EE-558, A Network Tour of Data Science
January 18, 2019

Flight network and airline alliances

Project Report on the "Flight Routes" Dataset

Group 6: Gabor Csordas, Nicolas Fontbonne, Maëlle Le Clainche, Marie Sadler

1 Introduction

We analyze the *Airline Route Mapper Route Database* which maps 3'321 airports (nodes) worldwide with their respective connections (67'663 connections corresponding to 19'256 edges) operated by 548 airlines [1]. Some of these airlines have joined each other to form passenger airline alliances which for instance facilitates flight connections for multi- stop flights. As of today three major airline alliances exist, **Star Alliance** (SA), **SkyTeam** (ST), and **One World** (OW) which regroup 26, 20 and 14 airlines, respectively (state of 2014 to be consistent with the flight dataset) [2].

The aim of our project is to unravel the airline alliances network and their global presence in the air route network. We try to understand to what extent the whole flight network is dominated by the alliances, if there are geographic differences in their distribution, and whether there is an underlying network property that allows to identify the major alliances from the available data without prior knowledge about their existence.

2 How prevalent are the alliances in the global flight route network?

2.1 Airlines distribution

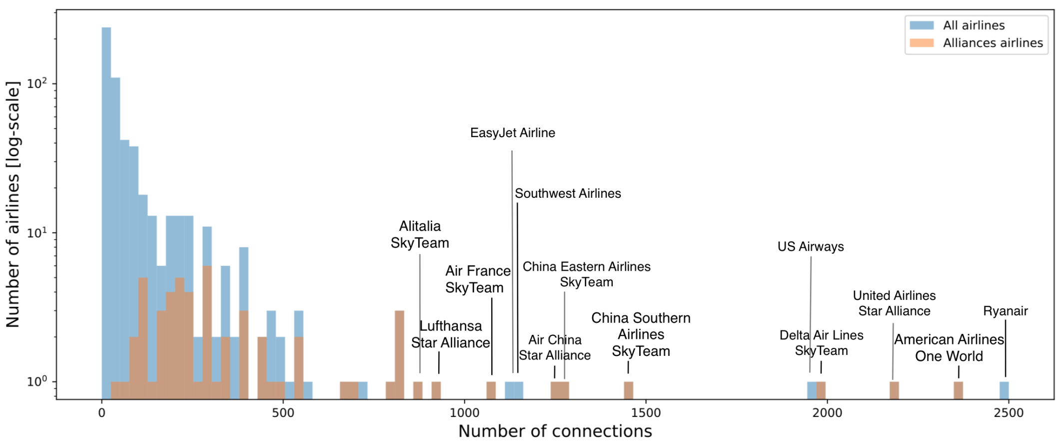

Figure 1: Number of connections operated by each of the airlines as indicated in the original dataset.

The original dataset *Airline Route Mapper Route Database* provides the airport connections and indicates by which airlines they are operated. This number should not be confused with the number of flights that are operated per day or per year, and which might give a different picture of how "big" an airline is in terms of passenger transport. While the number of connections when taking into account the different airlines is very large, the number of actual airport connections is much smaller (67'663 compared to 19'256) and in the following we will refer to the first case as the number of connections (weighted graph) and the second as the number of edges (unweighted graph).

In Figure 1 we see that most of the airlines offer very few flight connections, while a small number of airlines offer a very large number of connections. Even though the alliance's airline only represent 11% of all the airlines they operate 28% of the connections which is reflected by them being rather in the tail of the distribution.

The World Air Transport Statistics (WATS) records, among other statistics, the "Scheduled Passengers Carried" and in 2014 the top 5 airlines from highest to lowest were as follows: Delta Air Lines, Southwest Airlines, China Southern Airlines, United Airlines, American Airlines [3]. We notice that this statistics, which is more reflective of the actual number of flights, is slightly different from the ranking when only considering the number of connections.

2.2 Network properties of the alliance's flight routes

From our original graph, we create subgraphs for each of the 3 alliances as well as a subgraph for the three combined. In Table 1, the main network properties are listed and compared to the original graph. Each subgraph is plotted on the world map in Figure 5 shown in the Appendix.

