Airline clustering using complex network metrics and operational features

Victor Ciulei, Mihaela Mitici[2]

Faculty of Aerospace Engineering, Delft University of Technology, Kluyverweg 1, Delft, 2629 HS, Netherlands

Salvatore Miccichè[3]

Department of Physics and Chemistry - Emilio Segrè, Università degli Studi di Palermo, Viale delle Scienze Ed 18 90128, Palermo, Italy

Abstract

Airlines have evolved and adapted since their nascent times, through the deregularization period until todays technology-driven modern times. However, the categorization of airlines has remained unchanged and mainly refers to features, such as the business models, that are nowadays less pertinent than before. This paper proposes an unsupervised clustering methodology, based on a balanced mix of operational features and complex network metrics, that is quite able to detect in a robust way clusters of airlines. Moreover, the methodology is sufficiently powerful as to keep trace of different partitions of airlines due to the fact that different market conditions (such as summer timetable versus winter timetable) trigger different operational settings. The methodology is based on an ensemble of 12 individual algorithms whose results are then aggregated as to form a Consensus Index that in our case is based on a Hierarchical Agglomerative Clustering with average linkage. We analyze a set of 94 airlines in Europe through the entire year 2017 by using DDR2 data and open source data. The features that feed our algorithm concern information such as airline's fleet, number of operated flights, average off-block time, as well as a few metrics useful to describe the airport network associated to each airline. As a result we cluster airlines into four groups. We find that two of these groups are composed of airlines that are quite homogeneous from the point of view of the categorizations commonly used in the literature: Flag carriers and Low-Cost carriers. However, we also find two more clusters whose composition is quite heterogeneous when considering that categorization. We believe that our methodology well captures the way airlines group together when adapting themselves to the continuously changing needs of today's aviation market.

Keywords: airlines, clustering, Machine learning, Complex Networks
2010 MSC: ??, ??

February 10, 2021

[1]MSc. graduate, vciulei@gmail.com
[2]Associate Professor, Department of Air Transport Operations, m.a.mitici@tudelft.nl
[3]Associate Professor of Applied Physics, salvatore.micciche@unipa.it

1. Introduction

V:rename to full service airline, low cost airline and renounce carrier throughout the paper. keep only airline Current mainstream labels defining airline business models (flag carrier, network carrier, low-cost carrier, etc.) have become an ever-present part of modern air transport vocabulary and thus influence the industry and passenger perception. But these names were coined decades ago and may no longer be representative to the airline business models nowadays. As the economy changes, the markets shift and the airline sector does so, too: 78 out of the 110 low-cost airlines started between 1995 and 2011, went bankrupt, leaving the others to adapt in order to survive [1].

Such fast transformation is a natural follow-up of the market deregularization that started in the 1970s in US and 1993 in EU, when bilateral agreements started to be replaced with more permissive Freedoms of Air [1]. Both new and established carriers took advantage of these changes, with newer airlines being more flexible to adapt and focus easily on previously untapped market segments. One of the most famous example is that of Southwest Airlines [2]. It revolutionized air transportation by making flying affordable for a larger volume of passengers, mainly by homogenizing its fleet, charging for luggage and flying to smaller, cheaper airports where short turnarounds were possible.

Many airlines followed suit, some adapting the new business models to focus on a different niche. Previously established carriers (flag carriers) also had to adapt in order to retain market share [1]. Because they had legacy routes to defend and had to adapt to being de-nationalised, as well as dealing with a heterogeneous fleet, their transformation was slower but nevertheless constant. Another type of market entrant were the leisure airlines, that catered mostly to seasonal passengers going to holiday destinations. They are in some aspect similar to low-cost airlines because of their choices of fleet and pricing strategies, but also different due to their choice of destinations.

Therefore, an inexorable action-reaction cycle pushes the airlines to constantly adapt, improve and re-orient to maintain profitability. Another layer that is added on the classification is that of operational changes due to mergers, acquisitions or the different methods of leasing between airlines.

In this paper, 10 airline business model features are used to differentiate between airline business model types: operational, fleet and Complex Network features. The features feed the Unsupervised Machine Learning Clustering Ensemble which in turn generates a final clustering. Clustering ensembles are essentially a suite of different clustering algorithms or of the same algorithms with different parameters. They have evolved to be a better alternative to individual clustering algorithms, as the individual results are aggregated together in a robust, representative outcome that minimizes the biases that each algorithm is usually prone to [4]. The results obtained present the airline groups and analyze the contents of the groups in terms of business model (low-cost, flag carrier, leisure, regional) and size (large, medium, small). Furthermore, groups and subgroups features are compared and analyzed.

The main contribution of this paper lies in the unsupervised clustering of individual airlines, challenging and reinforcing the status-quo on airline business models and their meaning. In addition, it defines a clear methodology to cluster airlines and paves the way for identifying a complete set of airline features, that are used to classify airlines. Furthermore, a clear distinction in two of the four groups between large low-cost and large flag carriers has been obtained, making the proposed machine learning clustering methodology the first one that objectively differentiates between the two types.

[4] https://hdbscan.readthedocs.io/en/latest/comparing_clustering_algorithms.html

The remainder of this paper is structured as follows: Section 2 provides background on airlines, complex networks and machine learning, Section 3 presents the data used in the paper, Section 4 details the clustering methodology, Section 5 presents the results and a discussion, and Section 6 concludes the paper.

2. Related work and contributions

The Unsupervised Machine Learning approach has become a staple in modern data analysis, predictions and classification. It is used in many applications, such as classifying unstructured text and images [3]; classifying objects [4] or identifying cars from a moving target [5]. Although its applications are wide, this approach has not been used to classify airlines, most likely because of the absence of a clear machine learning methodology and a comprehensive set of quantitative parameters.

The need to accurately portray airlines has been recognized by several researchers in modern times. Perhaps one of the most comprehensive and coherent description of airline business model aspects can be found in the work of Belobaba et al. [6], that details airline specifics to the level of fleet components, route network development strategy, in-flight services and revenue management. It has contributed considerably to the understanding of airline operations but explicitly compares only two airline types: low cost and network carriers. Moreover, no extensive classification of a large set of airlines is present.

One of the first studies to compare and contrast airlines applies the POA (Product and Organizational Architecture) methodology on 10 numerical indices developed from a wide variety of airline data for 6 traditionally-known low-cost airlines [7]. These indices reflect aspects related to airline revenue, ticketing, network topology, seats, fleet, airports served or aircraft utilization, among others. Airlines are compared by an index benchmarking system with the 'top performer', resulting in a comparative analysis of airline business model aspects. This study provides a good basis for airline comparison and proposes a number of relevant indices, however it does not assign airlines to specific groups based on the indices as the sample size is rather low. Another similar study groups 43 low-cost airlines in 5 categories: Southwest copy-cats (based on [8]), subsidiaries, cost cutters, diversified charter carriers and state subsidised competing on price. Although entirely categorical and based on observation and expert judgement, it offers a sensible view on the state-of-the-art situation of airlines at the beginning of the 21st century and highlights how the low cost *model* is a better term to describe the variations in aspects of low cost carriers.

Another study [9] adapts the idea developed in [7] to emphasize the hybridization of 20 large LCCs in Europe. It uses a simple index based on the number of 'fulfilled' LCC criteria, such as having a low number of aircraft types, use of secondary airports, single-cabin aircraft, mostly point-to-point flights and no codesharing agreements. Based on the developed index, airlines are grouped into 4 distinct pools: pure LCC, hybrid carrier with dominating LCC characteristics, hybrid carrier with dominating full service carrier characteristics and full service airlines. The results underline the tendency of European LCCs to counter-react the cost-cutting of full service carriers and that airline hybridization is an ongoing phenomenon. A similar conclusion is presented by another empirical study [10], which implements a previously developed conceptual framework [11] whose goal is to systematize the comparison of airlines. The methodology is comprehensive and uses mainly categorical data types from three main business aspects: corporate logic (market target, geographical area), configuration of value chain activities (supply management, route network, fare structure) and assets (fleet, infrastructure and human capital). The methodology easily tracks business model convergence over time using a combination of categorical and quantitative data. One aspect of the methodology [11] that could be improved is that all metrics are compared on a 5-step scale, even though some of the metrics are binary. Finally, another study classifies airlines based on qualitative aspects by applying a new method called business canvas to airline operations [12]. The study includes 42 airlines based in multiple continents and uncovers 7 clusters...

This study classifies 94 European airlines in 4 groups by considering their fleet, operational, complex network (CN) metrics. A complex network is a network that has non-trivial topological characteristics, i.e. do not occur in random graphs. These can be, among others, specific degree, clustering and betweenness distributions, and the presence of communities or small-world properties.

The CN paradigm has been widely used to model real-world systems [13], such as the internet [14] or terrestrial transportation systems [15][16], the spread of infectious diseases [17], or social networks [**?**]. In ATM it is particularly useful as it sheds light on passenger mobility and reflects the characteristics typical to the airline business model [**?**] or market segment [18]. Many scientific articles have also characterized different networks within the ATM system: the network of navigation points and air routes [19] or the networks of airports [20][21][22][23]. For an overview on the Complex Network metrics in ATM, the reader is invited to consult [24].

3. Description of the data

.

The dataset used to conduct the analysis presented in this paper includes all the commercial flights executed in 2017 for which at least one airport (origin or destination) is located in the ECAC (European Civil Aviation Conference) member states [5]. This data has been obtained from Demand Data Repository (DDR2)[6]. A total of 6.675.475 flights are considered, corresponding to the year of 2017. More precisely, only flights with either the Origin or Destination (or both) in the ECAC area are considered. All the 861 airports included in this paper are shown in Fig. 1.

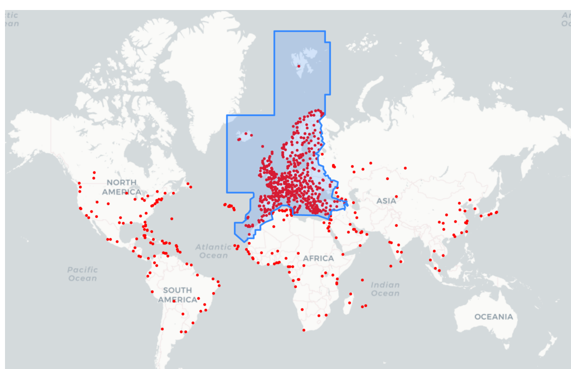


Figure 1: The airports considered in this paper (red dots) and the ECAC area.

A total of 94 airlines are considered. These airlines are registered in Europe, they complete at

[5]https://www.ecac-ceac.org/member-states
[6]https://www.eurocontrol.int/ddr

least 10 flights per month or have a fleet of at least 5 aircraft, and are not offering only charter services.

For each flight that an airline is performed, the following information is considered: actual arrival time, origin and destination airport, flight duration, aircraft type, AOBT (Actual Off-block Time) and operator. Among these flights, 7% of flights are identified as long-haul (more than 6 hours), 9.6% as medium-haul (less than 6 hours but longer than 3) and 83.4% are short-haul (less or equal to 3 hours). Theses flights are carried out with different types of aircraft. Using the RECAT-EU [7] categorisation scheme, 0.3% flights are carried out with an aircraft in the 'Super' category, 7.2% of the flights are carried out with an aircraft in the 'Upper Heavy' category, 1.3% in 'Lower Heavy', 64.1% in 'Upper Medium', 18.8% in Lower Medium and 0.9% in the 'Light' category.

