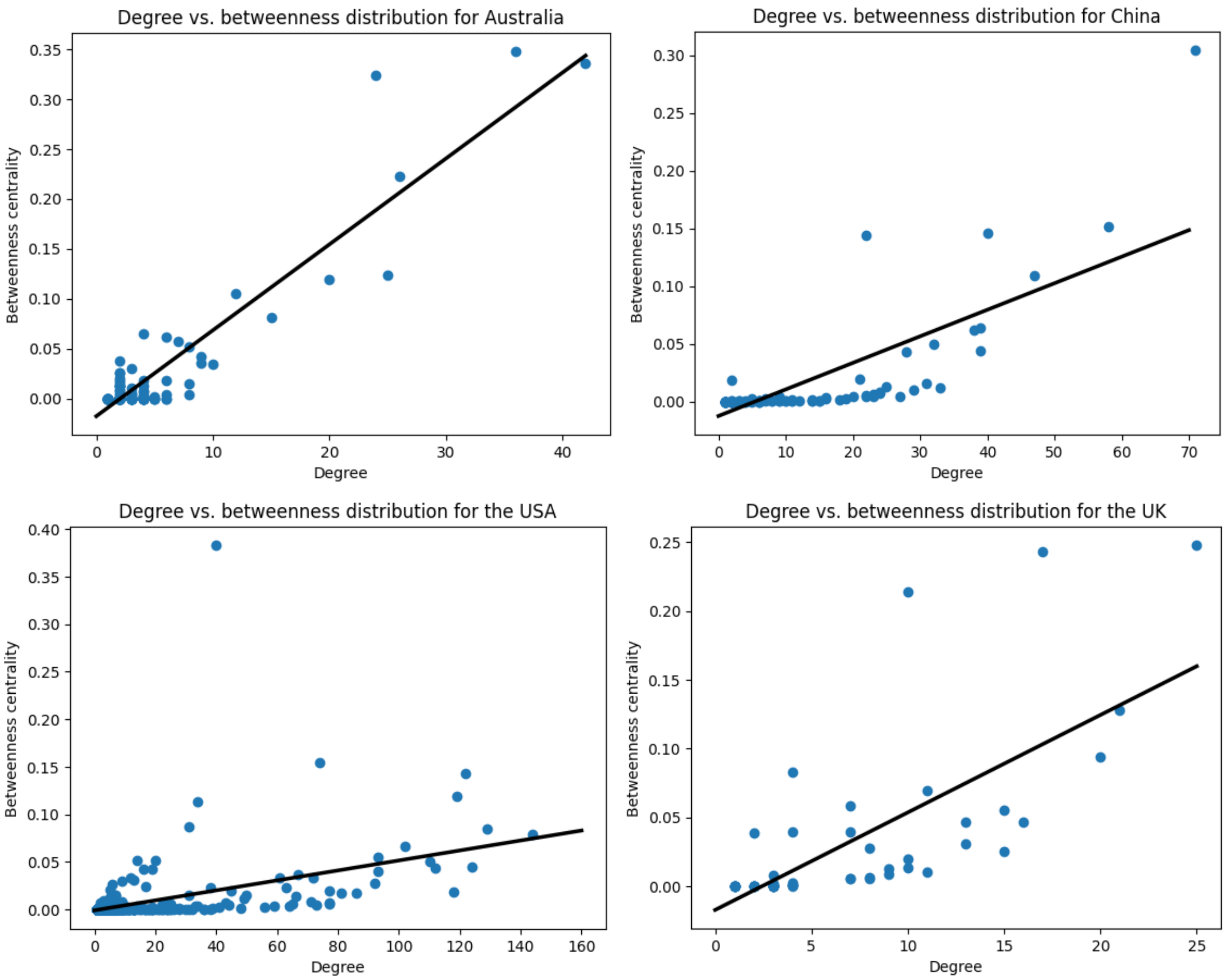

Figure 13: Degree vs betweenness distribution for each country

Assortativity (degree-degree correlation)

