



HAL
open science

Data-driven airline passenger flight supply comparison with optimal transport theory

Qian LIU, Paul Rochet, Chantal Roucolle

► **To cite this version:**

Qian LIU, Paul Rochet, Chantal Roucolle. Data-driven airline passenger flight supply comparison with optimal transport theory. 2026. <hal-05505239>

HAL Id: hal-05505239

<https://hal.science/hal-05505239v1>

Preprint submitted on 11 Feb 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

Data-driven airline passenger flight supply comparison with optimal transport theory

Qian Liu (刘倩),*† Paul Rochet*, Chantal Roucolle*

February 11, 2026

Abstract

This study presents an optimal transport (OT) framework to analyze airline supply in a pairwise manner, with an emphasis on strategic similarity and competitive intensity. By representing route-level supply as probability distributions and defining adapted cost structures, we develop a flexible approach to build comparative indices between airlines. We discuss three applications: route strategic positioning, competitive intensity, and operational cost. Applied to European domestic airlines, these indices reveal that mainline carriers exhibit greater strategic similarity due to their extensive networks, and low-cost carriers experience higher competitive intensity because of the geographic proximity of their routes. Full-service and low-cost carriers differ notably in operational strategies regarding frequency allocation and flight length choices. Compared with traditional multimarket contact (MMC) measures, the OT-based approach provides an interpretable, continuous metric that incorporates partial market substitutability through airport catchment areas. The framework can potentially be generalized to other transport systems, providing a robust methodological tool for measuring transporter competition, network structural changes, and strategic similarity across spatially distributed supply networks.

Keywords: Market competition measurement; Pairwise analysis; Competitive intensity; Strategic similarity; Multimarket contact; Business model convergence.

1 Introduction

Following the deregulation of air transport markets, first introduced in the United States in 1978, airlines have operated in increasingly competitive environments with greater strategic autonomy. This structural shift has led carriers to adopt heterogeneous network structures and capacity deployment strategies in order to attract passengers and maintain profitability. Because airlines primarily compete by adjusting the quantity of flights they supply, their competitive behavior can be well modeled by a Cournot competition framework (Brander and Zhang, 1990). Airlines differ not only in the routes they serve, but also in how they allocate frequency and seat capacity over time across their networks. These supply decisions reflect underlying business models, cost structures, and operational constraints, and they shape the intensity and nature of competition between carriers. In turn, these market dynamics influence fares, consumer welfare, and environmental outcomes (Schipper et al., 2003; Oum et al., 2005; Williams, 2022; Sun et al., 2024).

Accurate measurement of competitive intensity among airlines is essential for informing policy design, and supporting strategic network planning. Existing transport research has typically assessed airline supply similarity through network topology (Alderighi et al., 2007; Lordan et al., 2014), multimarket contact (MMC) measures (Zou et al., 2012; Liu, 2026), or classifications based on business models (Lohmann and Koo, 2013). Furthermore, traditional quantitative indicators of airline competitive intensity usually rely on route overlap measures, such as MMC, or on concentration metrics derived from market shares, including the Herfindahl–Hirschman Index (HHI) and concentration ratios (CR) (Lijesen et al., 2002; Kwieciński, 2017). While informative, these approaches remain insensitive to route characteristics such as spatial substitutability and seasonal demand similarity, which are non-negligible for capturing route heterogeneity in airline network analysis. Consequently, they provide a

*Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, France

†qian.liu@enac.fr

limited representation of the complex and multidimensional nature of competition in contemporary air transport markets. This motivates, therefore, the need for more flexible and accurate measures of strategic similarity and competitive intensity.

This paper proposes a novel OT-based framework to measure pairwise airline supply relationships and to capture strategic similarity, spatial multimarket competitive intensity, and operational cost differences. Optimal Transport (OT) theory (Villani et al., 2009) provides a rigorous mathematical framework for comparing probability distributions under flexible and interpretable cost structures. Originally developed to address mass transportation problems, OT has since been widely applied in economics, machine learning, and spatial analysis to quantify dissimilarities between complex distributions. For instance, in urban transport, OT has been used to design and optimize network structures (Leite and De Bacco, 2024); in applied economics, it has supported the solution of matching and allocation problems (Galichon, 2016); and in machine learning, it has enabled new approaches to diverse problems (Montesuma et al., 2024). Its capacity to incorporate domain-specific cost functions makes OT particularly suitable for modeling airline competition, where the notion of “distance” between strategies can reflect seasonal, spatial, or operational differences.

In our OT framework, airlines are represented by distributions of capacity or frequency across routes, while supply dissimilarity is quantified through three cost functions defined on the route space: (i) seasonal demand similarity, (ii) geographic route proximity, and (iii) route length differences. These cost structures enable a unified and multidimensional assessment of airline pairwise relationships. Each of them leads to the building of the specific OT based index.

Using European domestic flight data, we illustrate how the proposed OT-based indices provide meaningful insights into the top 20 airlines in terms of total capacity or frequency. First, the OT-based index I_1 effectively measures strategic similarity and highlights hybridization between low-cost carriers (LCCs) and full-service carriers (FSCs, or also referred to as mainline airlines) (Graham and Vowles, 2006; Daft and Albers, 2015; Magdalena and Bouzaima, 2021). Second, the OT-based index I_2 accounts for partial substitutability across routes and provides a continuous representation of multimarket competitive intensity. Third, I_3 captures key operational cost differences among airlines and establishes a foundation for future research on airline sustainability.

The contributions of this study are threefold. First, it proposes a flexible, data-driven index based on OT theory for analyzing airline pairwise supply relationships. Second, it expands airline competition analysis beyond route overlap by accounting for seasonal, spatial, and operational cost structures. Third, it offers empirical insights into the European domestic airline market. More broadly, the framework is transferable to other transport modes characterized by spatially and seasonally distributed supply.

The remainder of the paper is organized as follows. Section 2 reviews relevant literature and identifies the research gap. Section 3 introduces the general OT framework and defines three cost functions, each tailored to a specific research question. Sections 5, 6, and 7 present the empirical results corresponding to the three cost functions, including descriptions of the data and interpretation of the calculated index values. Section 8 provides a discussion of the findings. Finally, Section 9 concludes the paper and outlines directions for future research.

2 Related work

Structural and dynamic models of airline competition Structural models of airline competition are typically classified into price-based and quantity-based approaches. Brander and Zhang (1990) show that airline competition is better characterized by a Cournot framework, a conclusion later confirmed by Brander and Zhang (1993). Motivated by this evidence, we focus on quantity competition, which reflects airlines’ primary supply decisions and subsequently shapes pricing strategies. More recent structural work allows for simultaneous competition in prices and quantities (Grauberger and Kimms, 2016).

Subsequent research extends these frameworks to dynamic settings, such as using game-theoretic, econometric, or microeconomic models to analyze strategic interactions among airlines, particularly entry and exit decisions. For instance, Aguirregabiria and Ho (2010) develop an empirical dynamic game of airline network competition that incorporates hub-and-spoke entry deterrence through airline-specific profit functions. Building on this approach, Aguirregabiria and Ho (2012) propose a dynamic network competition model and introduce methods to reduce the dimensionality of the state space in

dynamic games and to address multiple equilibria in counterfactual analyses. More recently, [Ciliberto et al. \(2021\)](#) provide an econometric framework for estimating a game of simultaneous entry and pricing decisions in oligopolistic markets, allowing for correlation among unobserved fixed costs, marginal costs, and demand shocks. [Yuan and Barwick \(2024\)](#) develop an empirical framework for network competition based on a microeconomic model that estimates demand and cost functions. In their setting, airlines make market entry and flight frequency decisions in the first stage, followed by price competition to attract passengers in the second stage.