Table 1 shows an example of the data format of any flight.

Origin	Destination	Aircraft ID	Operator	Aircraft type	Profile
LSZH	EDDC	SWR918	SWR	F100	Table 2

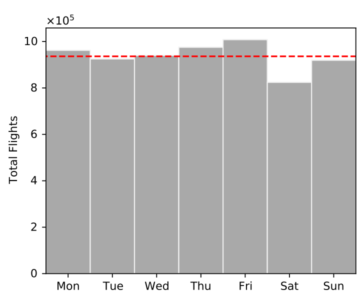
Table 1: DDR2 sample flight data fields used in the study

Each trajectory profile consists of multiple instances, recorded every few seconds, up to XX seconds. Table 2 below presents the fields of one such profile instance.

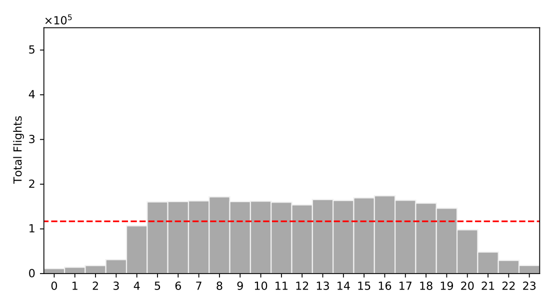
20170508164700	LSZH	DEGES2W	14	0	A	472729N0083253E
yyyymmddhhmmss	Navpoint	Sector	FL	Speed[kts]	Navpoint type	Latitude & Longitude

Table 2: Data fields of trajectory profile

Considering all 94 airlines, Fig. 2.b shows the total number of flights taking off every hour (the Actual Off-Block Time). Fig. 2.a presents the total number of flights taking off each day of the week. In both cases, the dotted red line represents the average. During the year 2017, on average, 18289 daily (relevant) flights were executed.



(2.a) Total flights during the days of a week



(2.b) Total hourly flights

[7] https://www.eurocontrol.int/sites/default/files/content/documents/sesar/recat-eu-released-september-2015.pdf

4. Airline clustering methodology

.

In this section a methodology to cluster airlines using unsupervised clustering algorithms is proposed. The methodology takes into account both network-based features, i.e., the characteristics of the networks that each airline is flying, as well as operation-based features that characterize the actual operations of these airlines. It involves various steps. Firstly, the data is normalized and

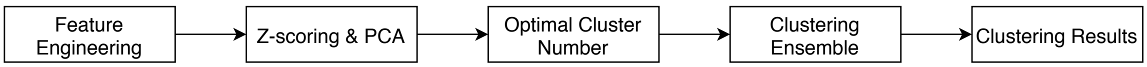


Figure 3: High-level overview of the clustering methodology

Principal Component Analysis is performed for dimensionality reduction. Next, an optimal number of clusters is determined. Having determined the best number of clusters, airlines are assigned to one of them using a Clustering Ensemble [25] method. Fig. 3 shows an overview of the steps followed for the clustering of airlines.

4.1. Feature Engineering

.

In this analysis, two types of features are considered: i) features based on the network that each airline is flying, and ii) features based on the operational characteristics of each airline.

i) Complex network features

Let A denote the set of airlines and let $a \in A$ be a specific airline in the set. The network of airline $a \in A$ is represented as an undirected, edge-weighted graph $G^a(V^a, E^a)$, where V^a is the set of all airports where a flight operated by airline a departs from/arrives at, and E^a is the set of links between origin-destination airports where airline a flies to and from. A link $e^a_{ij} \in E^a$ connects a vertex (airport) v^a_i with a vertex (airport) v^a_j. Let $w(e^a_{ij})$ denote the weight of the edge e^a_{ij}, i.e., the number of flights between airport i and airport j that airline a is flying. Based on this network, the following airline network features are considered:

1. The total number of airports that airline a departs from/arrives at, which is denoted by n_a, $n_a = |V^a|$.
2. The total number of origin-destination links that airline a is flying, which is denoted by $n_{a,l}$, $n_{a,l} = |E^a|$.
3. The average degree of the network of airline a, which is denoted by k_a, $k_a = \frac{n_{a,l}}{n_a}$.
4. The average clustering coefficient, which is denoted by C_a and defined as:

$$\bar{C}_a = \frac{1}{n_a} \frac{2|\{e^a_{jk} : v^a_j, v^a_k \in N^a_i, e^a_{jk} \in E^a\}|}{|N^a_i|(|N^a_i| - 1)},$$

where N^a_i is the neighborhood of vertex v^a_i,

$$N^a_i = \{v^a_j : e^a_{ij} \in E^a\}.$$

5. The average strength $\bar{S}_a$ of the network of airline a, which is determined as:

$$\bar{S}_a = \frac{1}{n_a} \sum_{i=1}^{n_a} \sum_{j=1}^{n_a} w(e^a_{ij}).$$

6. The average path length $\bar{L}_a$ of the network of airline a. Let $d(v_i^a, v_j^a)$ denote the shortest distance between vertices v_i^a and v_j^a. Then,

$$\bar{L}_a = \frac{1}{n_a(n_a-1)} \sum_{i \neq j} d(v_i^a, v_j^a), \;\; v_i^a, v_j^a \in V^a.$$

ii) Operation-based features

The following airline operation-based features are considered in our methodology:

1. The number of flights airline a completed between any airport v_i^a and $v_j^a, i \neq j$, which is denoted by $\bar{n}_{a,F}$, i.e.,

$$n_{a,F} = \sum_{e_{ij}^a \in E^a} w(e_{ij}^a).$$

2. The total number of aircraft types in the fleet of an airline a, which is denoted by $n_{a,K}$.
3. The fleet size of an airline a, which is denoted by $n_{a,AC}$.
4. The average daily aircraft utilization for an airline a, denoted by $\bar{U}_a$. Let $T_{c,a,d}$ denote the flight time, including taxi time, of an aircraft c of airline a which is performed during day d. Let $F_{c,a,d}$ denote the total number of flights aircraft c of airline a performs in day d. Then the average daily utilisation for aircraft c of airline a is defined as:

$$\bar{U}_{c,a,d} = \frac{T_{c,a,d}}{F_{c,a,d}}.$$

Then, average daily aircraft utilization for an airline a is:

$$\bar{U}_a = \frac{1}{n_{a,c}} \sum_{c=1}^{n_{a,AC}} \sum_{d=1}^{n_H} \frac{1}{n_{c,H}} \bar{U}_{c,a,d},$$

where $n_{c,H}$ is a number of days (period of time) that aircraft c has flown. An aircraft can fly a maximum of 365 days since the total considered period is the entire year of 2017.

Example of network-based and operations-based airline features

Below the network-based and operation-based features for four European airlines with different operational models and networks are illustrated. We consider the following aoirlines: KLM, Lufthansa, EasyJet and Ryanair. Table 3 shows the network and operations-based features of these four airlines in year 2017. EasyJet and Ryanair operate large point-to-point networks as it can be seen from the large number of links $n_{a,l}$ and the high average degree of the airlines' network k_a, while KLM and Lufthansa operate a network with one and two airport hubs, respectively. The strength S_a for KLM and Lufthansa is much larger than for Ryanair and EasyJet, which shows that on average, KLM and Lufthansa have many more flights per link than the other two airlines. The average path length $\bar{L}_a$ is similar for all four airlines. This means that, despite the differences in the network topology of these airlines, a destination airport can be reached from an origin airport with 2 flight legs on average. Lastly, KLM has a much lower average clustering coefficient $\bar{C}_a$, which implies that KLM has a network with one main airport hub (one vertex with a high degree), while the other three airlines have networks with several, communicating vertices of high degree.

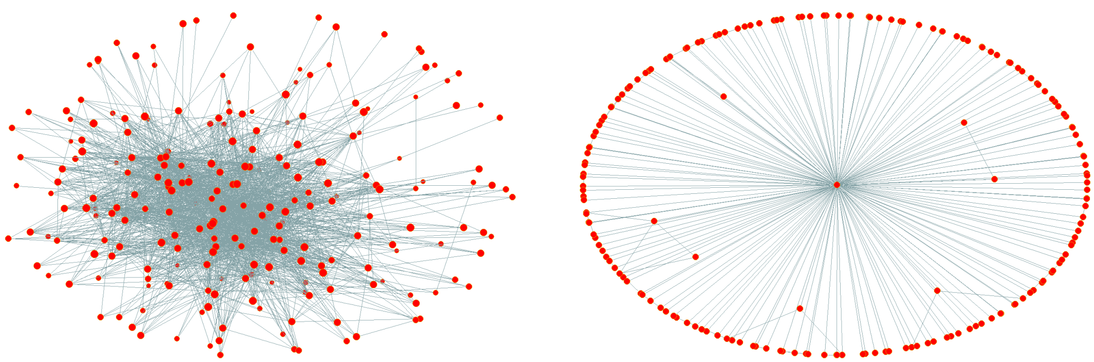
The results in Table 3 are further supported by Fig. 4 which shows the topology of the networks of the four airlines. In particular, Fig. 4 supports the clustering coefficient values obtained: in

	KLM	Lufhansa	EasyJet	Ryanair
n_a	185	240	180	230
$n_{a,l}$	208	434	1.153	2.086
$\bar{k}_a$	2,24	3,61	12,8	18,13
$\bar{C}_a$	0,19	0,5	0,54	0,5
$\bar{S}_a$	1.193	1.174	411	340
$\bar{L}_a$	2,08	2,13	2,24	2,19
$\bar{U}_a$	12,62	11,61	9,8	10,63
$n_{a,F}$	249.413	509.645	474.223	710.542
$n_{a,AC}$	196	359	319	419
$n_{a,K}$	18	20	6	4

Table 3: Example of network and operations-based features for four European airlines in year 2017.

fact, the network shows few triangles in KLM's network, since most flights happen between the hub and another airport. Ryanair's network on the other hand, has a much higher clustering coefficient as there are more connections between any two neighbours of a node. Lufthansa's network has a higher clustering coefficient than KLM's since it has 2 hubs, and the connections resulting thereof.

By analyzing the operations-based features, Table 3 shows that KLM and Lufhansa have a higher utilisation than Easyjet and Ryanair. Also, Ryanair has the largest, homogeneous fleet of aircraft, while KLM and Lufhansa have a much smaller fleet with a large variety of aircraft types.