Table 1: Network properties

	Whole flight network	**Alliance Network**	**Star Alliance**	**SkyTeam**	**One World**
Number of airports (nodes)	3'425	1'797	1'221	1'081	953
Weighted graph					
Number of connections	67'663	29'349	11'267	10'794	7'288
Average degree $< k >$	39.51	32.66	18.46	19.97	15.29
Maximum degree	1826	1405	587	1319	646
Unweighted graph					
Number of edges	19'256	9'656	4'385	4'006	3'063
Average degree $< k >$	11.24	10.75	7.18	7.41	6.43
Maximum degree	248	228	226	210	184
Diameter of the graph	13	9	9	9	9
Average clustering coefficient	0.49	0.53	0.47	0.49	0.44

We see that SA is the most prevalent among the three alliances. It serves the largest number of airports and has the most connections and edges. SA is followed by ST which serves a similar number of airports and has a similar number of connections and edges. However, its network is more dense, shown by a slightly higher average degree (7.41 vs 7.18). OW is the smallest alliance with 7'288 connections and its network is more sparsely connected and it has the lowest average clustering coefficient (OW: 0.44, SA: 0.47, ST: 0.49). The degree distribution in the global flight network follows a power law, and the network is characterized by the presence of a few very large hubs and a lot of nodes with only a few links. This behaviour is even more pronounced when we analyze the alliance-network alone. As expected, the alliance network is more concentrated and it forms a single component, whereas the whole network has 8 components. In addition, the diameter of the alliance network equals 9 which shows that the alliance network is less spread out than the whole flight route network which has a diameter of 13.

2.3 Alliances and big hubs

As already seen in the previous analysis, Figure 2 shows again that most of the airports are not served by an airline belonging to one of the alliances. The majority of airports in the alliance network are only served by one alliance (852 airports in total). However, a lot of airports are also served by all the three alliances, and interestingly this number is higher than the number of airports that are served by 2 alliances (432 compared to 513 airports). This distribution might hint the fact that big hubs are likely to be served by all the alliances, whereas small airports are only served by one alliance. This assumption is confirmed in Figure 3. When only considering the alliance's subgraph, there is a trend that airports with a larger number of connections are more likely to be served by the 3 alliances. However, there are also sparsely connected airports that are served by all the alliances. On the right plot in Figure 3, the degree is calculated on the whole network. In this case there is no increasing relation, but it rather seems that there is a threshold at a degree of around 100. Airports exceeding this number are served by all the

3 alliances (except for Stansted which is only served by OW and which owes this high degree to Ryanair which is not part of the alliances).

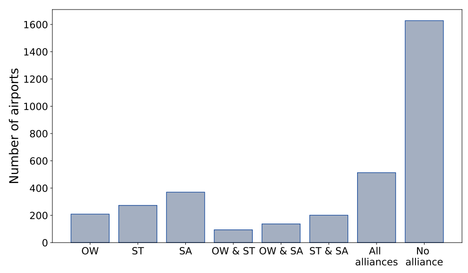

Figure 2: Number of airports served by the different alliances

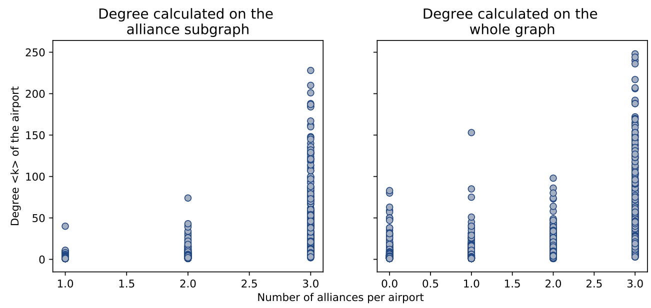

Figure 3: Degree of the airport (unweighted graph) plotted against the number of alliance's airlines serving it.