Despite their methodological rigor, structural and dynamic models face practical limitations. They are computationally intensive and rely on strong assumptions about both market conditions and airlines' strategic behavior. These assumptions restrict their scalability in complex market environments, particularly when analyzing interactions among many firms simultaneously. As a result, although these models are well-suited for counterfactual and policy analysis in specified settings, they are less effective at providing a broad, descriptive snapshot of the competitive landscape of the aviation market at a specific point in time, which can be useful for time-updated supervision.

Static measures: network structure, multimarket contact, and market concentration

When the objective is measuring rather than analysing strategic interactions, static models can be computed quickly and therefore provide a descriptive snapshot of the airline network structure. Research adopting a network perspective to characterize airline operation typically relies on topological and graph-based measures to quantify network complexity and connectivity, offering descriptive indicators of individual airlines' network structures ([Alderighi et al., 2007](#); [Sun and Wandelt, 2014](#); [Lordan et al., 2016](#)). More broadly, complex network theory has been widely applied in airline network analysis ([Lordan et al., 2014](#); [Verma et al., 2014](#); [Du et al., 2016](#)), contributing substantially to the study of network robustness and resilience. While these approaches emphasize the structural similarity of networks, they typically overlook the relationships between the networks of different airlines, such as overlapping routes or shared connections, focusing instead on each network individually.

Research on multimarket contact (MMC) explicitly accounts for airline interdependence across overlapping routes, showing that repeated interactions across markets can soften competition and may lead to collusion or mutual forbearance ([Ma et al., 2020](#); [Zou et al., 2012](#)). However, no widely accepted measure of multimarket contact exists. In general, MMC measures fall into two broad categories: count-based measures and probabilistic measures. Recent empirical work refines the count-based measures by weighting routes according to traffic frequencies ([Liu, 2026](#)). Nevertheless, most MMC measures implicitly assume homogeneous substitutability across markets, treating all routes as equivalent. For example, they ignore differences in demand across routes, which limits their ability to capture operational and market-specific heterogeneity.

An alternative static market analysis relies on concentration indices, such as the Herfindahl–Hirschman Index (HHI) and concentration ratios (CR), to measure competitive pressure on the market ([Lijesen et al., 2002](#); [Kwieciński, 2017](#)). Using panel data, empirical studies document the evolution of market concentration over time and examine how deregulation, competition, and mergers reshape industry structures ([Goetz, 2002](#); [Oliveira and Oliveira, 2018](#)). In the context of airline mergers, market concentration level and multimarket contact are central analytical dimensions, as their changes may expand firms' ability to exercise market power and affect industry performance ([Mazzeo, 2003](#); [Prince and Simon, 2009](#); [Zhang and Round, 2009](#); [Baumgarten et al., 2014](#)). By tracking different markets and time periods, the concentration index can provide a macro-level assessment of market competition. While concentration indices have regulatory significance, they reduce complex competitive interactions to a single market index and fail to capture the interdependencies between airline networks.

Airline business models Airlines are traditionally classified by business model, most commonly distinguishing between low-cost carriers (LCCs) and full-service carriers (FSCs). In practice, however, this dichotomy has become increasingly inadequate. Many airlines have adopted hybrid business models that combine elements of both LCCs and FSCs in order to broaden their customer base and remain competitive in more contested markets ([Lohmann and Koo, 2013](#)). This progressive hybridization of airline strategies ([Daft and Albers, 2015](#); [Magdalena and Bouzaima, 2021](#)), together with the strategic response of full-service carriers through the establishment of low-cost subsidiaries ([Wang et al., 2025](#)), has further blurred the boundaries between traditional categories. Consequently, static, label-based classifications have a declining capacity to reflect airline operational similarity, highlighting the need

for more flexible and accurate approaches to the analysis of airline clusters.

Optimal transport and distributional similarity in transportation Optimal Transport (OT) theory has become a principled framework for comparing complex distributions under flexible and specific cost structures (Villani et al., 2009; Peyré et al., 2019). For instance, in economics, OT has been applied to problems such as labor market matching, housing price formation, and marriage markets (Galichon, 2016). In transportation and mobility research, recent studies use OT to compare spatial demand patterns, network configurations, and mobility distributions (Li et al., 2010; Leite and De Bacco, 2024).

The key strength of OT lies in its ability to jointly capture distributional magnitudes and the geometry of the underlying space over which mass is reallocated. This feature makes OT particularly well-suited to airline supply analysis, where competition is driven by route-specific characteristics, such as seasonal demand patterns, operating conditions, and geographic positioning. Although Li et al. (2010) apply this framework to airline network planning, the use of OT for analyzing airline competition remains limited.

Research gap Taken together, the literature reveals a clear methodological gap in the measurement of airline competition. Structural and dynamic models offer rich insights into strategic behavior but are computationally demanding and not suited for providing timely, descriptive assessments of competitive conditions in complex markets. Static approaches, including network metrics, multimarket contact measures, concentration indices, and business-model classifications, are easier to implement but rely on restrictive assumptions. They either abstract from inter-airline network interdependencies, impose homogeneous substitutability across routes, compress competition into single-market aggregates, or depend on categorical labels that no longer reflect operational realities.

While Optimal Transport provides a flexible and theoretically grounded tool for comparing heterogeneous distributions in a structured space, its application has largely been limited to network planning or aggregate mobility analysis, rather than the measurement of strategic similarity or competitive intensity among airlines. As a result, there is a lack of empirical frameworks that combine the descriptive tractability of static measures with the ability to capture route-level heterogeneity, network interdependence, and continuous variation in airline strategies. This gap motivates the development of a more accurate approach to measuring airline competitive intensity.

3 Methodology: OT and three applications

In this section, we first introduce the general framework of optimal transport (OT) theory along with an intuitive explanation. We then formulate three distinct cost functions within this framework, each tailored to a specific analytical purpose. The cost functions are data-driven, which can be derived directly from observations without developing complex models, thereby ensuring transparency and realism of the analysis. It is worth noting that, in our illustrative examples, a route market is defined by an airport pair. The proposed framework can be readily extended to alternative market definitions, such as city-pair markets, or applied to analyze pairwise relationships among other transporters who serve routes, including maritime, rail, and urban transport operators.

3.1 General OT framework

Optimal transport (OT) provides a mathematical framework for comparing probability measures by quantifying the minimal “effort” required to transform one distribution into another under a defined cost. Originally formulated by Monge (1781) and later relaxed by Kantorovich (1942), OT characterizes this transformation through transport plans or couplings that redistribute mass while minimizing the total transportation cost. Especially, when the cost function is a metric, the OT cost defines a meaningful distance between distributions, known as the Kantorovich–Rubinstein metric or Wasserstein distance, which captures both magnitude and spatial arrangement of discrepancies between distributions. Owing to its solid theoretical foundation and versatility, OT has a wide range of applications across economics, physics, machine learning, and data analysis.