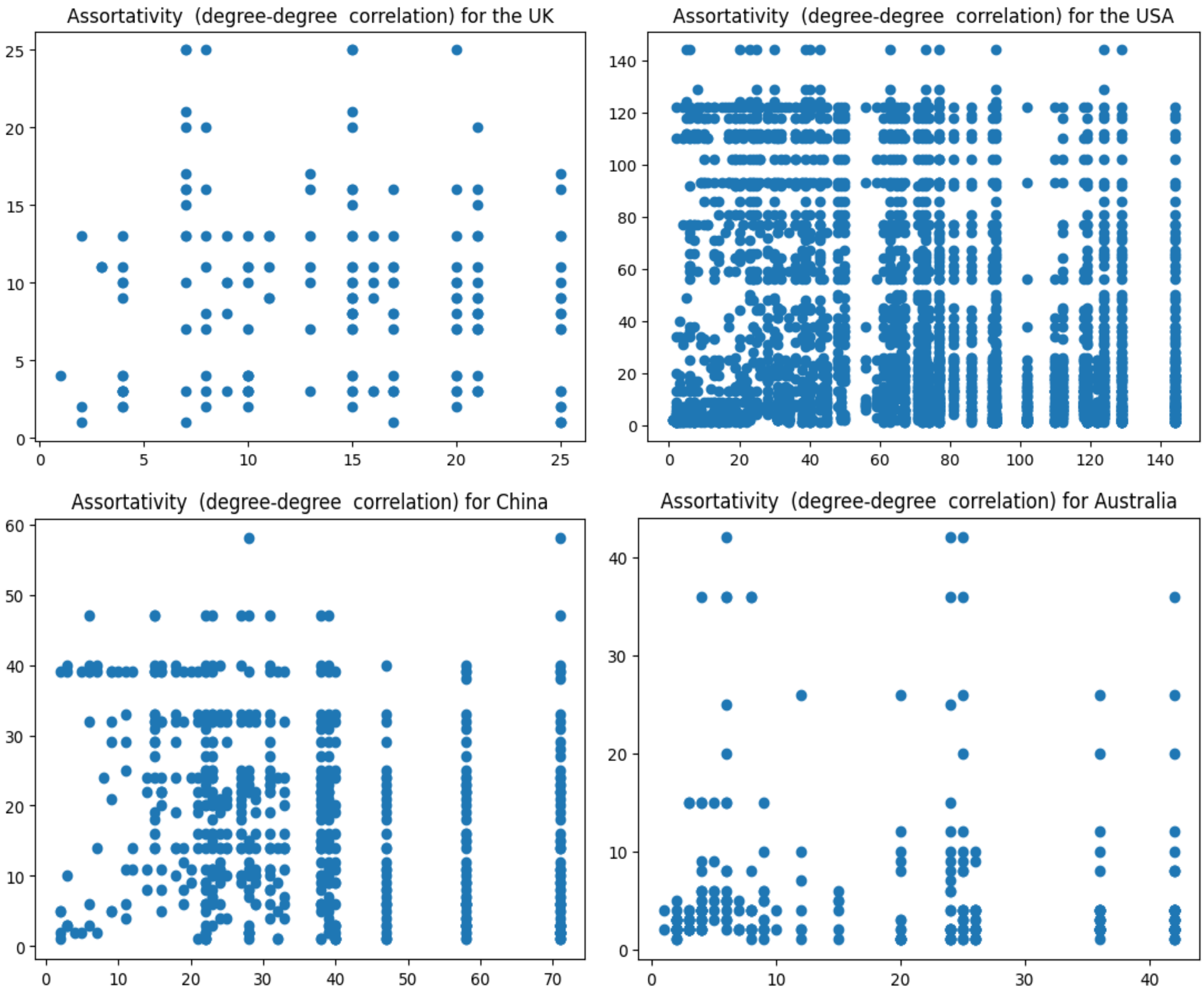

Figure 14: Degree-Degree correlation for each country

Country	Assotativity value
USA	-0,2023
Austalia	-0,2304
China	-0,3972
UK	-0,1791

Figure 15: Assortativity value based on the coefficient of Pearson

The core community size

<u>Clustering</u>

To process this part, an ascending hierarchical classification algorithm (clustering) was applied to each graph, the results of which are below, which highlighted the different communities of graphs.

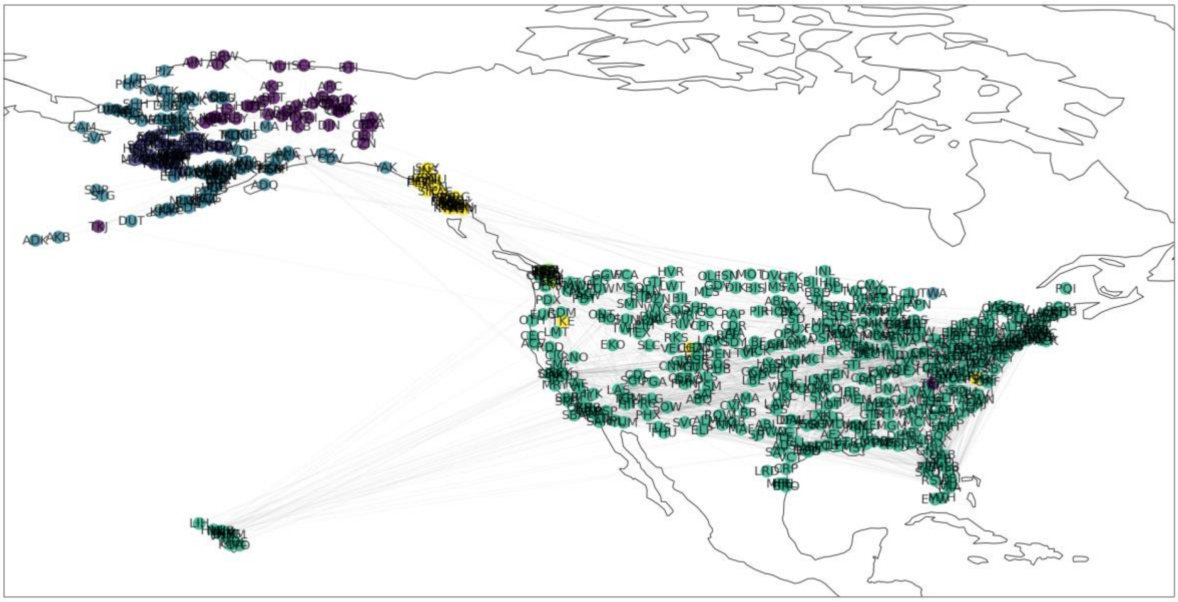

Figure 16: Clustering on USA airline network (4 clusters)