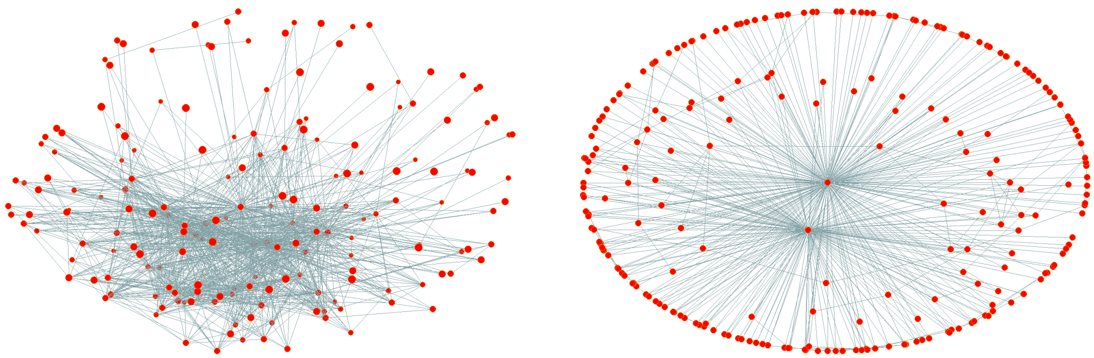
(4.a) Ryanair network graph

(4.b) KLM network graph

4.2. Feature pre-processing using PCA

The features obtained in Section 4.1 are next normalized [26]. This is needed since the scale of some features such as the number of airports of an airline, n_a, is of a much larger magnitude than features such as the clustering coefficient $\bar{C}_a$. Following normalisation, Principle Component Analysis (PCA) [27] is applied to remove the correlation between the features and reduce the dimensionality. This is because features such as the total number of flights of an airline, $n_{a,F}$, and the average strength $\bar{S}$ are highly correlated, i.e., the higher the number of flights, the higher the strength of the network.

Figure 5 shows the fraction of variance from the initial data that is retained when we consider several Principal Components. Five Principal Components are considered for this analysis, since



(4.c) EasyJet network graph (4.d) Lufthansa network graph

Figure 4: Airline networks displayed as circular graphs for flights in 2017 - each node corresponds to an airport, each link corresponds to an Origin-Destination link where at least one flight was completed.

this is the minimum number of components that ensures that more than 90% of the variance of the initial data set is retained.

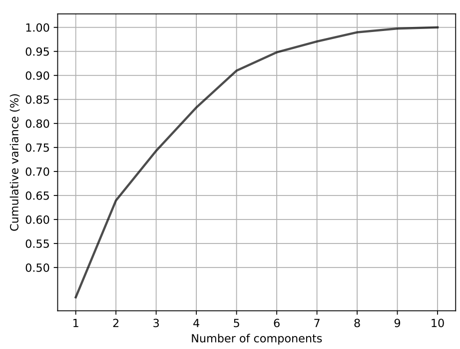


Figure 5: Cumulative variance of the original dataset retained when considering several Principal Components.

Table 4 shows the correlation between the five Principle Components considered and the initial network-based and operations-based features. The last row of Table 4 shows the percentage of variance explained by the 10 initial features. The results show that the number of origin-destination links of an airline, $n_{a,l}$, and the number of airports that an airline departs from/arrives at, n_a, explain more than 60% of the variance in the dataset considered.

4.3. Determining the number of clusters

.

Since the aim of this paper is to cluster European airlines, but the actual number of clusters in which these airlines should be clustered is unknown, the Consensus Index (CI) [28][29] is applied to determine an optimal number of clusters for the given data set.

	$n_{a,l}$	n_a	$\bar{C}_a$	$\bar{k}_a$	$\bar{S}_a$	$\bar{L}$	$n_{a,AC}$	n_k	n_F	$\bar{U}$
PC-1	0,41	0,42	0,19	0,33	0,16	0,02	0,42	0,27	0,43	0,23
PC-2	-0,31	-0,02	-0,3	-0,44	0,56	-0,33	0,2	0,36	0,11	0,13
PC-3	-0,02	0,11	-0,53	-0,07	0,02	0,81	0,12	0,14	0,01	-0,11
PC-4	0,11	0,06	-0,47	0	-0,18	-0,13	-0,11	-0,36	0,03	0,76
PC-5	0,21	-0,27	-0,35	0,19	0,25	-0,18	0,19	-0,52	0,36	-0,44
Variance ratio	0,44	0,2	0,1	0,09	0,08	0,04	0,02	0,02	0,01	0,01

Table 4: Principal Components - features correlations and percentage variance explained by each feature.

The Consensus Index (CI) is determined as follows. First, a range of possible number of clusters $K \geq 2$ is considered. For each K, a set of n clustering solutions is determined, $\{C_1^K, \ldots, C_n^K\}$, where C_i^K is a clustering solution with K clusters. Here, a clustering solution $C_i^K, i \in \{1, \ldots, n\}$ is determined using the K-means++ algorithm [30]. The performance of each K is determined based on a similarity metric $\mathcal{S}(.)$, as follows:

$$CI(\{C_1^K, \ldots, C_{n_K}^K\}) = \frac{1}{n_K(n_K-1)} \sum_{i \neq j} \mathcal{S}(C_i^k, C_j^K),$$

where $\mathcal{S}(.)$ determines the similarity between cluster C_i^k and C_j^k. Finally, an optimal number of clusters K^* is selected such that CI is maximized, i.e.,

$$K^* = \arg\max_{K \geq 2} CI(\{C_1^K, \ldots, C_{n_K}^K\}).$$

Similarity metrics

Two similarity metrics $S(.)$ are considered: the Adjusted Rand Index (ARI) [29] and the Adjusted Mutual Information (AMI) [29].

The Adjusted Rand Index (ARI) is a pair-counting method that quantifies the agreement level between clustering results of different clusters, i.e. whether data points across different clustering solutions belong to the same group. ARI is determined as follows: given a set of n elements and two clustering solutions $\mathbf{X} = \{X_1, \ldots, X_i\}$ and $\mathbf{Y} = \{Y_1, \ldots, Y_j\}$, the overlap between $\mathbf{X}$ and $\mathbf{Y}$ can be summarized in a contingency table, where $[n_{ij}] = X_i \cap Y_j$ (see Table 5 [29]). Using the contingency table between $\mathbf{X}$ and $\mathbf{Y}$, the following four scores are defined: N_{11} is the number of elements that are in the same cluster in $\mathbf{X}$ and $\mathbf{Y}$, N_{10} is the number of elements that are in the same cluster in $\mathbf{X}$ but not in $\mathbf{Y}$, N_{01} is the number of elements that are not in the same cluster in $\mathbf{X}$ but are in the same cluster in $\mathbf{Y}$, and N_{00} is the number of elements that are not in the same cluster in $\mathbf{X}$ and are not in the same cluster in $\mathbf{Y}$.

$\mathbf{U}\backslash\mathbf{V}$	V_1	V_2	$\ldots$	V_C	Sums
U_1	n_{11}	n_{12}	$\ldots$	n_{1C}	a_1
U_2	n_{21}	n_{22}	$\ldots$	n_{2C}	a_2
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
U_R	n_{R1}	n_{R2}	$\ldots$	n_{RC}	a_R
Sums	b_1	b_2	$\ldots$	b_C	$\sum_{ij} n_{ij} = N$

Table 5: Example of a contingency table $n_{ij} = |\mathbf{X_i} \cap \mathbf{Y_i}|$.

Then ARI of two clustering solutions $\mathbf{X}$ and $\mathbf{Y}$ is:

$$\mathrm{ARI}(\mathbf{X}, \mathbf{Y}) = \frac{2\left(N_{00}N_{11} - N_{01}N_{10}\right)}{\left(N_{00} + N_{01}\right)\left(N_{01} + N_{11}\right) + \left(N_{00} + N_{10}\right)\left(N_{10} + N_{11}\right)}. \tag{1}$$

The Adjusted Mutual Information (AMI) indicates how much can be deduced of a cluster $\mathbf{X}$ by looking at cluster $\mathbf{Y}$.

AMI of two clustering solutions $\mathbf{X}$ and $\mathbf{Y}$ is determined as follows [29]:

$$\mathrm{AMI}(\mathbf{X}, \mathbf{Y}) = \frac{MI(\mathbf{X}, \mathbf{Y}) - E\{MI(\mathbf{X}, \mathbf{Y})\}}{\sqrt{H(\mathbf{X})H(\mathbf{Y})} - E\{MI(\mathbf{X}, \mathbf{Y})\}}, \tag{2}$$

where $H(\mathbf{X}) = -\sum_{i=1}^{R} \frac{a_i}{N} \log \frac{a_i}{N}$ is the entropy of $\mathbf{X}$, N is the total number of objects being clustered, and MI is the Mutual Information, defined as:

$$MI(\mathbf{X}, \mathbf{Y}) = \sum_{i=1}^{R}\sum_{j=1}^{C} \frac{n_{ij}}{N} \log \frac{n_{ij}/N}{a_i b_j / N^2}.$$

Figure 6 shows the CI results when considering the similarity metrics ARI and AMI. For each $K \in [2, 14]$, 90% of the considered data set is randomly sub-sampled several times. Each subsample is then assessed against all other subsamples with by similarity measures (AMI and ARI) relating to the same K. Figure 6 shows the results of two independent runs (the results of the first run are shown in solid lines, those of the second run are shown in dotted lines). The highest CI is associated with an optimal number of clusters. Figure 6 shows that the largest CI is obtained for $K^* = 4$.

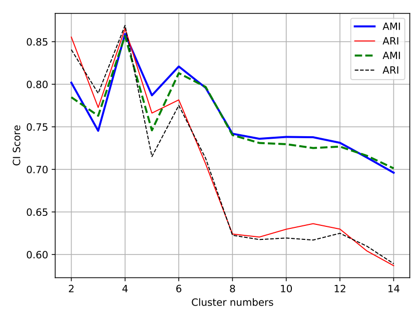


Figure 6: Consensus Index using ARI and AMI metrics.

4.4. Clustering Ensembles

In Section 4.3 an optimal number of clusters $K^* = 4$ has been determined. This section proposes a Clustering Ensemble [25] method to assign all airlines to one of the K^* clusters, implemented by an existing library [31]. A Clustering Ensemble generates a robust clustering solution from a a suite of clustering algorithms, which are aggregated into one final, representative clustering.

To obtain the final clustering, a consensus function is applied on the results of all individual clustering algorithms. Commonly considered consensus functions are Majority voting [32], K-nearest neighbours [33] or a Hierarchical Clustering (HC) algorithm [34]. The HC algorithms are a family of recursive algorithms that perform bottom-up or top-down recursive merging of data points while minimizing specific dissimilarity criteria. The advantage of the HC algorithms is that they allow for visual inspections of the structure of the result, and are deterministic, i.e. a partition can be selected by means of 'cutting' the resulting solution tree, which is referred to as a dendrogram. In this light, the Hierarchically Agglomerative Clustering (HAC) algorithm sub-family is used in this paper [35]. In particular, two HAC variants are used: HAC with a dissimilarity criteria calculated by the unweighted pair group method with arithmetic mean (UPGMA) [36] or average linkage, and HAC with a dissimilarity criteria calculated using the Ward [34] linkage method. Both methods are chosen because both include information from all components of a sub-cluster when calculating the (dis)similarity or distance criteria, as opposed to single [37] [38] or complete linkage [39] [38] which include only one of the extremes of each sub-cluster.