3 Which part of the world is dominated by who?

3.1 Geographic distribution of the alliances

Each airline has a country of origin and thus we can attribute the alliances to different countries. In addition, for each airport we can determine the major alliance serving it by comparing the node degree of each alliance. Figure 6 in the Appendix shows that each alliance is quite spread over the world. This seems to be a consequence of the fact that they want to provide the best experience to their client (the client can go almost anywhere they want with one alliance and thus benefit from a good deal). However, more importantly is that each state maintains exclusive sovereignty over the airspace above its territory as stated by historical conventions: This means that if an airline wants to land in a foreign country, there has to bilateral negotiations between the two countries, and more precisely between the governments and not the airlines themselves [4]. To overcome these negotiation hurdles, the SA formed in 1997, leading to the creation of OW and ST in 1999 and 2000, respectively by their competing airlines. The USA is an unavoidable country for each alliance: it is a big country with a lot of customers and hence strategically important for each alliance. There is no country with the 3 alliances in Europe, which is probably a consequence of the *European Common Aviation Area* (ECAA) which allows any company from any ECAA member state to fly between any ECAA member states airports.

When we compare the distribution of the airport and country attribution, we see that they match very well. In Table 2, it becomes clear that the countries where an alliance's airlines have their base have a high probability to be countries where the alliance dominates some airports. This seems logical, because if an airline has its base in a country, it will probably own a lot of the flights in this country. On the other hand, the alliances are present in a lot of countries where their airlines do not have their base, which of course is part of their expansion strategy.

Table 2: Comparison between the country and airport distribution

Number of countries where:	Star Alliance	One World	SkyTeam
Alliance's airlines have their base	25	15	15
Alliance dominates at least one airport	145	94	81
Country base & Airport domination	23	12	15

3.2 Airport expansion of the alliances predicted through label propagation

We developed an airport label propagation algorithm whose underlying concept is that airport not yet served by an alliance will be served by the alliance occurring with the highest frequency among its neighbours (see Figure 4). Each neighbour has one vote, independent of how connected it is, and an airport may be labelled by 2 or 3 alliances if there is a tie among the node degree between the alliances (at the step of the initial labelling) or a tie between the vote comparison (at the step of attributing the label).

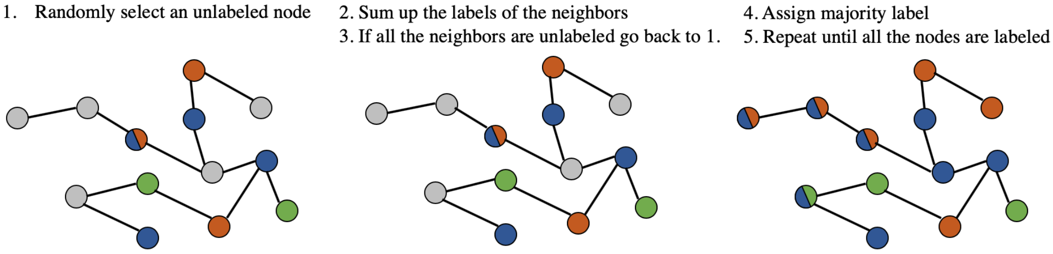

Figure 4: Scheme of the label propagation algorithm.

In a first phase, we validate our algorithm by only working on the alliance network and by randomly unlabelling half of the nodes. When we ran the algorithm, 883 out of 1797 were unlabelled airport and we got high accuracy scores. Since we allow the airports to have multiple labels, we calculate a score once for every alliance and an overall accuracy score. Here is an example to illustrate this: if an airport has a SA and a ST label, and our algorithm predicts only SA, their would be a full accuracy score for SA alone, but 0 for ST and the overall correctness. We obtain the following scores: SA: 81%, OW: 84%, ST: 85% and overall accuracy: 71%.

Next we proceed to label the 1600 airports not served by an alliance and that are in our largest connected component (globally only 28 airports are not part of it). Some geographic differences between the alliances become even more apparent when looking at the expansion map (Figure 7 in the Appendix): OW is more present in Australia and the West Coast of South America, ST in Indonesia and North Africa and SA on the East Coast of South America and Northern Europe.