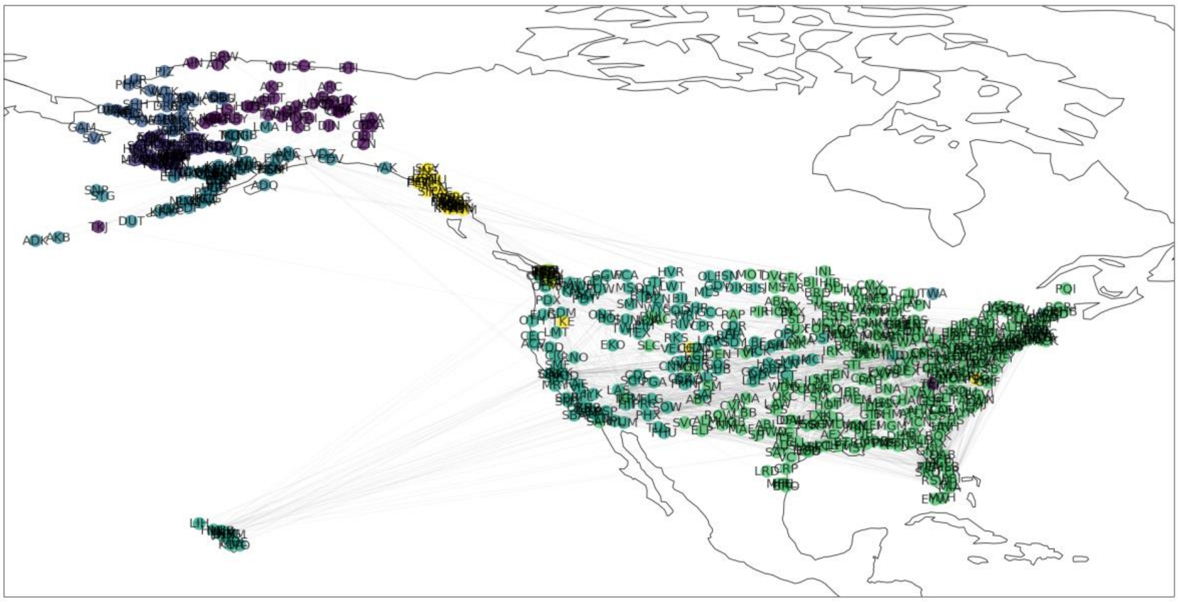

Figure 17: Clustering on USA airline network (5 clusters)