First, the HAC algorithm is briefly described by means of an example, consisting of a dendrogram and its affiliated matrix. The matrix is a representation of either pairwise dissimilarities or distances between the objects being clustered, while the dendrogram is the visual representation of the entire algorithm. Figure 7 shows such a dendrogram obtained with complete linkage [39], resulting from the distance matrices presented in Table 6. Each vertical line in the dendrogram corresponds to a step shown as a new similarity matrix: in the first step, points a and b are the closest, thus are merged in the group $\{ab\}$ and the distances to the remaining data points are recalculated using a complete linkage, i.e., the distance $d(ab, c) = max(d(a, b), d(a, c))$. The same steps are applied until all data points belong to the same cluster.

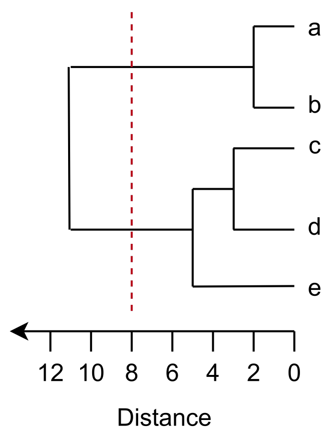


Figure 7: Example dendrogram corresponding to Table 6, cut at a distance of 8, resulting in two clusters: ab and cde

In this paper, the HAC algorithm is applied as a consensus function on the matrix representing the airlines' pairwise occurrences in the same group, instead of individual distances between airline features. As such, the consensus function aggregates the results from all individual partitions, into one, representative clustering.

Suite of clustering algorithms

The consensus function aggregates results from a suite of 6 clustering algorithms which generate 12 individual clustering solutions: one clustering solution is generated by KMeans++ [30],

	a	b	c	d	e
a	0				
b	**2**	0			
c	7	4	0		
d	9	3	3	0	
e	6	11	5	4	0

	ab	c	d	e
ab	0			
c	7	0		
d	9	**3**	0	
e	11	5	4	0

	ab	cd	e
ab	0		
cd	9	0	
e	11	**5**	0

	ab	cde
ab	0	
cde	**11**	0

Table 6: Example of distance matrix and the steps of recursive clustering using complete linkage. In bold the minimum distance at each step, corresponding to the two items being merged.

one clustering solution by Gaussian Mixture Models (GMMs) [40], three clustering solutions by the Hierarchical Agglomerative Clustering (HAC), given by combinations of distance (Manhattan and Euclidian) and average linkage [36]; and Euclidian distance with Ward linkage [34], one clustering solution by the Birch algorithm [41], two clustering solutions generated by the Spectral clustering algorithm (Manhattan and Euclidian distance) and four clustering solutions generated by the HDBSCAN algorithm [42], each one with the following pairs of min_sample and min_cluster_size: $(1, 3)$,$(1, 4)$,$(2, 3)$,$(2, 4)$.

Final clustering

With the 12 clustering solutions (12 ways in which the airlines are grouped in $K^* = 4$ clusters) obtained with the suite of 6 clustering algorithms, a final clustering solution is determined using the HAC algorithm [35]. Here, two dissimilarity criteria are investigated for the HAC algorithm: i) Ward criterion [34], which is based on the variance between sub-clusters and ii) the average linkage criterion [36], which is based on the average Euclidean distance between all points of the sub-clusters considered. Let C_W and C_A denote the generated clustering solutions with HAC using the Ward linkage and the average linkage method, respectively. Table 7 shows several performance indicators: compactness, separation, density, cohesion, dispersion, and the cophenetic correlation coefficient [43] [44] quantified by the following indices s_{dbw} [45], Point Biserial [46], Calinski-Harabasz [47], Dunn indices [48] [49] and the so-called 'modified Hubert' gamma statistic [50] for the C_A and C_W clustering solutions.

	min			max							
	Compactness, separation, density			Cohesion, dispersion						Separation	Similarity
	s_{dbw1}	s_{dbw2}	s_{dbw3}	Silh.	PBS	Calinski	Dunn indices			Hubert	Cophenetic
C_a	1,087	**0,77**	**0,577**	**0,204**	-0,806	**26,018**	0,019	0,101	**0,753**	**10,479**	**0,864**
C_w	**0,987**	0,856	0,642	0,189	**-0,797**	21,227	**0,024**	**0,159**	0,487	10,026	0,428

Table 7: Internal indices comparing C_A and C_W. In bold the best ranking clustering per index.

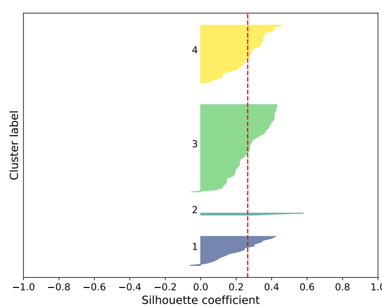
The cophenetic correlation coefficient [36] [51] and the average silhouette coefficient [52] in Table 7 are two of the main indices commonly used to assess clustering results. The cophenetic correlation coefficient [51] indicates how much the resulting dendrogram preserves the original distances between the unmodeled data points. In this case, this index shows that the C_A clusters are 86% similar to the initial, un-clustered data points, by far superior than the C_W clustering, which scores 43%.

The average silhouette coefficient [52] indicates how well an object fits with the cluster it has been assigned to and how well it is separated from the other clusters. However, Table 7 shows that the average Silhouette coefficient for C_A and C_W are similar.

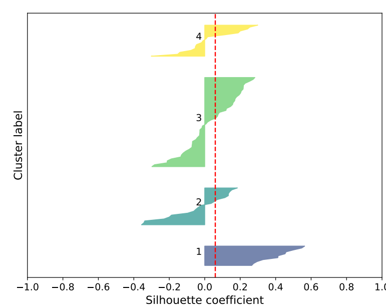
As such, the entire clustering is displayed by combining the silhouettes into single plots in Fig. 8.a and Fig. 8.b. Values close to -1 show points that cluster better with a neighboring group, while values close to 1 represent points that cluster well in their group. The results show that for C_A, almost all airlines seem to be well clustered in their groups, while in contrast, C_W shows that in three out of the four groups, a high number of airlines appear to be ill-clustered, corresponding to the area in the negative domain. Thus, C_A is chosen for the case study.

The other indices are briefly explained next. S_{dbw} is an index that encompasses information about the average compactness of clusters (i.e. intra-cluster distance), the total separation between the clusters (i.e. inter-cluster distance) and inter-cluster density. Two of the three variants of the index (Tong, Halkidiki and Kim) show that C_a performs better. The Point Biserial index is a measure of cohesiveness with values in $(-1, 1)$ and is defined by the difference between the mean value of inter-cluster distance and intra-clusters distances. Although C_W performs better than C_A, it does so by a small fraction. The Calinski-Harabasz index can be described as the multivariate analogue of Fisher's F statistic [53][54]. It clearly shows a better performance for C_A. Dunn's family indices shows how well separated the clusters in a clustering solution are, and although two out of three variants mark C_W as better performing, they do so with a small difference compared to C_A, while the third variant points to C_A by a large margin. Hubert's modified statistic [50], which is a measure of compactness, shows a better result for C_A, yet by a non-negligible advantage..

In conclusion, following the silhouette and the cophenetic correlation coefficient and the overall number of indices scoring better for C_A, the results show that by using the average linkage criterion, the HAC algorithm generates a better final clustering solution. As such, the final airline clustering is obtained with the average linkage criterion (C_A).



(8.a) C_a silhouettes



(8.b) C_w silhouettes

.

Figure 8: Average silhouette per clustering and silhouettes for each cluster (airline group)

4.5. Cluster characterization methodology

The clusters obtained with the above procedures can be characterized in terms of some selected attribute.

In fact, by using the methodology illustrated in [55], given the airlines in the clusters and the attributes of each airline, it is possible to assess the attributes that are more common in the cluster, with respect to a random null hypothesis. The idea is that these attributes are those characterizing the cluster.

Suppose to have a community of K airlines. Suppose that X out of K airlines are characterized by having a certain attribute A. Suppose that in a network of N airlines the attribute A can be associated to M out of N airline. Then the probability that X is observed by chance is given by the hypergeometric distribution H(X; N, M, K). For each attribute present in a cluster and for each cluster in the network one can, therefore, obtain a p-value. By considering an appropriate multiple hypotheses testing correction it is possible to investigate what are the attributes that result to be over-expressed in a cluster with respect to the null hypothesis.

In the present study the attributes considered were the four types of business model (low-cost, flag carrier, leisure, regional) usually defined [8].

A univariate 5% p-value threshold and the Bonferroni correction [56] for multiple comparisons were used.

We emphasize that the method does not simply select the attribute that are more present in the cluster. Rather, it selects those attributes that are more present with respect to a null hypothesis of random occurrence. This is exactly why we annex so much importance to the attributes selected with this procedure. In fact, we believe that if they are not randomly explained, then they might carry information relevant for the understanding of the system. Moreover, the fact that we are considering the Bonferroni correction for multiple test comparison implies that Type I (False Positives) errors are considerably reduced. This therefore ensures that the attributes we select are really characterizing the system, although we can not exclude that other characterizing attributes might be present although our methodology is not able to detect them.

5. Results

This section presents the airline clustering results obtained using the clustering methodology introduced in Section 4.3.

Before discussing the results, we assume the following airline definitions (see also Table 8). We label airlines as large (medium) if it has a fleet size of at least 100 ($(100, 50]$) aircraft, and a network with at least 300 (between $(100, 300]$) links and 150 (between $(50, 150]$) nodes. We label an airline as small if its fleet size is at most 50 aircraft and its network has at most 100 links and at most 50 nodes.

	Large airline	Medium airline	Small airline
$n_{a,AC}$	> 100	$(50, 100]$	≤ 50
$n_{a,l}$	> 300	$(100, 300]$	≤ 100
n_a	> 150	$(50, 150]$	≤ 50

Table 8: Definition of large, medium and small airlines.

As an example, airlines such as KLM, Ryanair, British Airways are considered large, Alitalia, Brussels Airlines, Blue Air and Volotea are considered medium, and Tarom, Czech Airlines and Braathens are considered small airlines.

[8] From the following links: https://en.wikipedia.org/wiki/List_of_low-cost_airlines, https://en.wikipedia.org/wiki/List_of_regional_airlines, https://en.wikipedia.org/wiki/List_of_government-owned_airlines,https://en.wikipedia.org/wiki/List_of_charter_airlines, https://www.eraa.org/membership/our-members and individual Wikipedia pages of airlines.

5.1. Airline clustering results

.