Having predicted the alliance label for each airport, we can even go one step further and assign alliances labels to airlines that are not part of the alliances. We do this by selecting the alliance with the highest frequency when counting together all the airports that the airline is serving. We tested this approach on the two airlines that joined the alliances after 2014 and that we did not include since our dataset is from 2014. From our algorithm both the airlines (Air India and Oceanair/Avianca Brazil) are predicted to join SA which indeed is who they joined.

4 Can the alliances be predicted from an airline network?

4.1 Building an airline network

In the previous chapters we used the alliances as ground truth. However, these alliances were founded in order to make the journey of intercontinental travellers more comfortable by providing connections at hubs, therefore the airline alliances must appear somehow in the network of flight routes.

Our idea is to construct the network of airlines based on specific properties of the flight routes network and determining the alliances with spectral clustering on this new graph where each node is an airline and the edges represent how strongly they are connected to each other. We use a very simple hypothesis to construct this new graph. We assume that an airline will offer more transfer options to the flights of those airlines that are in the same alliance than to those that are in another alliance. So we checked all the pair of flights where the destination airport of the first one is the same as the source airport of the second one. If the first one and the second one are maintained by two different airlines we increased the weight of their edge in the adjacency matrix of airlines by one.

The problem with our hypothesis is that the airlines that operate in the same region (Europe, North America, South America, Africa, Asia) are connected strongly to each other even if they are in different alliances. Therefore we improved our adjacency matrix by checking for all the airlines the region they operate in and setting the weight of the edges between two airlines operating in the same region to zero.

At this point we discovered that the result of the spectral clustering does not make any sense since huge airlines like American Airlines (2354 connections) and Delta (1981 connections) have much more flights than smaller but important airlines in alliances like Quantas (432 connections). Therefore the weight of edges between large airlines is larger than between smaller airlines regardless of the alliances. This gave us the idea to do some normalisation on the edges. We divided the weights by the degree of the node (sum of the weights in the row), so the adjacency matrix became asymmetric. In order to make it symmetric again we took the maximum of weight from node i to j and from j to i, thus the relative number of connections for the smaller airlines dominates the adjacency matrix.

4.2 Cluster analysis

Firstly, we analysed just the network of those airlines that are member of the alliances.

	1. cluster	2. cluster	3. cluster
One World	3 (21 % of OW)	11 (79 % of OW)	0 (0 % of OW)
Skyteam	7 (35 % of ST)	2 (10 % of ST)	11 (55 % of ST)
Star Alliance	14 (54 % of SA)	11 (42 % of SA)	1 (4 % of SA)

Table 3: The number indicates how many airlines of a given alliance are inside the cluster.

In spite of the simple hypothesis the clusters represent the alliances quite well (1. Star Alliance, 2. One World, 3. Skyteam). On the other hand we can see that the 2. cluster contains 42% of Star Alliance airlines and more importantly it has United Airlines which is the one the most important airlines of the Star Alliance. Actually the other Star Alliance airlines were clustered here because of United Airlines. The reason is that there are some South American (Aerolineas Argentinas, Copa Airlines, Avianca - Aerovias Nacionales de Colombia) and Asian airlines (All Nippon Airways, Air New Zealand, Cathay Pacific) that are almost equally connected to American Airlines (One World) and United Airlines (Star Alliance).

As the previous label propagation analysis, spectral clustering also allows to predict attributions of airlines to alliances. We did the experiment on 10 major airlines not part of the alliances with the two approaches and for 8 of them we got matching results (see details in the Table 4 in the Appendix). Since both analyses make use of the assumption that expansions to neighbouring airports are the most likely, or in other words, neighbouring connections are most likely to be inside a cluster, this result could be anticipated.

5 Conclusion

The analysis showed that the alliances are globally very present in the flight network: half the airports and half the connecting edges are served and operated by an alliance airline, even though the alliance's airlines only represent 11% of all the airlines. Big hubs are indispensable for them and airports exceeding a degree of 100 are served by all the alliances (except for the airport Stansted in London). The airlines are spread more or less equally over the whole world, once to provide the best offer to the customers and thus have increased sales figures, but also to overcome negotiation hurdles due to the airspace regulations which is done through so-called "code-sharing" of different airlines inside an alliance.