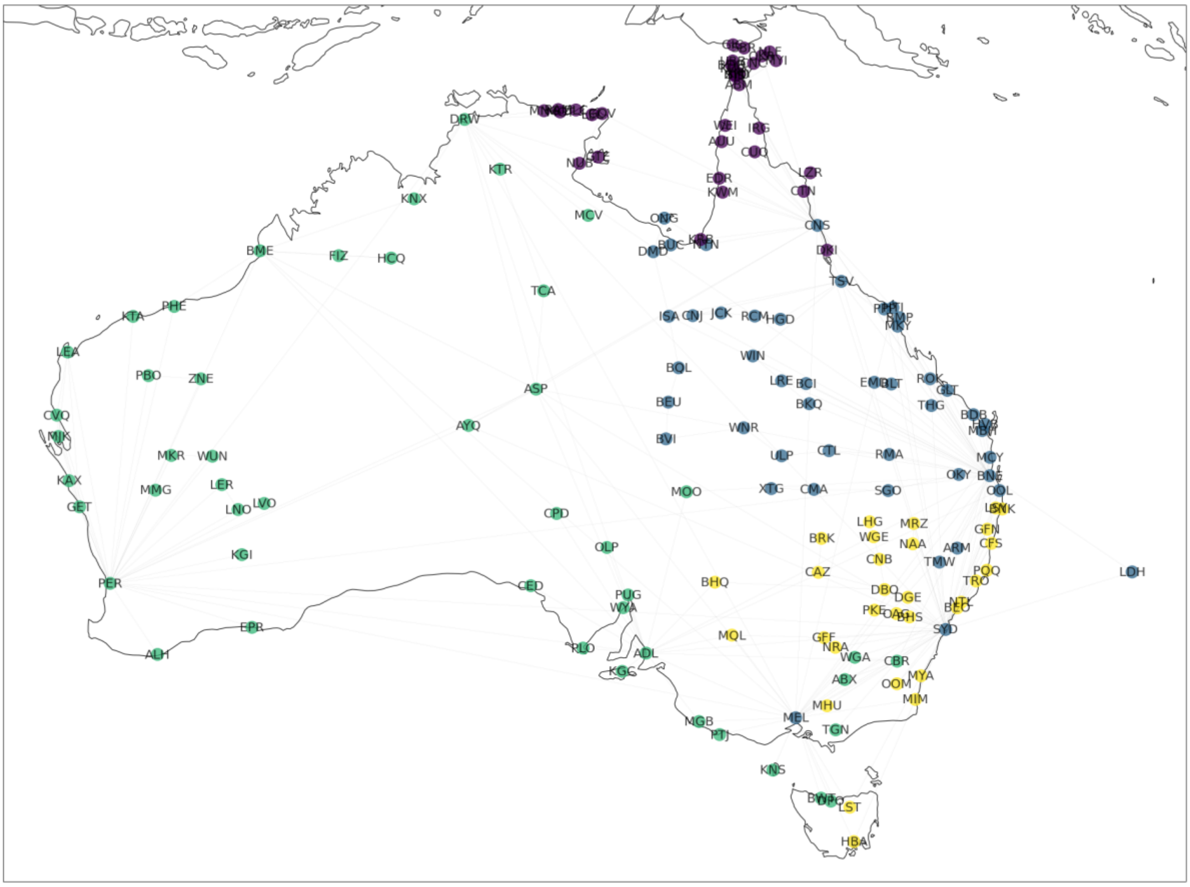

Figure 18: Clustering on Australia airline network (4 clusters)

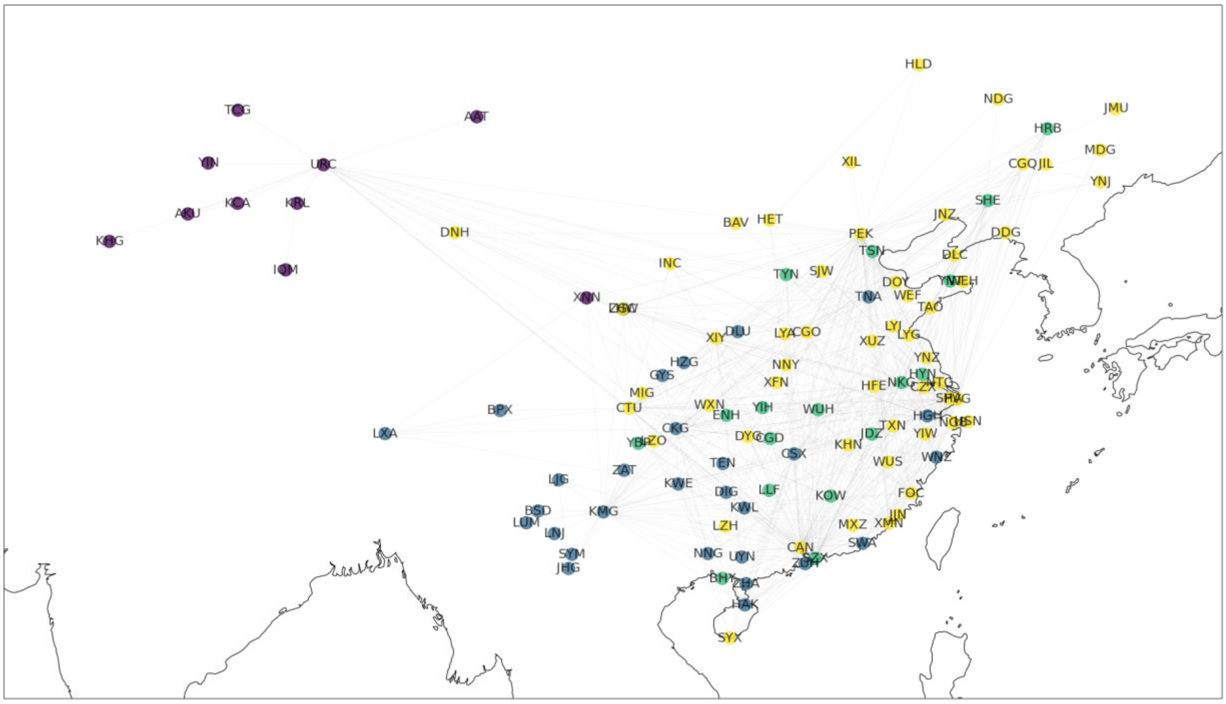

Figure 19: Clustering on China airline network (4 clusters)

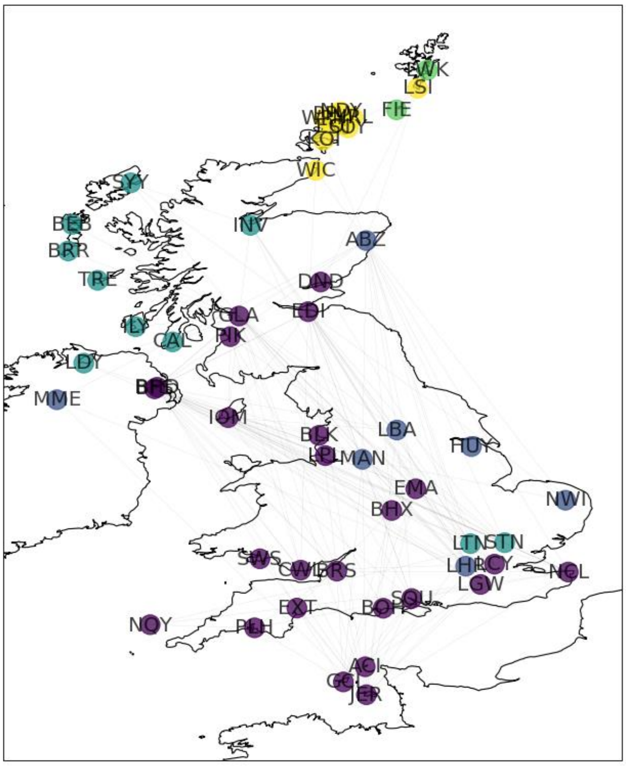

Figure 20: Clustering on the UK airline network (4 clusters)