Table 9 shows the resulting four airline clusters and the subclusters of groups 3 and 4. Clusters 1 and 2 consist of only large airlines, while cluster 3 consists of a mix of small and medium airlines. Group 4 is mostly made of medium-sized airlines and a few larger ones. The results of clusters 1 and 2 show that the proposed clustering methodology discriminates objectively between what have been commonly referred to as Low Cost Airlines (LCA) and Full Service Airlines (FSA). Until now, such classification for an airline has been done on ticket prices, on-board service type. In this study, we show that using Machine Learning clustering algorithms with operational and network features, most commonly referred LCAs are clustered together, while commonly referred to FSAs are also clustered together [6, chapters 3.4.2,5.2.2]

Sub-group	Airline	Size
Y1	Air France, Austrian Airlines, Alitalia, British Airways, Brussels Airlines, Lufthansa, Aer Lingus, Finnair, Iberia Airlines, KLM, LOT Polish Airlines, Scandinavian Airlines, British Airways Shuttle, Swiss International Air Lines, TAP Portugal, Turkish Airlines	16
Y2	EasyJet, Ryanair	2
Y3.1	Aigle Azur, Air Nostrum, Blue Air, Blue Panorama, Jet2.com, Luxair Loganair, Monarch Airlines, Primera Air Scandinavia, Transavia NL, Widerøe	11
Y3.2	Aegean Airlines, AlbaStar, Blue Islands, Darwin Airlines, Pegasus, Sunexpress DE, Sunexpress TK	7
Y3.3	Air Malta, Air Serbia, Astra Airlines, Belair Airlines, Air Baltic, Air Corsica, BA CityFlyer, Czech Airlines, Croatia Airlines, Edelweiss Air, Ellinair, Cobalt, Fly One, Atlantic Airways, Sky Wings Airlines, Iberia Express, Icelandair, Bulgaria Air, Montenegro Airlines, Air Moldova, NextJet, Helvetic Airways, People's Vienna, Tarom, Sky Express, Stobart Air, Thomas Cook Airlines, Transavia France	28
Y4.1	Adria Airways, Air Europa, CityJet , Condor Flugdienst, Eurowings, Eastern Airways, Flybe, Germania, Germanwings ,Hop!, Meridiana, TUI Fly Belgium, Norwegian Air Shuttle, Neos, Laudamotion, Malmö Aviation, TUI Airways, Volotea, Vueling Airlines, Wizz Air	20
Y4.2	Air Dolomiti, Chalair, Sun Air of Scandinavia, Twin Jet, Ukraine International Airlines	5
Y4.3	Aurigny Air Services, Albawings, Braathens Regional Airlines, Virgin Atlantic, XL Airways	5

Table 9: Airline subgroups components for subgroups of groups 3 and 4

In order to better analyze the inner structure of the clusters, Fig 9 displays the co-occurrence matrix and its associated dendrogram. The co-occurrence matrix is a $n \times n$ matrix, with n the number of airlines. An entry a_{ij} in the co-occurrence matrix is the fraction of times airline i co-occurs in the same group as airline j with respect to the total number of times individual clustering partitions were obtained, i.e. 12. For the associated dendrogram, a cut at a distance of 0.55 is used to identify the main $K^* = 4$ clusters. The cut is identified by taking the first value that results in $K^* = 4$ from an iterative cutting with a 0.01 incremental step. In our analysis, K* = 4 was reached for the first time when the cophenetic distance $d = 0.55$. The resulting 4 clusters are illustrated in yellow or light areas in the co-occurrence matrix in Fig 9. The dendrogram also shows, from bottom to top, cluster 1 (purple), cluster 2 (red), cluster 4 (green), cluster 3 (orange).

Figure 9 shows that cluster $Y1$ (green)can be considered as composed of 16 airlines. Most of them are popular carriers very well established in their own countries. In fact, as also shown in

Table 10, the trademarks of group 1 are: a large fleet with high average aircraft utilization and many aircraft types, an average path length closest to $Y2$, a low average network degree and average clustering, due to the hub&spoke topology, and a high average strength, signs of intensely flown OD pairs. Based on the characteristics mentioned in section 4.5, group $Y1$ will be referred to as the Full Service Airlines (FSA) group [57].

Group $Y2$ (purple) consists solely of Ryanair and easyJet. These airlines operate many flights,have a large fleets of aircraft, use few aircraft types and their networks have a low average strength and high average degree. As such, group $Y2$ will be referred to as the low-cost group [6], based on the aforementioned behaviour of the average features.

Group $Y3$ (orange) can be considered as composed of three main sub-clusters, when cutting at the distance threshold of 0.3. This is the maximum distance at which a cut results in three subgroups for group $Y3$ (largest subgroups visible in the dendrogram). This group is the largest and consists of a mix of medium-sized airlines, regional airlines that offer both scheduled and charter flights, airlines specialized in long-haul routes (Virgin Atlantic, XL Airways) and medium sized airlines from predominantly eastern European countries.

Group $Y4$ (violet) is the second-largest cluster and can be considered as composed of three main sub-clusters, when cutting at the distance threshold of 0.28. Similar to group $Y3$, this is the maximum distance at which the three main subgroups visible in the dendrogram are obtained. This cluster involves a mix of airlines with different business models.

5.2. Sub-cluster analysis

Table 10 shows how the average features behave across clusters and sub-clusters. The results show that groups $Y1$ and $Y2$ are the largest of the four groups by fleet, network and number of flights. At a first glance it is interesting to notice that the average block time $\bar{U}$ is slightly larger for the second group which operates flights between European OD pairs, while the first group also encompasses airlines that also operate intercontinental flights. This means that notwithstanding the fact the airlines in cluster $Y1$ serve a potentially larger market, airlines in group $Y2$ have developed a policy by which they have a more intensive usage of aircraft, on average. this is probably a key point for explaining how successful these companies have been in the last few decades.

When it comes to network topology, the average path length is fairly similar for all four groups. This indicates that, on average, it takes two legs to cross from one airport to another, meaning that the airlines have managed to optimize their connections, on average, irrespective of the topology of their own network. However the groups' topology is clearly different, indicated by the differences in average degree $\bar{k}$, strength $\bar{S}$ and clustering coefficient $\bar{C}$. Group 2 has a bigger network in terms of links and airports, and airport nodes are connected by more links than those of group $Y1$. However, these links are flown at an intensity of 45% compared to those of group 1. $\bar{C}$ gives another insight: even though all networks have an $\bar{L}$ of approximately 2, the airlines show two distinct choices: one with higher clustering and less hubs (groups $Y2$ and $Y4$), and one with lower clustering and a tendency towards more hubs (groups $Y1$ and $Y3$).

Groups $Y3$ and $Y4$ are rather large and several subgroups can be observed within those large groups. Table 10 shows the average feature values for all groups, and for each three of the subgroups in groups $Y3$ and $Y4$, referred to as $Y3.1, Y3.2, Y3.3$ and $Y4.1, Y4.2, Y4.3$ respectively.

The airline sub-clusters are presented in Table 9.

Subcluster $Y3.1$ involves LCCs, Regional and Leisure airlines, predominantly medium-sized. The only large airline present in this subcluster is Luxair, which is medium by fleet and large by network.

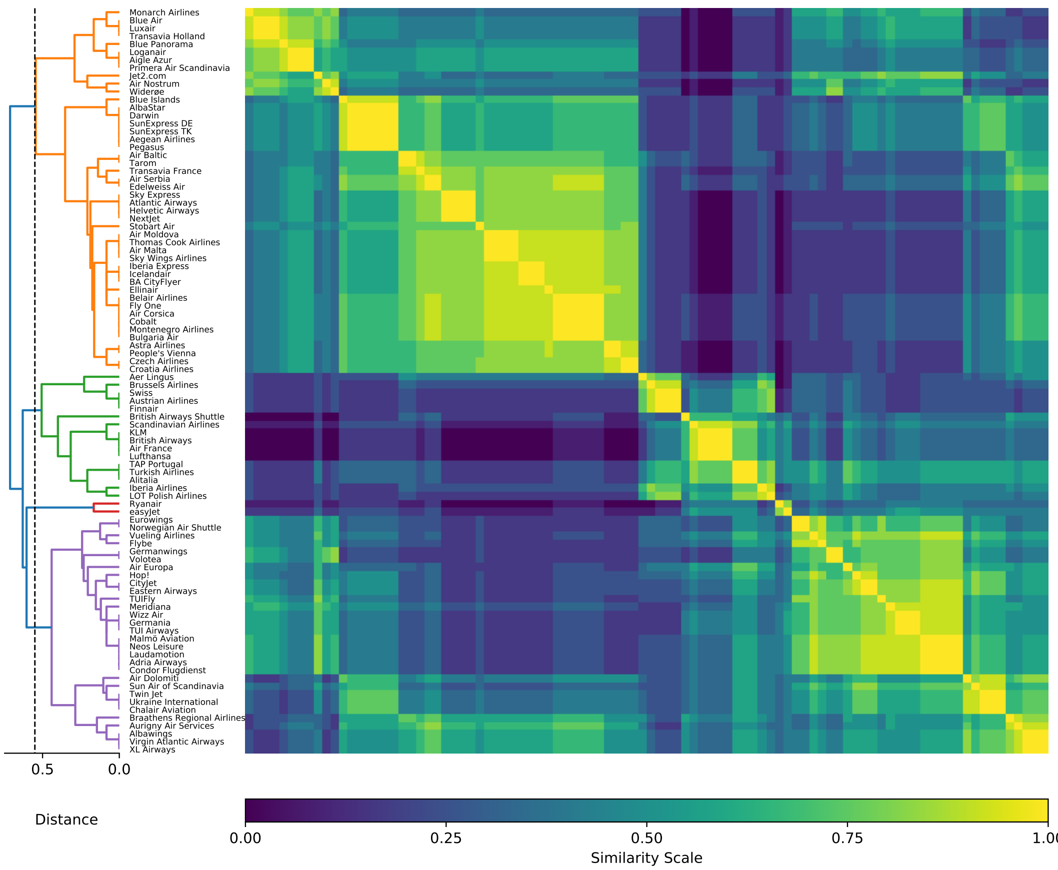


Figure 9: The clustering ensemble co-occurrence matrix with a similarity band and the associated dendrogram.

Groups	n_L	n_A	$\bar{U}$	n_F	n_{AC}	n_K	$\bar{k}$	$\bar{L}$	$\bar{S}$	$\bar{C}$
1	235	143	9,8	180187	178	17	3,2	2,13	825	0,33
2	1620	205	10,22	592383	369	5	15,48	2,22	376	0,52
3	119	61	7	23275	30	6	3,68	2,2	190	0,37
3.1	227	85	8,18	43356	40	7	5,36	2,23	186	0,5
3.2	119	57	5,95	19771	38	4	4,06	2,47	134	0,28
3.3	76	52	6,49	16263	24	6	2,93	2,11	205	0,34
4	326	105	8	48602	68	10	5,22	2,34	225	0,37
4.1	452	136	8,14	66282	91	12	6,45	2,33	161	0,47
4.2	107	55	4,12	12969	23	5	2,93	2,66	347	0,13
4.3	37	28	8,94	13515	24	5	2,6	2,06	357	0,23

Table 10: Average airline feature values for the resulting groups

The second subcluster $Y3.2$ is smaller and consists of a balanced mix between small regional airlines and medium LCCs, with one medium flag carrier standing out here as well: Aegean Airlines.

The third subcluster $Y3.3$ is the largest of group $Y3$ and of the entire dataset, consisting mainly of small and medium-sized airlines. The majority is formed by national carriers of the Baltics and eastern European countries (Romania, Croatia, Czech Republic etc). The rest of the group is split evenly between small regional airlines and small and medium leisure airlines. Interesting to notice that four out of the 13 nowadays defunct airlines are in this sub-group: Astra Airlines, Belair and Sky Wings; additionally with People's Vienna operating two destinations only by the end of 2020. Since this subcluster contains most of the other known Full Service Airlines, it is compared with group $Y1$, consisting of large FSAs. The network sizes of subgroup 3.3 are three times smaller and have more than seven times less aircraft and 10 times less flights on average, than group $Y1$. Interesting to notice however that subgroup $Y3.3$ has also three times less aircraft types than group $Y1$, and 3 out of 4 complex network metrics have similar values, except of the average strength $\bar{S}$ which is dependent on the number of flights.