However, there are also geographic differences which arise from the different airlines having different countries of origins. Through label propagation, we predicted the expansion of the alliances which enhanced these differences: SA is more likely to expand in Northern Europe and the East Coast of South America, OW in Australia and the West Coast of South America and ST in Indonesia and North Africa.

Raising the hypothesis that airlines inside alliances offer more transfer options from a particular airport to other airports than airlines that are in another alliance, allowed to construct an airline network from which, through spectral clustering, we could identify the alliance's clusters quite well.

In a last part, we predicted which alliances major airlines are likely to join, once through the spectral clustering and once through the label propagation approach and very coherent results were obtained.

It remains to see, whether the predictions will be confirmed in the future or if on the other hand the alliance business will experience a decline.

References

[1] Airport, airline and route data. `https://openflights.org/data.html`. Accessed: 2018-09-19.

[2] Alliance data. `https://www.hopper.com/articles/860/a-guide-to-the-three-major-airline-alliances-star-alliance-oneworld-and-sky-team`. Accessed: 2018-12-11.

[3] World Air Transport Statistics (WATS). Scheduled passengers carried, 69th edition. `https://web.archive.org/web/20150618070729/http://www.iata.org/publications/pages/wats-passenger-carried.aspx`. Accessed: 2019-1-6.

[4] Robert F Barron. Code-share agreements: a developing trend in us bilateral aviation negotiations. *Ind. LJ*, 72:529, 1996.