It is interesting to note that for each network, the identified clusters are geographically close. Indeed, it seems understandable that people who travel on domestic flights do not necessarily go far away (work, weekends), hence the links between the different airports that are geographically close.

To identify the core community of a graph from the clusters, we can use a centrality indicator such as the degree of centrality.

The degree of centrality measures the "importance" of a node in the graph based on the number of links it has with other nodes. The more links a node has, the more central it is considered to be in the graph.

Proximity centrality measures how far a node is from other nodes in the graph. The closer a node is to other nodes, the more central it will be considered in the graph.

To identify the core community from the clusters, we can:

3) Apply a clustering method to identify clusters or groups of flights that have close links to each other.

2)Compute the centrality indicator for each node in the graph.

3)Identify clusters that have a high degree of centrality or proximity centrality. These clusters will be considered as part of the central community.

The Core Periphery structure

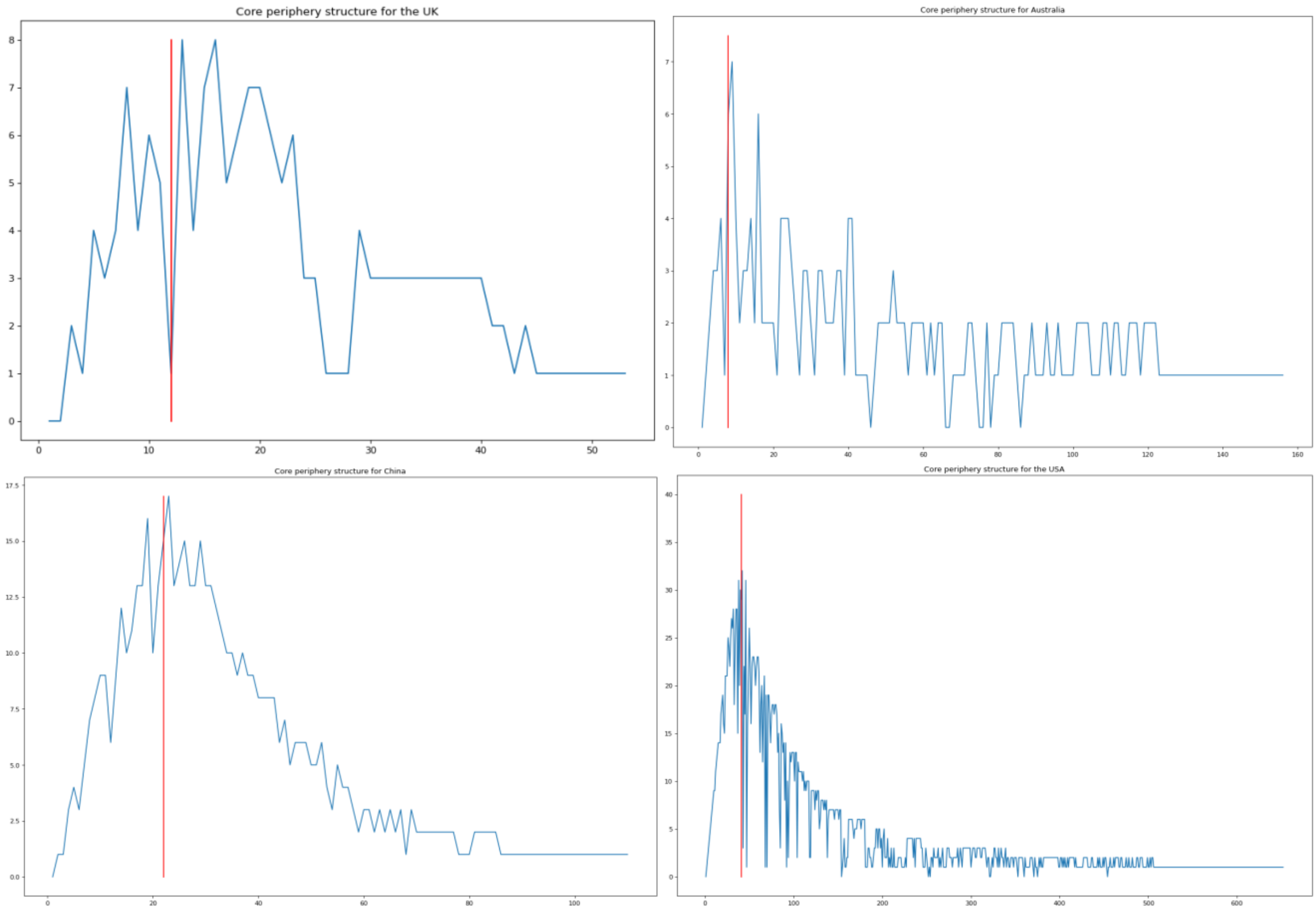

Figure 21: Core Community structure for each country

Country	networkx method	Core Periphery Structure with smoothing curves	Clustering (inaccurate)
USA	38	38	291
China	25	22	54
UK	18	15	25
Australia	7	6	48

Figure 22: summary table of the core community size

Analysis of the results

To analyze the results obtained previously, an analysis at different scales will be presented: macro-scale (statistical analysis), 2) meso-scale (community analysis), 3) node-level (centrality analysis).