Given two probability measures μ, ν defined on a space X , Kantorovich’s formulation of OT corresponds to finding a coupling T that minimizes a total transport cost from μ to ν . More precisely, let $\Gamma(\mu, \nu)$ denote the set of probability distributions on $X \times X$ with marginals μ, ν , and $c : X \times X \rightarrow [0; +\infty)$ a cost function, the OT problem can be expressed as minimizing

$$T \mapsto \int_{X \times X} c(x, x') dT(x, x') \quad \text{subject to} \quad T \in \Gamma(\mu, \nu). \quad (1)$$

If X is a finite space, a distribution μ on X can be identified with a non-negative valued vector $\mu = (\mu_x)_{x \in X}$ such that $\sum_{x \in X} \mu_x = 1$, and the OT problem can be formalized as a discrete optimization program:

$$\min_{T \in \Gamma(\mu, \nu)} \sum_{x, x' \in X} c(x, x') T(x, x'). \quad (2)$$

3.2 Airline comparisons

A central idea in this work is to identify an airline’s supply as a distribution of either seat capacity or flight frequency across all routes. Consequently, we define the following OT framework. Let R denote a set of routes and consider an airline operating on R which allocates a supply $f_r \geq 0$ to each route $r \in R$. Letting $f = \sum_{r \in R} f_r$ denote the total supply of the airline, we consider its associated supply distribution $\mu = (\mu_r)_{r \in R}$ where $\mu_r = f_r/f$ is the normalized supply of the airline for the route $r \in R$. For the cost function $c : R \times R \rightarrow [0; +\infty)$ on the route space, the similarity between two airlines with associated distributions μ, ν relative to $c(\cdot, \cdot)$ is defined by the total transport cost

$$I(\mu, \nu) := \min_{T \in \Gamma(\mu, \nu)} \sum_{r, r' \in R} c(r, r') T(r, r')$$

where $\Gamma(\mu, \nu)$ denotes the set of all transport plans $T : R \times R \rightarrow [0; +\infty)$ with marginals μ, ν , satisfying the conditions

$$\forall r \in R, \sum_{r' \in R} T(r, r') = \mu_r \quad \text{and} \quad \forall r' \in R, \sum_{r \in R} T(r, r') = \nu_{r'}.$$

This approach provides a general tool to compare airline supplies from various perspectives, with the focus of the comparison encapsulated in the cost function $c(\cdot, \cdot)$. We discuss three examples of applications, each associated with a particular choice of cost function, to assess the similarity in commercial strategy, competitive intensity and operational cost between airlines in the European flight network.

We either use capacity-based distributions for market and demand-oriented aspects (Sections 5 and 6) or frequency-based distributions when focusing on structural differences in business models (Section 7). In Section 5, we analyze how airlines respond similarly (or differently) to route-level passenger demand, which is supported by annual capacity variation profiles. Section 6 discusses a new way to measure airline multimarket competitive intensity, by treating geographically closer routes as likely substitutes and therefore imposing stronger competitive pressure on the airlines operating them. In these contexts, comparing offered seat capacity is appropriate because it reflects the actual supply made available to meet market demand over a given time period. Seat capacity equals the number of flights multiplied by the seats per aircraft, thus capturing both the frequency of service and the type of aircraft chosen. By contrast, in Section 7, frequency is used to represent airline supply because it highlights a key distinction between LCCs and FSCs. Traditional LCCs generally compete through point-to-point operations, characterized by fewer frequencies, higher load factors, and fleets with limited aircraft variation. FSCs, however, rely on higher frequencies supported by hub-and-spoke networks and business-oriented demand (Rozenberg et al., 2014). Using the OT framework with frequency distributions, therefore, reveals these structural differences and offers insights into distinguishing LCCs from FSCs.

Within this framework, the required data consist of seat capacities or flight frequencies for the airlines over a given period (to construct μ and ν), along with additional information needed to define the cost function, which must be tailored to the specific analytical objective of using the OT model. Below, we introduce three types of cost functions, each emphasizing a different aspect of airline operation, and we interpret the resulting OT costs. The assumptions underlying each cost matrix are presented together with the corresponding cost function.

3.2.1 Route strategic positioning

It is commonly agreed that in European air markets, competition generally emphasizes frequency on business-oriented routes, usually operated by mainlines, where travelers value multiple daily departures for flexibility. Low-cost or leisure-focused airlines typically operate a single aircraft type and offer a single cabin class (economy) to simplify operations, reduce costs, and increase seat occupancy, though they usually provide fewer flight frequencies. Slot constraints at major airports and intermodal competition from high-speed rail also influence these strategies, occasionally prompting carriers to balance frequency and aircraft capacity to supply their total capacity. Therefore, to examine strategic similarity in route-level supply, which is grounded in airport-pair market demand, we focus on the total capacity offered by airlines, represented by the distributions μ and ν introduced above.

Within this framework, we define the cost of transporting one unit of mass from one route to another as the annual dissimilarity in market demand between the two routes. This cost is an informational measure reflecting how distinct the demand conditions are across markets, and thus how costly it is, from a strategic perspective, for an airline to reallocate capacity between them. Conceptually, we proxy market demand of one route by the total capacity supplied by all airlines on this route. Under the assumption that the load factor is on average identical for airlines and constant over the year, weekly observed capacities might reflect weekly demand.

The dissimilarity in market demand is measured as follows. Each route r is characterized by the distribution of its normalized weekly capacity over the year. Formally, the weekly total (i.e., all airlines combined) seat capacity of a route $r \in R$ is denoted by $S^r = (S_n^r)_{n=1, \dots, 52}$ where n indicates the week number. For two routes $r, r' \in R$, we compute the Wasserstein distance, say $W(S^r, S^{r'})$ between S^r and $S^{r'}$ associated to the distance $|n - n'|$ between two weeks $n, n' \in \{1, \dots, 52\}$. The idea behind this choice of cost function is to convey that two routes that are similar from an airline strategy point of view are expected to have comparable seat distributions over the year.

Building on this route-level metric, we adopt a hierarchical optimal transport formulation to characterize the dissimilarity in airline supply strategies arising from their emphasis on different route demand profiles. Specifically, we define an index based on the total OT cost using the Wasserstein distance between weekly seat capacities as the cost function:

$$I_1(\mu, \nu) \stackrel{\text{def}}{=} \min_{T \in \Gamma(\mu, \nu)} \sum_{r, r' \in R} W(S^r, S^{r'}) T(r, r').$$

Here, $W(\cdot, \cdot)$ captures seasonal demand dissimilarities, while μ and ν denote the normalized capacity distributions of the two airlines on the European domestic market. This nested construction allows us to compare airlines at the strategy level while explicitly accounting for heterogeneity in route-level demand patterns.

3.2.2 Competitive intensity

Two routes are considered substitutable when they are geographically close and offer comparable generalized travel conditions (price, travel time) to passengers. For instance, Paris Charles de Gaulle (CDG) airport and Paris Orly (ORY) airport may serve as substitutes for certain travelers, as their catchment areas are the same. Similarly, destinations such as Toulouse and Carcassonne can be considered as potential substitutes when efficient rail connections link the two cities within a reasonable travel time. In this situation, if the airfare to Carcassonne is low enough to offset the extra ground travel time, the alternative will be acceptable for passengers planning to travel to Toulouse.