In group $Y4$, the first sub-cluster $Y4.1$ consists of 20 airlines, six of which are currently defunct (Adria Airways, CityJet, Flybe, Germania, Germanwings and Meridiana). Looking at its composition, it consists of an equal mix of medium and large airlines and is rather heterogeneous in terms of airline types: all of them are present in similar numbers.

The second sub-cluster $Y4.2$ consists of only 5 airlines, predomintantly small and regional. The odd one is Ukraine's flag carrier, but when comparing the individual feature values, there are not that large discrepancies ($n_L = 66$, $n_A = 54$,$\bar{U} = 6.2$, $n_F = 30067$, $n_AC = 54$, $n_K = 10$, $\bar{k} = 2,44$, $\bar{L} = 2,49$, $\bar{S} = 455$ and $\bar{C} = 0,05$).

Similarly, the last sub-cluster $Y4.3$ consists of mostly small low-cost airlines (3/5) but the airlines are too small for a relevant comparison with group 2.

5.3. Analysis of airline features

In this section, the network and operational features are analyzed in the context of the obtained clusters and sub-clusters.

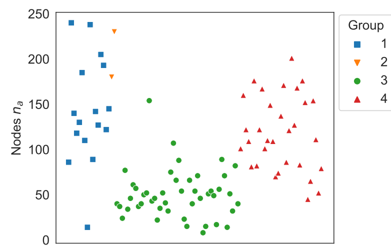
Figure 10.a shows the number of airports (destinations) of each airline. Here, an airline in cluster $Y3$ has a lower number of airport nodes in its network, of up to 100 nodes, while groups $Y2$, $Y4$ and $Y1$ all have between 50 and 250 airport nodes. A lower number of airport nodes for an airline in cluster $Y3$ is expected, as discussed in section 5.2.

The number of links in Figure 10.b shows more similar values among groups, with group $Y2$ clearly separated from the rest, group $Y3$ again with the lowest values, group $Y1$ with a large heterogeneity, values being between few links and roughly 700, and group $Y4$ spread between approximately 100 and 1000 links.

The number of aircraft types in Fig 10.c follows a similar pattern with the number of airports, the main difference being that values of group $Y2$ are much lower relative to the rest. Similarly, the fleet size follows similar behaviour as the network size, only more widely spread out for group $Y1$.

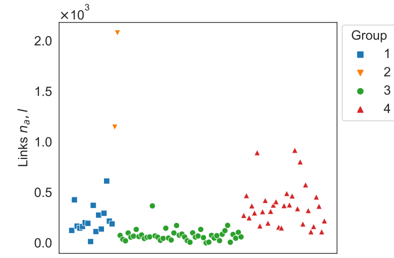
The fleet sizes are depicted in Fig 10.d and follow a similar pattern as with the number of links. As expected, major carriers from groups $Y1$ and $Y2$ exhibit the largest fleets, however a few airlines in group $Y4$ also have large fleets of more than 150 aircraft, with Eurowings showcasing a fleet of 226 aircraft in 2017. Only 8 of the 94 airlines considered in this study have more than 200 aircraft in their fleet.

The block time is shown in Fig 10.e and is one of the most spread-out features for all groups, with a remarkable 16 and 18 - hours average utilization for XL Airways and Virgin Atlantic - a feature to be expected due to the high number of long-haul flights. The total number of flights in 2017 for each airline is presented in Fig 10.f and follows a similar behavior as the fleet size and number of aircraft types, with group $Y1$ spread over half an order of magnitude, group $Y2$ with the highest average number of flights, group $Y3$ with the lowest average number of flights, and group $Y4$ with a maximum of 200.000 flights by Vueling.

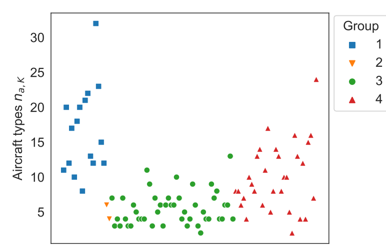


(10.a) Number of airport nodes per airline

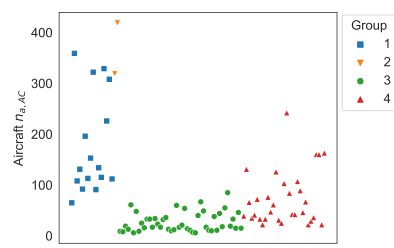
types.pdf



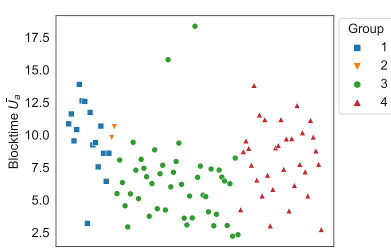
(10.b) Number of links per airline



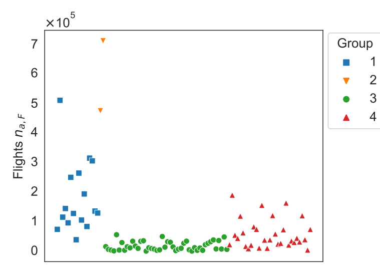
(10.c) Aircraft types per airline



(10.d) Fleet size per airline



(10.e) Average block time per airline



(10.f) Total flights per airline

Figure 10: Individual operational features per airline groups

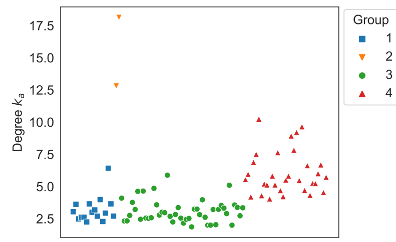
The complex network features are presented in Figure 11 and here a few remarks are in order, too. The average degree shown in Fig 11.a points to a more compact group $Y1$, when compared to the same groups' operational features. This happens most likely because of the predominantly hub&spoke networks that flag carriers have. Surprisingly however, the network of Scandinavian Airlines has an average degree of 6, due to the fact that the carrier operates from 3 hubs. Other high values for group $Y3$ are observed for subgroup $Y3.1$ mainly due to the presence of low-cost carriers. The average clustering coefficient is shown in Fig 11.b and together with the average path length, varies the most within each group. Interesting to notice that one flag carrier in group $Y1$ (Scandinavian Airlines) has a higher average clustering than any of the major low cost carriers in group $Y2$.

Sub-group $Y4.1$ differentiates itself from the rest of group $Y4$ with a high clustering coefficient due to the large point-to-point networks of the large leisure and low-cost carriers. The average path length in Fig 11.c paints a well-known picture for flag carriers, with more values in the group closer to 2 due to the hub&spoke topology. Group $Y2$ also shows that, despite a completely different topology, the network can be optimally traversed by flying on average 2.2 legs. Group $Y3$ varies the most, spreading average path values between the minimum (Albawings) and the maximum (Sun Air of Scandinavia) of the entire data set. Finally, the average strength in Fig 11.d shows that, even though group $Y2$ has the most flights flown overall, it flies its OD pairs less intensely as the entire group $Y1$ and a quarter of group $Y3$. On the contrary, group $Y4$ maintains the same variance for strength as with the total number of flights.

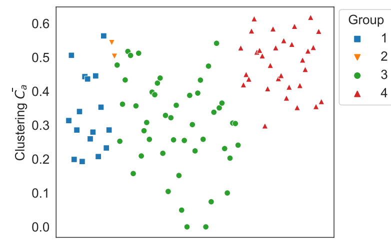
The individual features show that, for group $Y1$, values vary a lot for most features except the number of links in the network, the average path and the average degree. Surprisingly, all components of group $Y1$ show higher strength values compared to group Y2, which has the largest number of flights compared to any of the groups and accounts for nearly 18% of all flights considered in the data set, even if the group consists of only two airlines. Group $Y3$ has some of the smallest networks and least amount of flights, but boasts some of the highest average aircraft utilization per airline and presents high variance for most of the complex network features. This is somewhat to be expected as the group is made of three subclusters that consist of airlines with rather different business models. Finally group $Y4$ shows an inverse behaviour compared to group $Y3$: only 3 features (number of destinations, aircraft types and average block time) are further spread out; in the remaining 7 cases, the maximum difference between airline features is constant, yet on average larger than the maximum spread in group $Y3$.

5.4. Airline cluster characterization

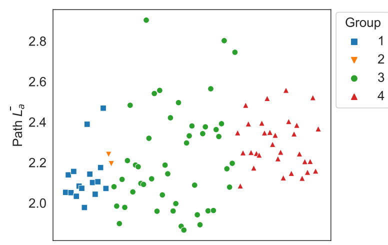
Given the main four clusters reported in Table 9 we have considered the methodology briefly illustrated in Section 4.5 in order to investigate whether such clusters can be characterized in terms of the following attributes (see Appendix A): Low cost carrier (LCC), Flag carrier, Regional carrier, Leisure/Charter carrier. The results are shown in Table 11. The table reports all tests that have been performed, depending on the presence of airlines with a certain attribute in a certain cluster. The second column gives the size of the considered cluster. The fourth column indicates how many airlines with a certain attribute (column 3) are present in our set of 94 airlines. The fifth column gives the number of airlines in the cluster with that attribute. the p-value shown in the seventh column must be compared with the Bonferroni threshold shown in the last column. Each cluster has its own Bonferroni threshold. In fact, since the number of attributes might change from cluster to cluster, the number of performed tests is in principle different for each cluster and therefore the Bonferroni threshold changes. We considered an univariate 5% threshold.



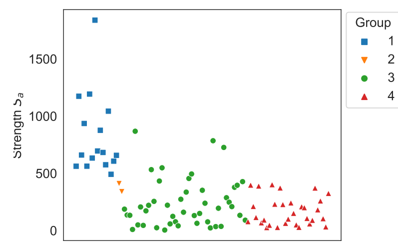
(11.a) Average degree per airline



(11.b) Average clustering coefficient



(11.c) Average path length per airline



(11.d) Average strength per airline

Figure 11: Individual complex network features per airline

As a result, we have that the first cluster $Y1$ is characterized by the attribute ‘Flag Carrier’ and the second cluster $Y2$ is characterized by the attribute ‘Low Cost Carrier’.

The Bonferroni correction is quite severe and is the most conservative choice that can be done when correcting for multiple test comparison. A less stringent correction is given by the False Discovery Rate (FDR). When applying this correction to our data, we obtain the same results as with the Bonferroni correction.

At the level of subclusters, our methodology is able to characterizie subcluster $Y4.2$ in terms of the ”Regional carrier” attribute.

By summarizing, we find that two cluster, i.e. cluster $Y1$ and cluster $Y2$, are composed of airlines that are quite homogeneous from the point of view of the categorizations commonly used in the literature: Flag carriers and Low-Cost carriers. However, we also find two more clusters whose composition is quite heterogeneous when considering that categorization. This is why we find that none of these categorizations is strong enough as to give statistically validated results, not even at the level of sub-clusters, where indeed we have only one statistically robust characterization.