Macro-scale

From the distribution of airports in the world (figure1), we can see that the airport network of the United States is extremely developed throughout the U.S. including Alaska, while that of China is more scattered (the west and the coast is much more developed than mainland China) which influence the network of both countries. For example, China's highest degree node is 71 versus 140 for the United States. The United Kingdom is a country fairly well connected everywhere in the territory, it is less dense because the territory covered is infinitely smaller than China or the United States. The situation in Australia is quite homogeneous at first sight of the graphs

For each graph, we can visually see that some airports are much more connected than others and have more or less connections. This highlights the Kagging Pauper effect. This translates into airports that have fewer links/arcs (and therefore less centrality) than other airports in the graph.

The Kagging pauper effect is a concept that has been developed in the context of graph theory to explain why some parts of a network are less connected than others.

Specifically, the Kagging pauper effect can be observed in the air transportation graph when some airports have fewer direct flights to other airports than others. This can be due to several factors, such as the geographical distance between airports, their size or their hub status.

For example, if an airport is located in a remote or isolated area, it is less likely to have direct flights to other airports than if it is located in a more populated or better served area. Similarly, a smaller airport will have fewer direct flights to other airports than a larger airport or hub.

Meso-scale

The mesoscopic scale can be analyzed from the communities that have been formed by clustering and hierarchical ascending classification.
For each country, it can be noted that the clusters of communities in an air transport graph are geographically close.
Indeed, in the case of air transportation, airports that are geographically close are generally better connected to each other and have a higher probability of serving similar destinations. For example, two airports located in the same or neighboring regions are more likely to have direct flights between them and to serve the same destinations than if they were far apart.

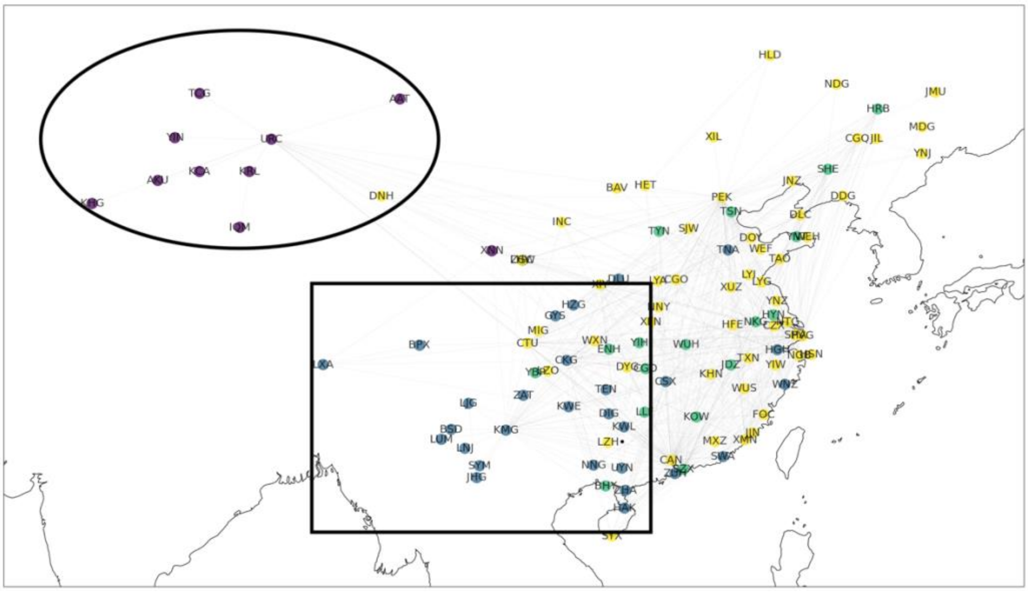

Figure 23: Illustration with the communities of China

For the two framed areas, we see that the clusters formed are quite concentrated in one geographical area. However, it is interesting to note that when a group of airports is very far from the rest (such as the purple cluster), the community will be even more concentrated in this geogrpahic area. While we see that when the area is dense (rectangular area), this is still true but there are also nodes in the area that do not belong to the community.

Another interesting aspect is the two graphs of communities in the United States according to the number of clusters. Going from 4 to 5 clusters allows to mark the difference between the West Coast and the East Coast. 4 clusters were not enough to highlight this distinction.

Node scale

A node-scale analysis can be performed quantitatively from the results of question 2.