In the study of multimarket contact (MMC) and multimarket competition, traditional measures typically rely on simple market overlap counts. While informative, this approach neglects competition arising from proximate or adjacent markets that may function as effective substitutes. To address this limitation, we next construct a geography-based measure of route substitutability.

Let $d(\cdot, \cdot)$ be the geographical distance between airports, with route r identified by the unordered pair $\{O, D\}$ as its origin and destination airports. We define the mismatch distance between routes $r = \{O, D\}$ and $r' = \{O', D'\}$ as

$$d_2(r, r') \stackrel{\text{def}}{=} \min \left\{ d(O, O') + d(D, D'), d(O, D') + d(D, O') \right\}.$$

We take the minimum of the summed distances because both access and egress journeys influence passenger airport choice (Kouwenhoven, 2008; Hess, 2016; Allard and Moura, 2016). Moreover, $d_2(\cdot, \cdot)$ satisfies the properties of a metric (see Appendix A).

While the route distance d_2 captures geographic separation of routes, directly using it as a transport cost may overweight a small number of distant routes, even when two airlines primarily compete in geographically proximate markets. To mitigate this issue, we simply transform geographic distance into a ‘reject probability’ that measures the extent to which two routes are not mutually accepted as substitutes:

$$c_2(r, r') \stackrel{\text{def}}{=} 1 - \exp\left(-\frac{d_2(r, r')}{\lambda}\right), \quad \lambda > 0.$$

This implies that the two routes are less likely to be substitutes as the total mismatch distance d_2 increases. The model c_2 assumes identical access facilities for all airports, regardless of location. We then define the corresponding index:

$$I_2(\mu, \nu) \stackrel{\text{def}}{=} \min_{T \in \Gamma(\mu, \nu)} \sum_{r, r' \in R} c_2(r, r') T(r, r').$$

The transformation $c_2(\cdot, \cdot)$ maps geographic distance into the unit interval $[0, 1]$, where values close to zero indicate high spatial proximity. The exponential form ensures that sufficiently distant routes contribute negligibly to competitive interaction, reflecting the limited substitutability perceived by passengers. Because d_2 aggregates origin and destination distances, it may decay rapidly; the scale parameter λ moderates this effect. The choice of λ must be adjusted depending on the expert judgment and the specific focus of the analysis. For illustration, the exponential term $\exp(\cdot)$ can be interpreted as the probability that a passenger is willing to switch from one route to another, while $1 - \exp(\cdot)$ represents the corresponding cost or reluctance of such a transfer. For example, $\exp(\cdot) = 0.9$ indicates a high willingness to substitute, with only about 10% of passengers perceiving the transfer as too costly.

It can be understood that two airports are in competition if their catchment areas overlap (Lieshout, 2012). We assume identical access facilities and a catchment area of 100 km for all airports, which is a common practice in European airport studies (International, 2023). We propose, therefore, to set $\lambda = 50$, which provides a reasonable rate at which the reject probability c_2 diminishes with geographic distance. With $\lambda = 50$, if the combined origin–destination distance is $d_2 = 20$ km, we obtain $c_2 = 1 - \exp(-0.4) \approx 0.33$, which reflects the close geographic proximity of the two routes. By contrast, a value of $c_2 \approx 0.9$ corresponds to an aggregated distance d_2 of approximately 115 km, a range that remains feasible through rail connections within reasonable travel time. When $c_2 \approx 0.99$, the distance d_2 increases to around 230 km. Beyond this range, routes with combined origin–destination distances exceeding 230 km contribute negligibly to substitution, consistent with limited passenger willingness to consider such alternatives.

3.2.3 Operational cost

We next consider absolute route length differences as a simplistic measure of airline reallocation cost across routes. For a given demand, frequency choices are closely tied to aircraft deployment decisions: larger aircraft typically enable lower service frequencies, whereas smaller aircraft are associated with higher-frequency operations. From a business model perspective, operating a small number of flights with large aircraft constitutes a fundamentally different strategy from operating many flights with smaller aircraft, even when total capacity is identical. Accordingly, using frequency rather than capacity mitigates confounding effects related to fleet size, ensuring that the OT measure primarily captures differences in scheduling patterns rather than scale effects.

In this context, absolute differences in route length serve as a proxy for the operational cost of reallocating flights between routes, capturing variations in factors such as fuel cost and en-route charges. When considered alongside frequency distributions, these route length differences offer a typical representation of heterogeneity in airline service quality and operating costs across markets, reflecting the strategic choices embedded in airline business models and cost structures.

Formally, we define the route-level cost as

$$d_3(r, r') = |\text{len}(r) - \text{len}(r')|,$$

where $\text{len}(r)$ denotes the length of route r in km. The corresponding OT-based dissimilarity index between airlines is then given by

$$I_3(\mu, \nu) \stackrel{\text{def}}{=} \min_{T \in \Gamma(\mu, \nu)} \sum_{r, r' \in R} d_3(r, r') T(r, r').$$

4 Data

For the first two applications, calculating I_1 and I_2 , we extract airline passenger flight schedules from OAG. The OAG dataset provides detailed flight schedule information, including carrier codes; market definitions at the airport, city, and country levels; aircraft codes; seat capacity; service frequency; and timestamps. We focus on the European domestic market, defined by flights whose origins and destinations are both located within the 28 countries (Table 1) that simultaneously belong to the European Economic Area (EEA) and the Schengen Area. Note that Liechtenstein has no airport, but is included for completeness. The inclusion of the EEA ensures economic comparability across member states, while the Schengen area facilitates passenger movement, allowing a consistent analysis of airline route networks.

| | | | |
|------------------|-----------------|---------------|----------------------|
| AT – Austria | BE – Belgium | BG – Bulgaria | CZ – Czech Republic |
| DE – Germany | DK – Denmark | EE – Estonia | ES – Spain |
| FI – Finland | FR – France | GR – Greece | HR – Croatia |
| HU – Hungary | IS – Iceland | IT – Italy | LI – Liechtenstein * |
| LT – Lithuania | LU – Luxembourg | LV – Latvia | MT – Malta |
| NL – Netherlands | NO – Norway | PL – Poland | PT – Portugal |
| RO – Romania | SE – Sweden | SI – Slovenia | SK – Slovakia |

Table 1: European domestic market countries and their ISO 3166-1 alpha-2 country codes.

We focus on the top 20 airlines for 2024 in terms of annual total seat capacity offered. Table 2 lists these airlines together with their OAG carrier-type labels (Mainline or Low-cost), IATA codes, and headquarters country codes.

| Low-cost (L) | Mainline (M) |
|--|------------------------------------|
| D8 – Norwegian Air International (IE, Ireland) | A3 – Aegean Airlines (GR) |
| DY – Norwegian Air Shuttle (NO) | AF – Air France (FR) |
| EW – Eurowings (DE) | AY – Finnair (FI) |
| FR – Ryanair (IE) | AZ – ITA Airways (IT) |
| U2 – easyJet (GB, United Kingdom) | IB – Iberia (ES) |
| V7 – Volotea (ES) | KL – KLM Royal Dutch Airlines (NL) |
| VY – Vueling (ES) | LH – Lufthansa (DE) |
| W4 – Wizz Air Malta (MT) | OS – Austrian Airlines (AT) |
| W6 – Wizz Air (HU) | SK – Scandinavian Airlines (SE) |
| | TP – TAP Air Portugal (PT) |
| | UX – Air Europa (ES) |

Table 2: Top 20 airlines by total annual capacity in 2024, with OAG carrier type labels, IATA airline codes, airline names, and headquarters country codes.