A Network of the three alliances

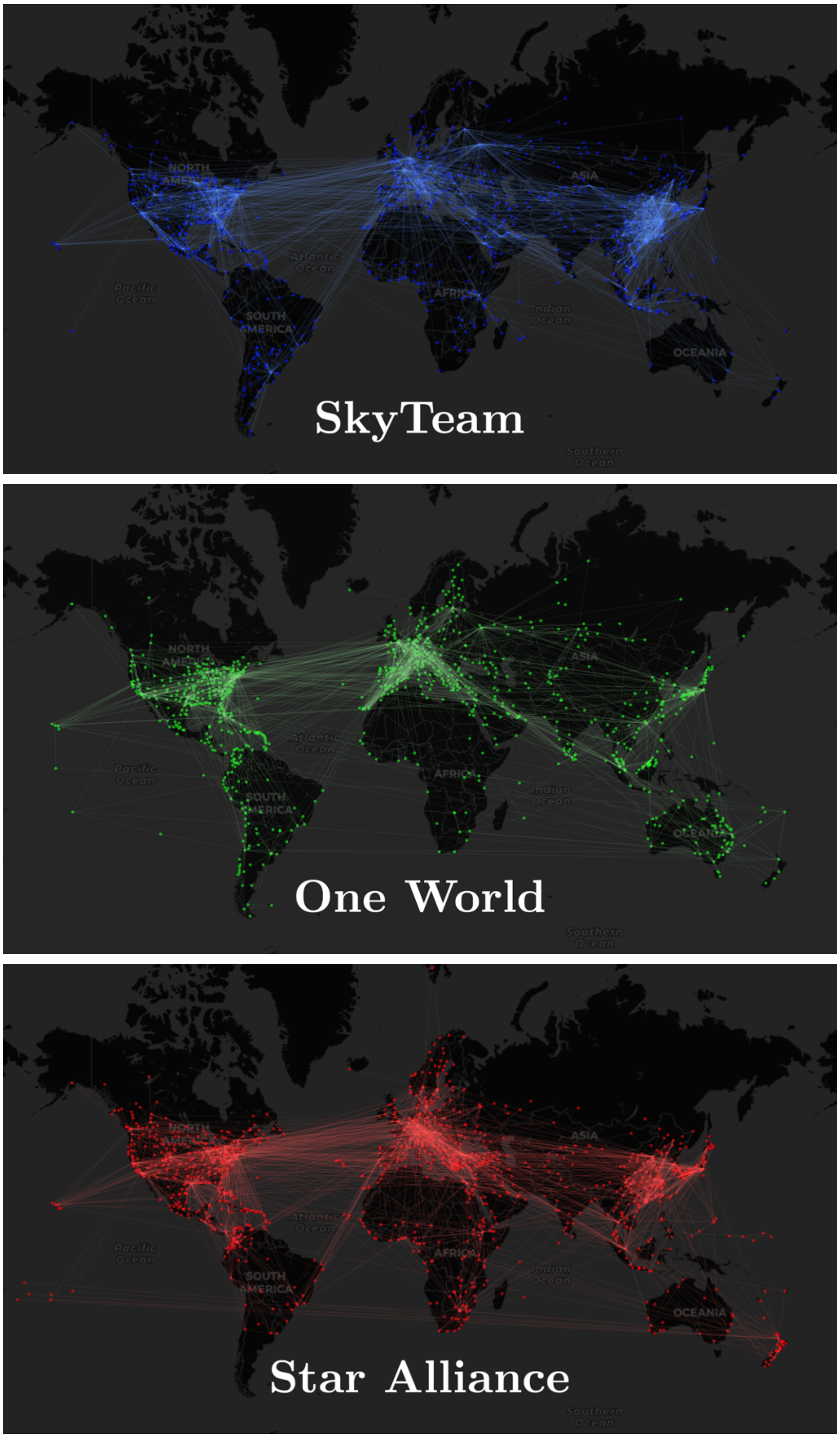

Figure 5: Network of the three alliances. Each dot represents an airport where at least one flight is operated by the alliance studied, and the edges are the flights connecting the airports.

B Alliance attribution's map

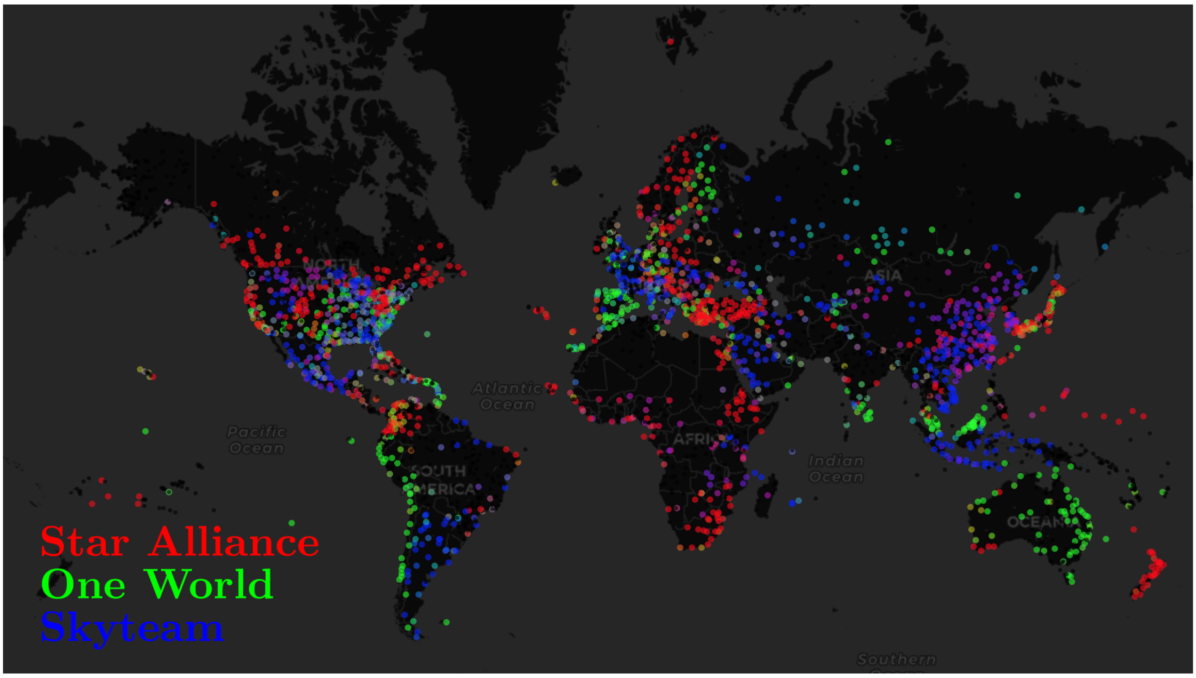

Figure 6: Alliance attribution's map. We associate each alliance to a color (Star Alliance: red, One World: green, SkyTeam: blue. Then the value of these channels is proportional to the number of flights from and of to these airports. This means that a reddish airport has a lot more flights from the Star Alliance than the other alliances.