By analyzing a degree distribution curve as an inverse function, or a descending line, one can deduce that the graph has a hierarchical structure and that some nodes are much more connected than others. For example, in the case of a transportation network, hubs could be important airports that are connected to many destinations, while other nodes could be less important airports.

By considering the degree distribution, the graphs have a hierarchical structure, with a number of nodes having a very high degree (these nodes are often called “hubs”) and most of the other nodes having a relatively low degree (local airport).

It is interesting to note that an airport like London Heathrow (the largest airport in the UK in terms of passengers and global (international) connections) is not considered a hub on the graph. If this seemed surprising to me at first, it is actually quite logical because London Heathrow is an international airport and we only work on domestic flights. Thus, we can underline that LHR is not the airport with the most connections when it comes to UK domestic flights. If we had taken into account the world traffic, the assortativity should be higher because the hubs are connected to each other, we can mention for example New York (JFK) and London Heathrow (LHR).

By plotting the degree vs. betweenness centrality distribution on a graph, one can visualize how these two measures are related in the graph and how they can be used to understand the structure of the graph and the relationships between the nodes.

For example, if the distribution of degree and midpoint correlates strongly (i.e., nodes with high degree also have high midpoint), this may indicate that nodes with high degree are also heavily involved in many paths between other nodes in the graph. In the case of air traffic, this could mean that some airports are highly connected and play an important role in the flow of flights between other airports. This is the case for each of the networks.

By analyzing the assortativity of a graph, we can infer information about the structure of the graph and the relationships between the nodes. For example, in the case of air traffic, a positive assortativity could mean that airports with a large number of flights tend to be related to each other, while a negative assortativity could indicate that airports with a large number of flights tend to be related to airports with a smaller number of flights.
In our graph examples, negative assortativity could mean that airports with similar characteristics (e.g., large size or high activity) are less likely to connect to each other by flights than airports with different characteristics. This could be the result of different marketing strategies or airline development policies, or environmental or geographic factors. For example, the widely adopted aerial strategy of the hub and spoke organization is probably the most likely explanation, The hubs connecting mainly with regional airports. This makes some sense in our study, where only domestic flights are considered. If you look at the graphs, you can often see that there is one hub that serves many smaller regional airports.

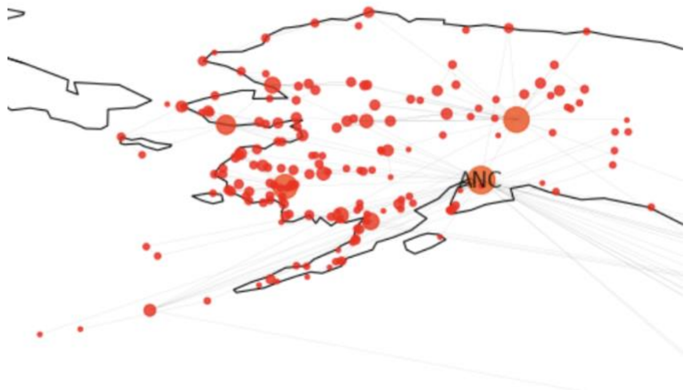

Figure 24: the hub and spoke organization illustrated with Alaska

Thus we see that the Anchorage airport can be seen as an alaska hub that serves the local alaska airports.

It is notable in this example speaking of Alaska, the flights are mostly short distances, small aircraft may be more suitable due to their ability to land at small airports and their lower operating cost.

Exploration on the fuel increasing prices ans its consequences

Lets seehow will be changing the fuel price impact the distance penalty and the formation of random graphs that mimic the airline network?

Rising fuel prices may result in higher aircraft operating costs, which may result in an increase in the distance penalty (i.e., long-haul transportation costs may become higher). This could impact the decision of some airlines to fly long distances, which could lead to a change in the structure of the air transportation network.
This could lead to a change in the structure of the air transport network, as companies would serve nearby airports, thus increasing the formation of distinct clusters on a macroscopic scale. On a more local scale, the larger nodes will be more impacted and may have fewer connections. For more local nodes, it depends on the distance between a hub and the local node. If they are far from a hub, then they might get cut off from the graph.

Figure 25: Illustration of the connections between the state of Hawai and the continental United States

For example, if transportation costs were to rise and the airlines were to discontinue service between Hawai'I and the West Coast airports, then Hawai'I's airports would be isolated from the rest, they could sleep a sub-network among themselves, isolated from the main network. The West Coast airports would be hubs with a lower degree of service like San Francisco (SFO) or Los Angeles (LAX).
It is important to note that these impacts will depend on the sensitivity of passenger demand to changes in fuel prices and the ability of airlines to pass these costs on to passengers.

On the graphs of Australia, China and the UK , the long-distance routes often link the hubs together. For these graphs, we can think that the increase in fuel costs could eliminate the connections between hubs. These hubs like Perth (PER) or Melbourne (MEL) would become regional hubs only. In this case, assortativity could be even more negative than previously found, as airports with a large number of connections will tend to be linked to airports with a smaller number of flights.