To focus on operationally relevant routes and reduce noise from irregular flights with less market demand, we retain only routes that meet the following criteria:

- Operated for at least four months within the year, to exclude highly irregular seasonal routes.
- Annual frequency of at least 52 flights, equivalent to an average of at least one flight per week across all airlines.

Applying these criteria, the 2024 data is reduced to 5,658 routes in total, with 1,312 filtered out, and 4,346 routes retained (77%).

For I_1 and I_2 , we set μ and ν as the normalized capacity distributions of two airlines across all European routes. The resulting OT measure I_1 captures dissimilarities in how airlines allocate service over time, reflecting differences in supply strategies: whether airlines prioritize similar types of routes characterized by comparable seasonal variations in capacity. The resulting measure I_2 captures the multimarket competitive intensity among airlines.

For the third application I_3 , we also use passenger flight data. Consistent with the previous applications, we restrict the analysis to European domestic routes within the countries listed in Table 1. The top 20 airlines ranked by total annual frequency in 2024 are reported in Table 3. We calculate I_3 using 2024 data. Unlike the settings previously presented for Sections 5 and 6, where airline supply strategies are represented by normalised capacity distributions, this application relies on normalised frequency distributions across routes. This choice is intentional, designed to differentiate between operational characteristics and total capacity effects. In the earlier demand-oriented applications, capacity is the relevant metric because airlines primarily compete in quantities within a Cournot competition framework, and capacity decisions primarily influence prices and consumer welfare. Under such conditions, a route operated infrequently with large aircraft and one operated frequently with smaller aircraft may be equivalent in capacity, yet they differ substantially in operating costs, fleet composition, and environmental impact. Focusing on frequency, therefore captures variations in scheduling intensity and operational design independently of aircraft size. In this framework, frequency serves as a proxy for supply quality while retaining information relevant to cost structures. Since aircraft choice is closely related to route distance, the cost between routes is specified as a function of their length difference. Consequently, the resulting optimal transport index I_3 measures the cost of transitioning between airline frequency deployment strategies, which can be interpreted as an operational and financial adjustment cost.

| Low-cost (L) | Mainline (M) |
|-----------------------------------|------------------------------------|
| DY – Norwegian Air Shuttle (NO) | A3 – Aegean Airlines (GR) |
| EW – Eurowings (DE) | AF – Air France (FR) |
| FR – Ryanair (IE) | AY – Finnair (FI) |
| U2 – easyJet (GB, United Kingdom) | AZ – ITA Airways (IT) |
| VY – Vueling (ES) | IB – Iberia (ES) |
| W4 – Wizz Air Malta (MT) | KL – KLM Royal Dutch Airlines (NL) |
| W6 – Wizz Air (HU) | LH – Lufthansa (DE) |
| | LO – LOT Polish Airlines (PL) |
| | NT – Binter Canarias (ES) |
| | OS – Austrian Airlines (AT) |
| | SK – Scandinavian Airlines (SE) |
| | TP – TAP Air Portugal (PT) |
| | WF – Widerøe (NO) |

Table 3: Top 20 airlines by total flight frequency in 2024, with OAG carrier type labels, IATA airline codes, airline names, and headquarters country codes.

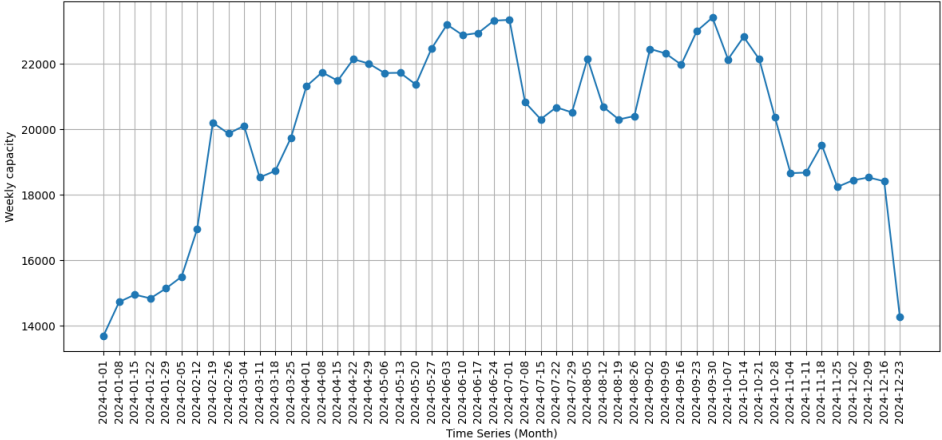
5 Application 1: Comparing route strategic positioning based on weekly demand

5.1 Cost matrix analysis

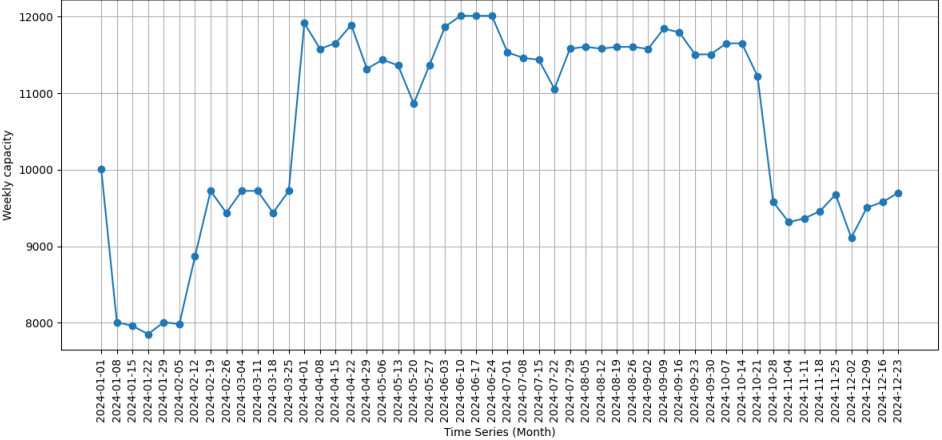
Route-level costs are computed based on the route total capacity weekly variation in 2024, and airline supply dissimilarities in terms of focusing on different route seasonal profiles are evaluated using I_1 . We analyze the cost matrix $W(r, r')$ defined in Section 3.2.1 from two aspects: (i) an illustrative example to understand route similarity, and (ii) the empirical distribution of the Wasserstein distance $W(r, r')$ across all route pairs.

Similar routes We first illustrate route similarity through an example of routes exhibiting highly similar seasonal capacity patterns. For any given route r , the most similar route can be identified

as route r' that minimises the Wasserstein distance $W(r, r')$. For instance, the route CDG–FLR is identified as the most similar to CDG–FRA, as it attains the minimum Wasserstein distance among all routes relative to CDG–FRA. This result is consistent with the visual evidence presented in Figure 1. The weekly total seat capacity offered by all carriers on both routes follows a remarkably similar seasonal evolution over the year. In particular, both routes exhibit comparable seasonal fluctuations and peak periods, suggesting that the Wasserstein distance effectively captures route similarities in demand-driven capacity adjustments rather than merely absolute traffic volumes.