This puts forward the idea that our methodology better captures the way airlines group together when adapting themselves to the continuously changing needs of today’s aviation market.

cluster	MM	attribute	KK	XX	NN	p-over	p-under	thresh BONF
Y3	46	Leisure	14	9	94	0.170	0.938	0.007
Y3	46	Regional	20	11	94	0.360	0.806	0.007
Y3	46	LCC	18	8	94	0.753	0.436	0.007
Y3	46	Flag	34	15	94	0.820	0.313	0.007
Y4	30	Leisure	14	5	94	0.481	0.744	0.006
Y4	30	Regional	20	9	94	0.127	0.952	0.006
Y4	30	LCC	18	8	94	0.161	0.937	0.006
Y4	30	Flag	34	2	94	1.000	0.000	0.006
Y1	16	Flag	34	16	94	4,802 10^{-9}	1.00	0.05
Y2	2	LCC	18	2	94	0.035	1.00	0.05

Table 11: Cluster characterization: the table reports all tests that have been performed, depending on the presence of airlines with a certain attribute in a certain cluster.Further explanation in the text.

5.5. Seasonality analysis

As with any other industry, air transportation consists of cycles, two of such cycles being the summer and winter scheduling season, whereby airlines and airports update their slots[9]. This occurs at the beginning and ending of the daylight saving time, and in our case it coincided with the dates of 26.03.2017 and 29.10.2017.

We have therefore split our dataset according to these two cycles. In particular we have considered the summer season from 26.03.2017 to 29.10.2017. We did not consider the winter period because we had no data relative to 2016 and 2018 and we did not deem appropriate to perform our analyses fpr two disjoint period as a whole.

The clustering methodology has been applied to the reduced dataset as described in Section 4. In fact, we considered the Z-scores and 5 Principal Components have been extracted from the initial 10 features as well as the optimal cluster number has been determined by using the Consensus Index Methodology. This gave us again 4 possible main clusters with another peak at $k = 7$ suggesting another possible partition in subclusters, as illustrated in Figure 12.

[9]https://www.iata.org/contentassets/4ede2aabfcc14a55919e468054d714fe/calendar-coordination-activities.pdf

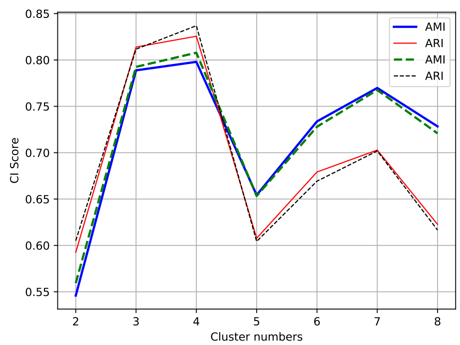


Figure 12: Consensus index scores for the optimal cluster numbers during summer season

Similar to the case when the entire year of 2017 is considered (Table 7), the average linkage yields overall better results in terms of silhouettes, cophenetic correlation coefficient and the rest of the internal validation indices (not shown in this paper). Thus, Table 12 presents the partition of the airlines into the 4 main clusters for the summer season dataset. Analogously, Figure 13 displays the co-occurrence matrix for the summer season clustering, with the cut associated to the above mentioned 4 groups.

Group	Airline	Size
S1	Air France, British Airways, Lufthansa, KLM, Scandinavian Airlines, Turkish Airlines	7
S2	Austrian Airlines, Brussels Airlines, Aer Lingus, Finnair, Iberia, LOT, Swiss	6
S3	easyJet, Ryanair	2
S4	Aigle Azur, Adria Airways, Air Europa, Aegean Airlines, Air Malta, Air Nostrum, Air Serbia, Ukraine International, Aurigny Air Services, Alitalia, Astra Airlines, CityJet, Flybe, Blue Air, Blue Panorama , Braathens Regional Airlines, Air Baltic, BA CityFlyer, Condor Flugdienst, Chalair Aviation, Czech Airlines, Croatia Airlines, Air Dolomiti, Darwin, Edelweiss Air, Ellinair, Eurowings, Jet2.com, Eastern Airways, Fly One, Germania, Sky Wings Airlines, Germanwings, Hop!, Iberia Express, Icelandair, Meridiana, TUIFly, AlbaStar, Luxair, Loganair, Bulgaria Air, Montenegro Airlines, Air Moldova, Monarch Airlines, Norwegian Air Shuttle, Laudamotion, NextJet, Helvetic Airways, Pegasus, Primera Air Scandinavia, Tarom, Malmö Aviation, Sky Express, British Airways Shuttle, Stobart Air, Sun Air of Scandinavia, SunExpress DE, SunExpress TK, TAP Portugal, Thomas Cook, Twin Jet, TUI Airways, Transavia NL, Transavia FR, Virgin Atlantic Airways, Vueling Airlines, Volotea, Widerøe, Wizz Air, XL Airways	77

Table 12: Airline names resulting from the clustering procedure for the summer season dataset.

A total of 92 out of 94 airlines have been included in the summer season clustering based on similar criteria as before: to have flown an average of at least 10 flights per month or have a fleet of at least 5 aircraft. The two airlines not included are Neos and Cobalt, meaning that these two airlines have slightly more flights during the winter season which put them above the minimum of 10 flights per month during the entire year.

The clustering results indicate that cluster $S2$ consists now of medium-large FSAs, while cluster $S1$ can be mapped almost directly to a subcluster of cluster $Y1$. Except for large LCCs, grouped again together in summer cluster $S3$, the rest of the airlines seem more undefined when considering only the 7 months period of the summer season. One interesting point is that TAP Portugal and Alitalia are no longer part of the FSA clusters $S1$ and $S2$, meaning that their operations deviate from those of the other FSA during summer and start to be more similar to the other airlines.

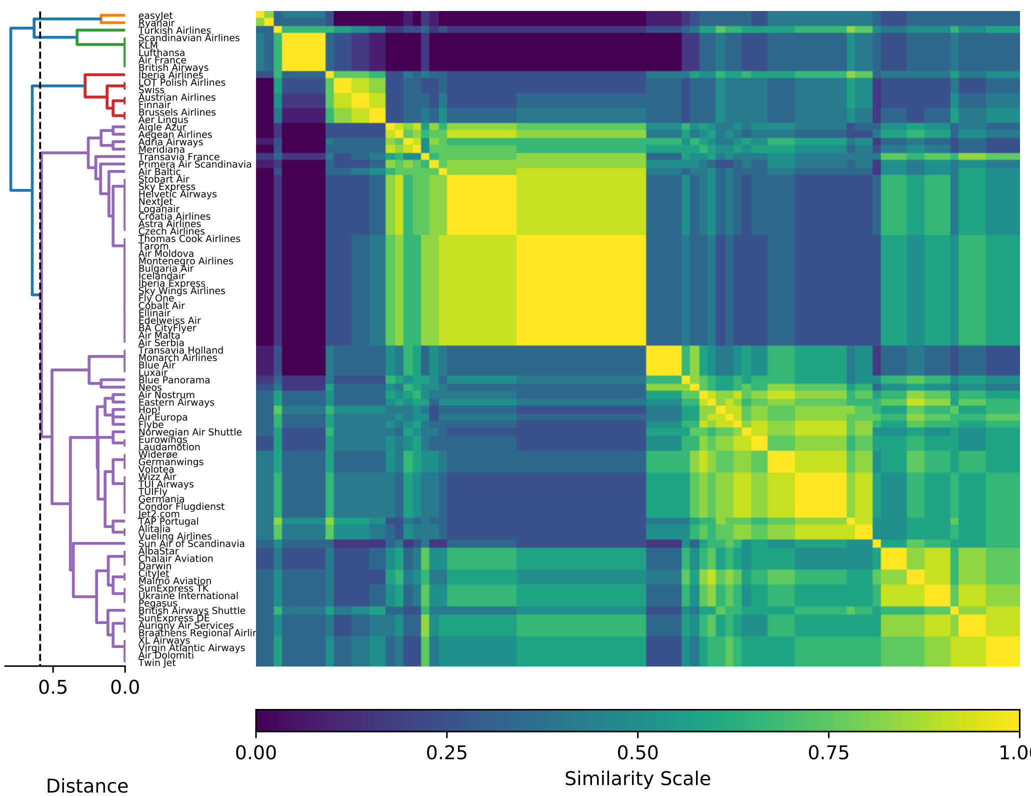


Figure 13: The clustering ensemble co-occurrence matrix for summer season dataset.

The above results suggest that our methodology, based on a balanced mix of operational features and complex network metrics, is quite able to detect in a robust way groups of airlines. Moreover, it is sufficiently powerful as to keep trace of different partitions of airlines due to the fact that different market conditions (summer versus winter) trigger different operational settings.

6. Conclusion

In this paper, an unsupervised clustering ensemble methodology is proposed, which clusters 94 European airlines into 4 groups. The number of groups have been defined by assessing what would best fit given the given dataset, consisting of 10 operational, fleet and complex network features of airlines, extracted from data covering all the flights that crossed the ECAC area in 2017. This classification provides the aviation community an updated picture of what actually can be referred to as 'real' low-cost carriers or flag carriers and clearly differentiates between the two business models in a holistic manner based on a set of objective characteristics extracted solely from airline operations and their networks. Besides characterizing the identified clusters in a unified manner, the methodology identifies a set of objective features that have been proven relevant in the issue of classifying airlines and paves the way towards an updated nomenclature for all European carries.

Regarding the content of the groups, two of the four resulting groups contain well-defined airline types: what are currently labeled as large low-cost and medium-large network (flag) carriers, making the clustering methodology, to the best of our knowledge, the first one that accurately differentiates between the two types. The other two groups, although larger and more heterogeneous, contain well-defined subclusters, consisting predominantly of one type of airline (characterized by business models such as flag carriers, leisure, low-cost, regional) and the size of the airlines (large, medium, small). The airlines appearing in minority in such sub-clusters, considered 'unexpected' entrants, are precisely an example of how the nomenclature does not keep up with the need of airlines to adapt and change their business models to survive the fast-changing economic landscape.

On the other hand, the presence of such large groups with well-defined subclusters also highlights the need for using more features in the classification, that will potentially increase the granularity of result. Some viable features include the average length of a flight or whether most destinations lie in the same country(ies), which could identify better regional airlines. Other feature examples includes the average spatial deviation from planned route, delays or the age of the fleet.

[1] K. Mason, W. G. Morrison, I. Stockman, in: Liberalization in Aviation: Competition, Cooperation and Public Policy; Forsyth, P., Gillen, D., Hüschelrath, K., Niemeier, HM, Wolf, H., Eds (2016) 141–156.

[2] S. A. Morrison, Journal of Transport Economics and Policy 35 (2001) 239–256.

[3] J. Gllavata, R. Ewerth, B. Freisleben, in: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., volume 1, IEEE, pp. 425–428.

[4] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, R. R. Choudhury, in: Proceedings of the 10th international conference on Mobile systems, applications, and services, pp. 197–210.

[5] R. G. Wijnhoven, P. H. de With, in: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), IEEE, pp. 2077–2083.

[6] P. Belobaba, A. R. Odoni, C. Barnhart, The global airline industry, Wiley, 2009.