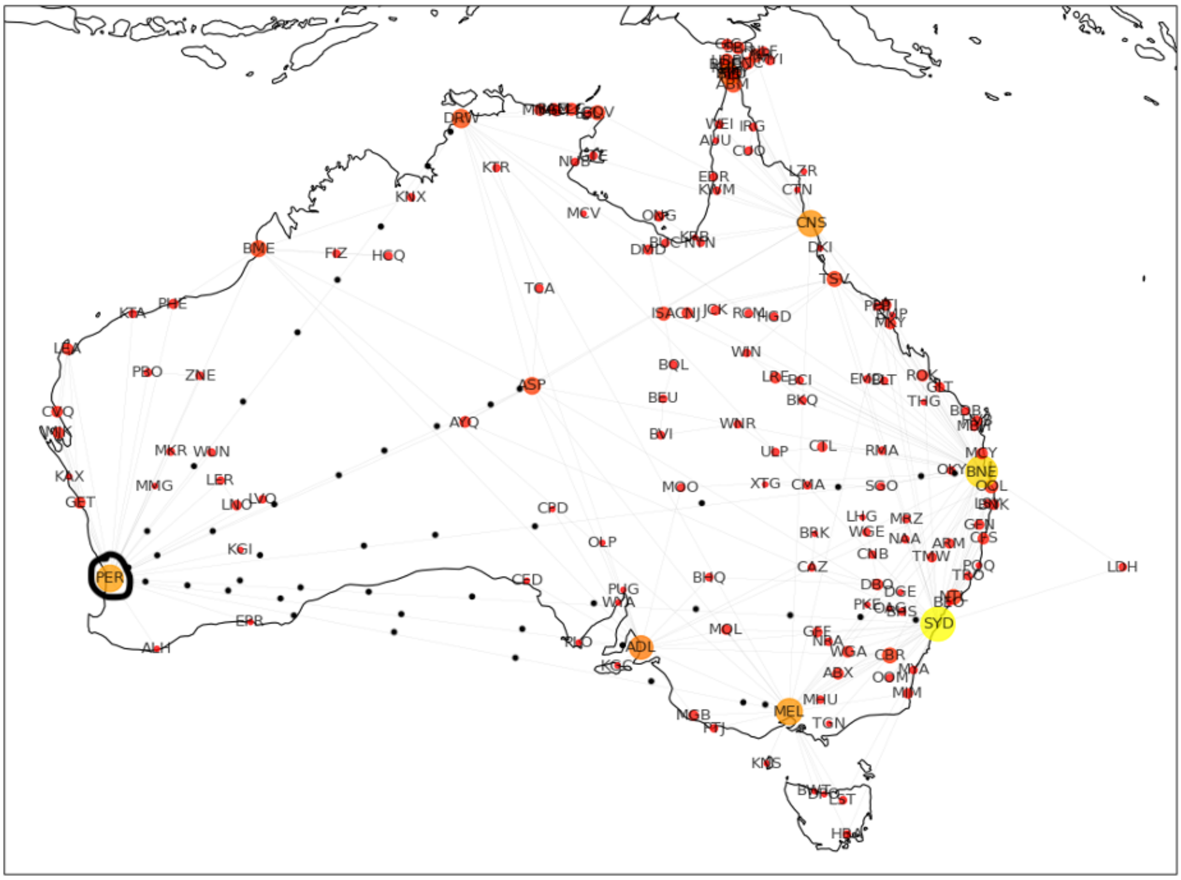

Figure 26: Illustration of the long distance connections of Perth airport.

The Perth airport is a good example of what has been said above. Its main long-distance connections are with other hubs. It is these connections that would be most impacted by an increase in fuel costs. If a clustering algorithm were applied, the cluster (green) that includes Perth would be even more focused on Perth and the smaller airports in the region.

Conclusion

In this study, we analyzed air traffic in graph form to better understand the structure and dynamics of the air transportation network. To do so, we used several graph analysis methods, such as node centrality analysis and node community analysis.

Our results show that the air transportation network is highly connected, with a large number of nodes having a high degree of centrality. We also identify several node communities, which are clustered around important air transportation hubs.

In conclusion, the graphical analysis of air traffic has allowed us to better understand the structure and dynamics of the air transportation network. These results could be useful for airlines wishing to optimize their transport network and for air transport regulators wishing to better understand the issues of this sector.

References

[1] Chen, L., Wang, L., & Zhang, Y. (2018). Empirical analysis of airline network evolution and its impacts on regional connectivity. Transportation Research Part E: Logistics and Transportation Review, 116, 82-101.

[2] Gao, L., White, C. C., & Tsai, J. J. P. (2011). A dynamic network analysis of the global airline industry. Transportation Research Part A: Policy and Practice, 45(1), 49-62.

[3] Erlander, S. G., & Ni, B. J. (1999). A structural model of air transport demand. Transportation Research Part B : Methodological, 33(8), 643-658.

[4] Weisi Guo, Bogdan Toader, Roxana Feier, Guillem Mosquera, Fabian Ying, Se-Wook Oh, Matthew Price-Williams, Armin Krupp (2019). Global Air Transport Complex Network: Multi-Scale Analysis. https://scholar.google.co.uk/citations?view_op=view_citation&hl=en&user=lhC7GvEAAAAJ&cstart=20&pagesize=80&citation_for_view=lhC7GvEAAAAJ:kvJssbFybhEC