(a) Weekly capacity evolution of route CDG–FRA.



(b) Weekly capacity evolution of route CDG–FLR.

Figure 1: Illustration: CDG–FLR is identified as the most similar route to CDG–FRA, as measured by the route Wasserstein distance.

Cost matrix Beyond the individual example, we analyze the global structure of the cost matrix. For the year 2024, a total of 4346^2 pairwise route Wasserstein distances are computed, and their empirical distribution is shown in the histogram in Figure 2.

The distribution is markedly right-skewed, indicating that only a limited fraction of route pairs exhibit low Wasserstein distances and therefore strong seasonal similarity in capacity patterns. Quantitatively, the median Wasserstein distance is 4.98, with the first and third quartiles located at 2.41 and 7.64, respectively. This implies that 50% of route pairs have a distance lying within a relatively wide interquartile range, highlighting substantial variability even among moderately similar routes. In contrast, the upper tail of the distribution extends to a maximum value of 39.08, reflecting pairs of routes with highly different seasonal patterns.

Overall, these results underline the pronounced heterogeneity of airline routes in terms of seasonality and market characteristics. The observed distribution confirms that the cost matrix is highly

discriminative: while a small subset of routes shares comparable seasonal dynamics, the majority differ significantly, making the Wasserstein distance a suitable metric for capturing meaningful structural differences across routes.

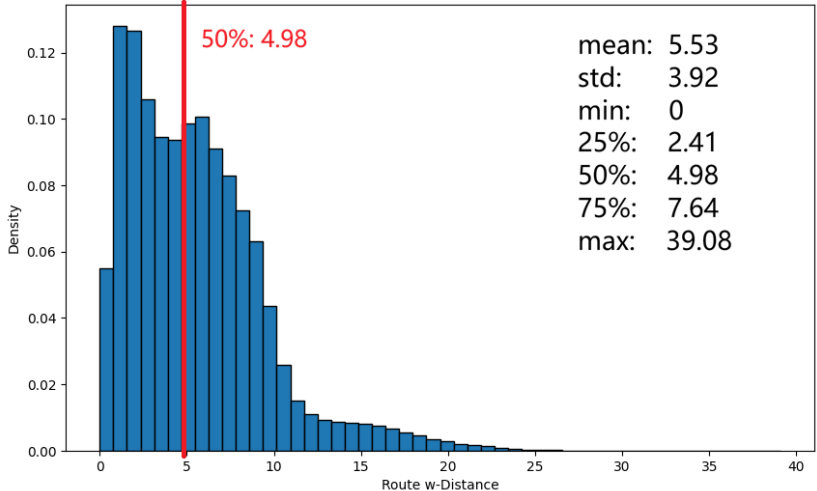


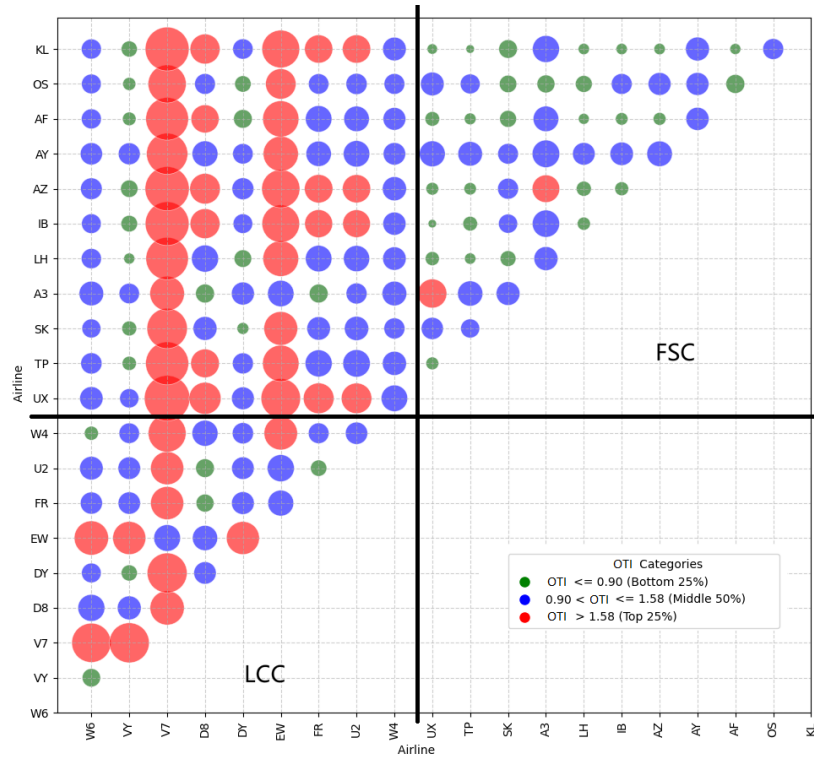
Figure 2: Histogram of the route cost matrix. Each entry corresponds to the Wasserstein distance between a pair of routes, computed using the ℓ_1 distance on week indices.

5.2 Top 20 airlines market overview

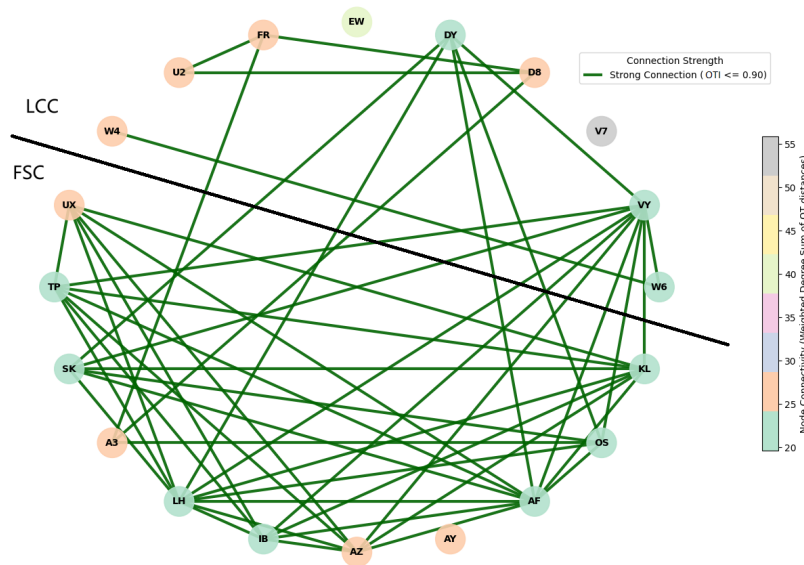
We analyze the top 20 airlines (Table 2) with the largest total seat capacity across the selected 4,346 routes in 2024. Pairwise airline dissimilarities are measured using I_1 , and the resulting values are visualised in Figure 3. Specifically, Figure 3a presents the OT index in a bubble-matrix representation: the size of each bubble is proportional to the magnitude of the OT index, such that smaller bubbles indicate a smaller difference (therefore greater similarity) between two airlines, while larger bubbles reflect stronger strategic dissimilarity. To facilitate interpretation, the OT index values are also categorised into three intervals based on their empirical distribution. Green bubbles correspond to values in the first quartile (below the 25th percentile), representing the most similar airline pairs. Blue bubbles denote values between the 25th and 75th percentiles, capturing intermediate levels of similarity, while red bubbles represent values above the 75th percentile, corresponding to the most dissimilar airline pairs. Figure 3b provides a complementary network-based representation, focusing exclusively on airline pairs whose OT index values fall within the first quartile. In this graph, nodes represent airlines and edges indicate strong strategic similarity as measured by low OT index values. The color intensity of each node reflects the sum of its edge weights as the weighted degree centrality, that is, the cumulative I_1 values linking the airline to all others.