[7] K. J. Mason, W. G. Morrison, Research in Transportation Economics 24 (2008) 75–84.

[8] R. Doganis, Airline business in the 21st century, Routledge, 2005.

[9] R. Klophaus, R. Conrady, F. Fichert, Journal of Air Transport Management 23 (2012) 54–58.

[10] J. Daft, S. Albers, Journal of Air Transport Management 46 (2015) 3–11.

[11] J. Daft, S. Albers, Journal of Air Transport Management 28 (2013) 47–54.

[12] M. Urban, M. Klemm, K. O. Ploetner, M. Hornung, Journal of Air Transport Management 71 (2018) 175–192.

[13] A. Cook, H. A. Blom, F. Lillo, R. N. Mantegna, S. Micciche, D. Rivas, R. Vazquez, M. Zanin, Journal of Air Transport Management 42 (2015) 149–158.

[14] R. Pastor-Satorras, A. Vázquez, A. Vespignani, Physical Review Letters 87 (2001) 258701.

[15] S. Porta, P. Crucitti, V. Latora, Environment and Planning B: Urban Analytics and City Science 33 (2006) 705–725.

[16] P. Sen, S. Dasgupta, A. Chatterjee, P. Sreeram, G. Mukherjee, S. Manna, Physical Review E 67 (2003) 036106.

[17] R. Pastor-Satorras, A. Vespignani, Physical Review Letters 86 (2001) 3200.

[18] A. Bombelli, B. F. Santos, L. Tavasszy, Transportation Research Part E: Logistics and Transportation Review 138 (2020) 101959.

[19] C. Kai-Quan, Z. Jun, D. Wen-Bo, C. Xian-Bin, Chinese Physics B 21 (2012) 028903.

[20] R. Guimera, L. A. N. Amaral, The European Physical Journal B 38 (2004) 381–385.

[21] M. M. Hossain, S. Alam, Journal of Air Transport Management 60 (2017) 1–9.

[22] G. Bagler, Physica A: Statistical Mechanics and its Applications 387 (2008) 2972–2980.

[23] Z. Xu, R. Harriss, GeoJournal 73 (2008) 87–102.

[24] M. Zanin, F. Lillo, European Physical Journal: Special Topics 215 (2013) 5–21.

[25] S. Monti, P. Tamayo, J. Mesirov, T. Golub, Machine learning 52 (2003) 91–118.

[26] A. B. Graf, S. Borer, in: Joint pattern recognition symposium, Springer, pp. 277–282.

[27] C. Ding, X. He, in: Proceedings of the twenty-first international conference on Machine learning, p. 29.

[28] N. X. Vinh, J. Epps, in: 2009 Ninth IEEE International Conference on Bioinformatics and BioEngineering, IEEE, pp. 84–91.

[29] N. X. Vinh, J. Epps, J. Bailey, Journal of Machine Learning Research 11 (2010) 2837–2854.

[30] D. Arthur, S. Vassilvitskii, in: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, pp. 1027–1035.

[31] T. Ronan, S. Anastasio, Z. Qi, P. H. S. V. Tavares, R. Sloutsky, K. M. Naegle, Journal of Machine Learning Research 19 (2018) 1–6.

[32] A. Fred, in: International Workshop on Multiple Classifier Systems, Springer, pp. 309–318.

[33] C. Domeniconi, B. Yan, in: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., volume 1, IEEE, pp. 228–231.

[34] J. H. Ward Jr, Journal of the American statistical association 58 (1963) 236–244.

[35] K. C. Gowda, G. Krishna, Pattern recognition 10 (1978) 105–112.

[36] R. R. Sokal, Univ. Kansas, Sci. Bull. 38 (1958) 1409–1438.

[37] R. Sibson, The Computer Journal 16 (1973) 30–34.

[38] T. A. Sorensen, Biol. Skar. 5 (1948) 1–34.

[39] D. Defays, The Computer Journal 20 (1977) 364–366.

[40] D. Reynolds, Gaussian Mixture Models, Springer US, Boston, MA, 2009, pp. 659–663.

[41] T. Zhang, R. Ramakrishnan, M. Livny, ACM Sigmod Record 25 (1996) 103–114.

[42] L. McInnes, J. Healy, S. Astels, The Journal of Open Source Software 2 (2017).

[43] R. R. Sokal, F. J. Rohlf, Taxon (1962) 33–40.

[44] F. J. Rohlf, D. R. Fisher, Systematic Biology 17 (1968) 407–412.

[45] J. Tong, H. Tan, Journal of Electronics (China) 26 (2009) 258–264.

[46] R. F. Tate, The Annals of mathematical statistics 25 (1954) 603–607.

[47] T. Caliński, J. Harabasz, Communications in Statistics-theory and Methods 3 (1974) 1–27.

[48] J. C. Dunn, Journal of Cybernetics 3 (1973) 32–57.

[49] J. C. Dunn, Journal of Cybernetics 4 (1974) 95–104.

[50] L. J. Hubert, J. R. Levin, Psychological bulletin 83 (1976) 1072.

[51] R. R. Sokal, F. J. Rohlf, Taxon (1962) 33–40.

[52] P. J. Rousseeuw, Journal of computational and applied mathematics 20 (1987) 53–65.

[53] R. A. Fisher, Journal of the Royal Statistical Society 85 (1922) 87–94.

[54] R. A. Fisher, in: Breakthroughs in statistics, Springer, 1992, pp. 66–70.

[55] M. Tumminello, S. Micciche, F. Lillo, J. Varho, J. Piilo, R. N. Mantegna, J. Stat. Mech. (2011) P01019.

[56] O. J. Dunn, Journal of the American statistical association 56 (1961) 52–64.

[57] E. Pels, Research in transportation economics 24 (2008) 68–74.

Appendix A. List of airlines

This appendix contains the list of airlines and some information about their business model [10]
The four main types are flag, LCC, regional and leisure or charter. Where the type of airline was unclear, its operations are described based on the manner in which the airports from the list of destinations [11] are mentioned.

Code	Group	Airline	Type	Defunct
AFR	1	Air France	Flag	0
AUA	1	Austrian Airlines	Flag	0
AZA	1	Alitalia	Flag	0
BAW	1	British Airways	Flag	0
BEL	1	Brussels Airlines	Flag	0
DLH	1	Lufthansa	Flag	0
EIN	1	Aer Lingus	Flag	0
FIN	1	Finnair	Flag	0
IBE	1	Iberia Airlines	Flag	0
KLM	1	KLM	Flag	0
LOT	1	LOT Polish Airlines	Flag	0
SAS	1	Scandinavian Airlines	Flag	0
SHT	1	British Airways Shuttle	Flag	0
SWR	1	Swiss	Flag	0
TAP	1	TAP Portugal	Flag	0
THY	1	Turkish Airlines	Flag	0
EZY	2	easyJet	LCC	0
RYR	2	Ryanair	LCC	0
AAF	3	Aigle Azur	Scheduled	1
AEE	3	Aegean Airlines	Flag	0
AMC	3	Air Malta	Flag	0
ANE	3	Air Nostrum	Regional	0
ASL	3	Air Serbia	Flag	0
AZI	3	Astra Airlines	Regional	1
BCI	3	Blue Islands	Regional	0
BHP	3	Belair Airlines	Leisure	1
BMS	3	Blue Air	LCC	0

[10] From the following links: https://en.wikipedia.org/wiki/List_of_low-cost_airlines, https://en.wikipedia.org/wiki/List_of_regional_airlines, https://en.wikipedia.org/wiki/List_of_government-owned_airlines,https://en.wikipedia.org/wiki/List_of_charter_airlines, https://www.eraa.org/membership/our-members and individual Wikipedia pages of airlines.

[11] for example https://en.wikipedia.org/wiki/List_of_Cobalt_Air_destinations

BPA	3	Blue Panorama	Leisure	0
BTI	3	Air Baltic	Flag	0
CCM	3	Air Corsica	Regional	0
CFE	3	BA CityFlyer	Flag	0
CSA	3	Czech Airlines	Flag	0
CTN	3	Croatia Airlines	Flag	0
DWT	3	Darwin	Regional	1
EDW	3	Edelweiss Air	Leisure	0
ELB	3	Ellinair	Leisure	0
EXS	3	Jet2.com	LCC	0
FCB	3	Cobalt Air	Seasonal + scheduled	1
FIA	3	Fly One	LCC	0
FLI	3	Atlantic Airways	Flag	0
GSW	3	Sky Wings Airlines	Leisure	1
IBS	3	Iberia Express	Flag	0
ICE	3	Icelandair	Flag	0
LAV	3	AlbaStar	Leisure	0
LGL	3	Luxair	Flag	0
LOG	3	Loganair	Regional	0
LZB	3	Bulgaria Air	Flag	0
MGX	3	Montenegro Airlines	Flag	0
MLD	3	Air Moldova	Flag	0
MON	3	Monarch Airlines	Leisure +scheduled	0
NTJ	3	NextJet	Regional	0
OAW	3	Helvetic Airways	Regional	0
PEV	3	People's Vienna	Leisure	0
PGT	3	Pegasus	LCC	0
PRI	3	Primera Air Scandinavia	Leisure	1
ROT	3	Tarom	Flag	0
SEH	3	Sky Express	Regional	0
STK	3	Stobart Air	Regional	0
SXD	3	SunExpress DE	LCC	0
SXS	3	SunExpress TK	LCC	0
TCW	3	Thomas Cook Airlines	Leisure	0
TRA	3	Transavia Holland	LCC	0
TVF	3	Transavia France	LCC	0
WIF	3	Widerøe	Regional	0
ADR	4	Adria Airways	Flag	1

AEA	4	Air Europa	Scheduled	0
AUI	4	Ukraine International	Flag	0
AUR	4	Aurigny Air Services	Flag	0
AWT	4	Albawings	LCC	0
BCY	4	CityJet	Regional	1
BEE	4	Flybe	Regional	1
BRX	4	Braathens Regional Airlines	Regional	0
CFG	4	Condor Flugdienst	Leisure	0
CLG	4	Chalair Aviation	Regional	0
DLA	4	Air Dolomiti	Regional	0
EWG	4	Eurowings	LCC	0
EZE	4	Eastern Airways	Regional	0
GMI	4	Germania	Seasonal + leisure	1
GWI	4	Germanwings	LCC	1
HOP	4	Hop!	Regional	0
ISS	4	Meridiana	Leisure	1
JAF	4	TUIFly	Leisure	0
NAX	4	Norwegian Air Shuttle	LCC	0
NLY	4	Laudamotion	LCC	0
NOS	4	Neos	Leisure	0
SCW	4	Malmö Aviation	Regional + charter	0
SUS	4	Sun Air of Scandinavia	Regional	0
TJT	4	Twin Jet	Regional	0
TOM	4	TUI Airways	Leisure	0
VIR	4	Virgin Atlantic Airways	Scheduled long - haul	0
VLG	4	Vueling Airlines	LCC	0
VOE	4	Volotea	LCC	0
WZZ	4	Wizz Air	LCC	0
XLF	4	XL Airways	Scheduled long - haul	0

Table A.13: Airline names, type, corresponding group, type of operations and whether in December 2020 they were defunct or not.