C Network expansion's maps

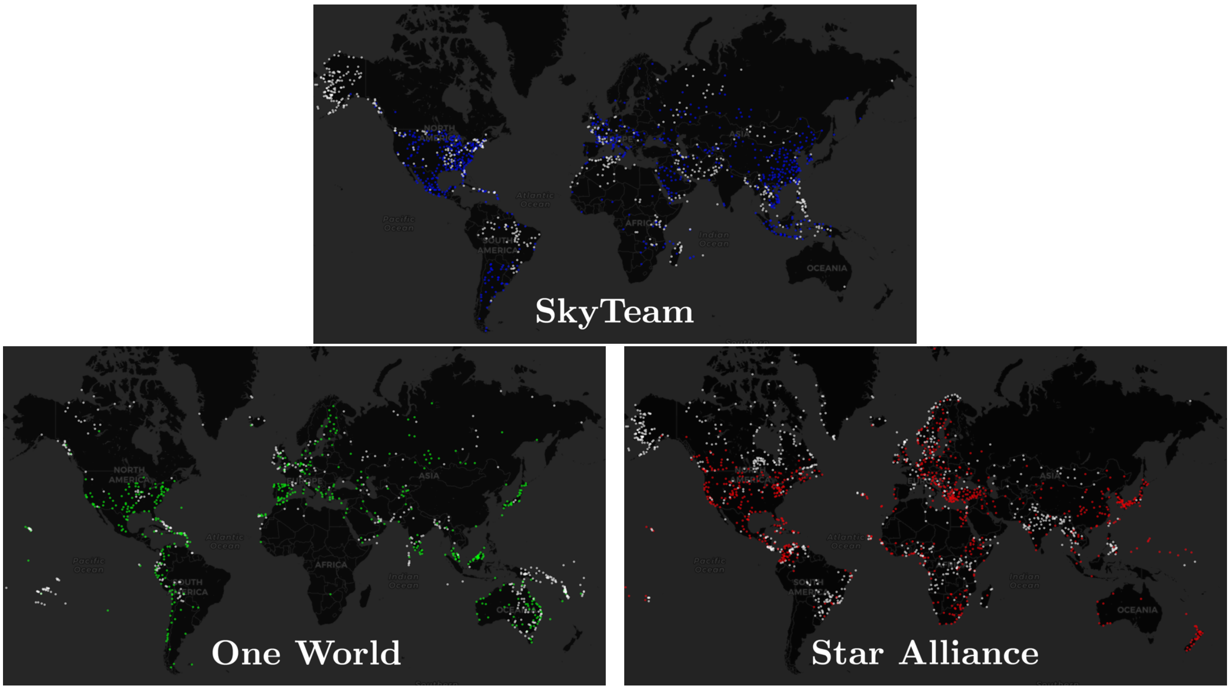

Figure 7: Network expansion's maps: the coloured dots are the airports where the alliance currently dominates, and the white dots are the airports predicted to be dominated by the alliance.

D Prediction of airline attributions

Table 4: Alliance's attribution of major airlines predicted through spectral clustering and label propagation.

Airline (IATA)	Spectral clustering	Label Propagation
JetBlue Airways (B6)	OW	ST
Ryanair (FR)	OW	OW
US Airways (US)	OW	ST
Emirates (EK)	SA	SA
Etihad Airways (EY)	SA	SA
easyJet (U2)	SA	SA
Wizz Air (W6)	SA	SA
AirAsia (AK)	ST	ST
Virgin Atlantic Airways (VS)	ST	ST
Southwest Airlines (WN)	ST	ST