Several structural patterns emerge from these visualizations, revealing insights into airline network strategies:

1. **Subgroup differentiation within the LCC segment:** The LCC segment exhibits identifiable subgroups. Norwegian (DY) and Vueling (VY) align closer to FSCs, showing a hybrid strategy that blends low-cost operations with relatively extensive networks. Another cluster includes easyJet (U2), Ryanair (FR), and Norwegian Air International (D8), which exhibit similar capacity allocation patterns across routes with comparable seasonal demand fluctuations, likely reflecting a strategic focus on high-demand short- and medium-haul markets in Europe. A separate cluster, consisting of Wizz Air (W6) and Wizz Air Malta (W4), demonstrates a regional concentration in overlapping regional markets. Meanwhile, airlines such as Eurowings (EW) and Volotea (V7) appear more isolated in the graph of Figure 3b, reflecting special route selection strategies that differentiate them from the broader LCC patterns. This is further confirmed by Figure 3a, where V7 and EW display high I_1 values relative to most other airlines.



(a) I_1 indices of airlines represented by bubble size and categorical colors. The matrix is symmetric, such that, for example, the KL-V7 entry is equal to the V7-KL entry.



(b) Network representation of airline pairs within the lowest 25% of I_1 values.

Figure 3: Illustration of I_1 values among the top 20 airlines, computed using a cost function based on route-level Wasserstein distances of annual capacity variation.

2. **Higher connectivity among FSCs versus LCCs:** In Figure 3b, FSCs exhibit a higher degree of connectivity than LCCs, reflecting greater similarity in their route portfolio structures. This pattern arises from their extensive hub-and-spoke networks, which connect multiple geographic markets with overlapping frequencies and various route demand profiles. Such a design produces high airline supply similarity, as FSCs tend to follow broadly comparable strategies across regions. By contrast, LCCs usually focus on select high-demand or leisure-oriented routes, introducing heterogeneity and weaker connections among them. Notably, some LCCs, such as Vueling (VY) and Norwegian (DY), deviate from typical segment patterns, displaying connectivity closer to FSCs.
3. **Interpretation of node coloring in Figure 3b:** Nodes are colored based on the sum of the I_1 values computed across *all* airline pairs. For a given airline (e.g., KL), the node color corresponds to the sum of the I_1 values in the KL row of the matrix shown in Figure 3a, irrespective of whether those pairwise relationships are displayed as edges in the network graph. While the network visualization only retains connections corresponding to the lowest 25% of I_1 values, the node coloring reflects the airline’s overall similarity profile relative to all other carriers. Consequently, nodes associated with lower total weighted degree centrality (e.g., nodes in green: TP, SK, LH, IB, etc.) represent airlines that are, on average, more similar to their peers, highlighting structurally central players whose network strategies closely align with multiple competitors. Conversely, nodes with higher total I_1 values (e.g., V7 and EW) correspond to airlines with more distinctive route portfolio structures, emphasizing strategic differentiation. Node colors therefore provide a complementary perspective to the displayed connections, capturing global similarity patterns that extend beyond the subset of links shown in the graph.

The main contributions of this application are twofold. First, it proposes the use of the Wasserstein distance between route-level capacity distributions as a measure of market demand differentiation across routes. Second, it introduces a nested OT framework to quantify strategic dissimilarity between airlines. Together, these contributions provide a novel quantitative perspective for analyzing airline supply strategy similarity and the hybridization of airline business models, thereby helping to bridge traditional distinctions between mainline and low-cost carriers.

6 Application 2: Comparing competitive intensity based on route geographic proximity

This section extends the analysis of airline supply similarity by incorporating the spatial dimension of route networks. Whereas Application 1 focuses on seasonal capacity patterns, this application assesses the geographical proximity of airlines’ operating routes. By embedding spatial distance into the OT framework, we can quantify multimarket competition in physically close markets, even if there is no overlap on the same routes.

6.1 Cost matrix analysis

To incorporate route spatial proximity into the OT framework, we construct a cost matrix based on geographical distance. For any two routes, the cost d_2 is defined as the minimum great-circle distance (GCD) between the sum of their respective origins and destinations. As a result, geographically close routes, even if they do not share identical endpoints, incur a low transportation cost, reflecting their potential substitutability from a market perspective. By construction, the cost function d_2 satisfies the properties of a metric (see Appendix A).

The empirical distribution of d_2 is shown in Figure 4. The distances span a wide range, from very short separations of 2 km to values exceeding 11,000 km, with a median of about 2,431 km and an average of 2,709 km. This indicates that a large share of European routes are geographically far apart, highlighting the substantial spatial dispersion of airline networks across the continent.

To translate geographical distance into an economically meaningful proximity cost, we apply a transformation to get the proximity cost $c_2 \in [0, 1]$ defined in Section 3.2.2. The role of λ is illustrated in Figure 5. The higher the λ value, the smoother the cost distribution, and the greater the spatial tolerance range for routes to be considered alternatives. For example, with $c_2 = 0.99$ fixed, setting

$\lambda = 50$ implies that routes separated by less than 230 km ($d_2 < 230$) are treated as substitutable, setting $\lambda = 200$ for a separation of 921 km, whereas $\lambda = 500$ considers routes within 2,302 km as comparable. Consequently, λ acts as a policy-relevant tuning parameter that determines the spatial granularity of competition: smaller values emphasize route overlap, while larger values capture broader regional competition.

Unsurprisingly, the rejection probability c_2 equals 1 for nearly all route pairs. Only 11,626 out of 9,441,685 pairs have c_2 values lower than 0.99, accounting for roughly 0.12% of all possible combinations. This pattern reflects the substantial geographical separation of routes across Europe, consistent with Figure 4.

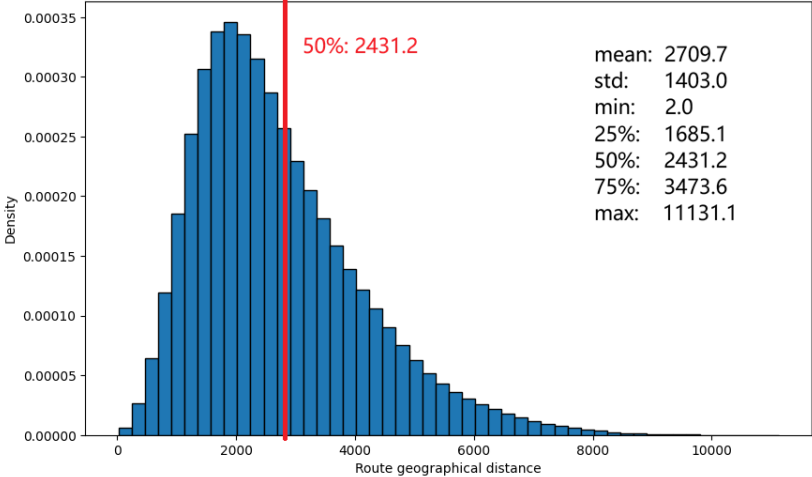


Figure 4: Histogram of great-circle distances between routes in 2024.

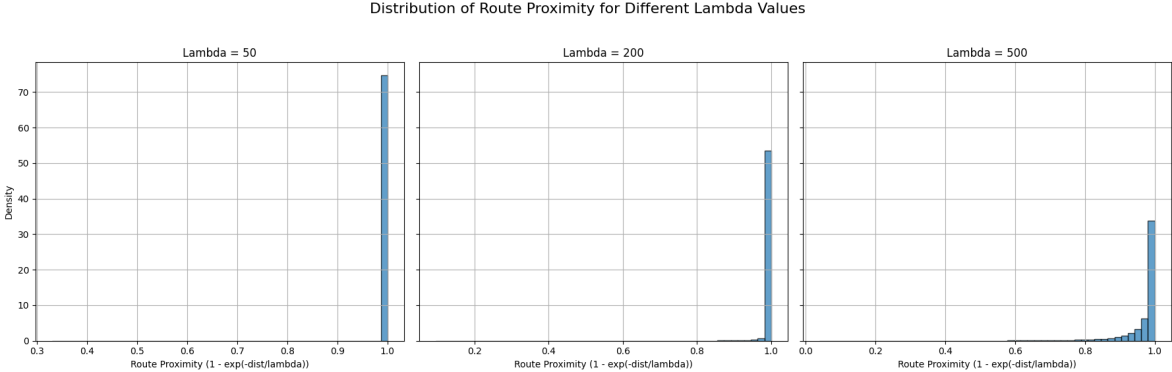


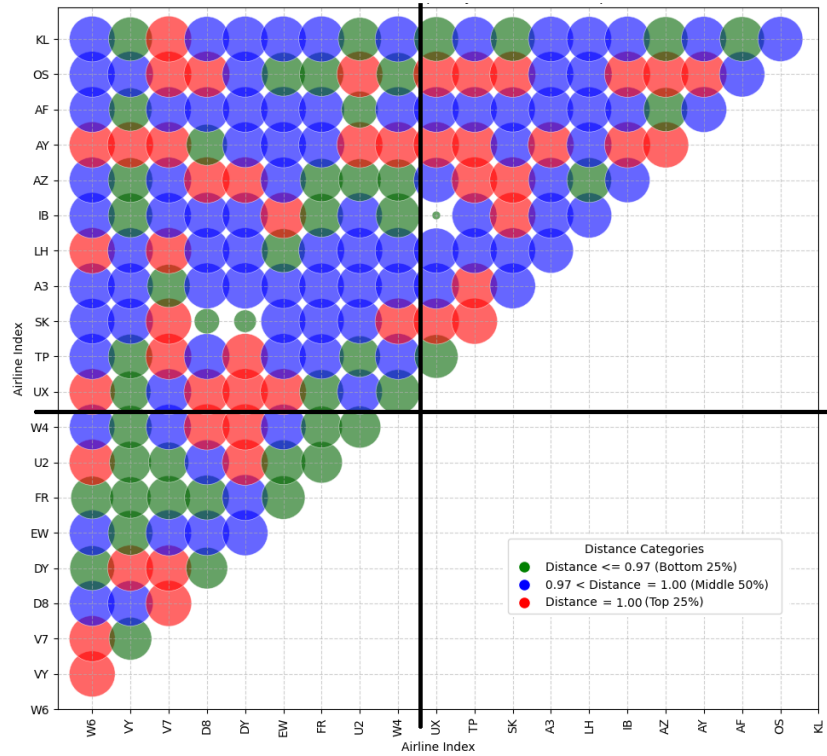
Figure 5: Illustration of how variations in λ affect the distribution of the route proximity cost c_2 .

6.2 Top 20 airlines market overview

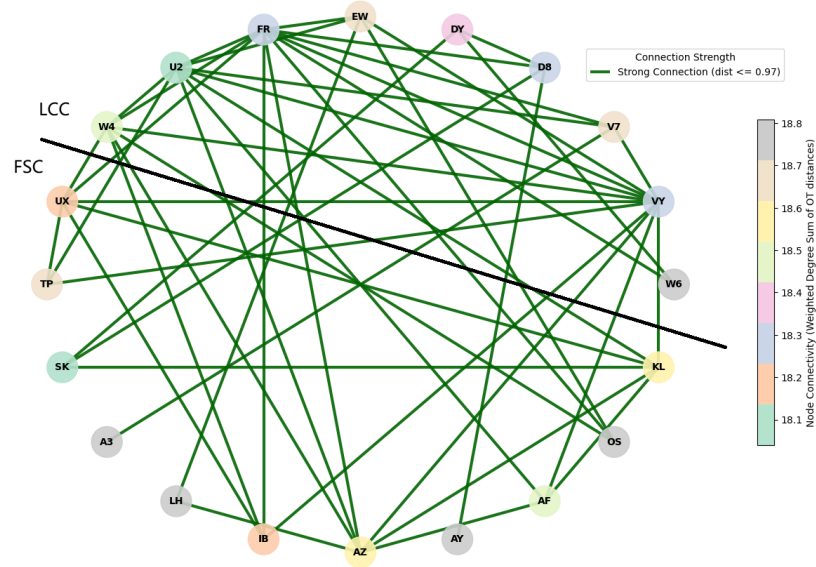
To translate airline network geographical distances into a measure of multimarket competition intensity, we compute the proximity-based index I_2 . The resulting values are shown in Figure 6a, using the same interpretation as Figure 3a. Lower values of I_2 indicate that two airlines operate on geographically close routes, whereas higher values suggest weaker spatial overlap and less direct competition. The corresponding network representation, retaining only the lowest 25% of I_2 values, is displayed in Figure 6b, highlighting the most intense spatial competition relationships.

Several patterns emerge from the proximity-based multimarket competition analysis:

- Identification of primary competitors:** Figure 6a allows the identification of an airline’s main competitors by examining its smallest I_2 values. For example, Iberia’s (IB) primary competitor is Air Europa (UX), largely because they operate overlapping regional networks within



(a) I_2 indices of airlines represented by bubble size and categorical colors.



(b) Network representation of airline pairs within the lowest 25% of I_2 values.

Figure 6: Illustration of I_2 values among the top 20 airlines, capturing airline multimarket competition intensity.

Spain and share the same hub in Madrid. Low values indicate pairs of airlines with geographically proximate route networks, reflecting a high likelihood of spatial competition. This method captures competitive relationships more accurately than simple route overlap, accounting for potential substitutability of routes.

- **Competitive clustering of low-cost carriers:** Figure 6b shows that LCCs form denser competitive clusters than FSCs. In fact, LCCs concentrate capacity on a limited set of high-demand, profitable, and leisure routes. This spatial overlap intensifies competition among LCCs. In contrast, FSCs are more concentrated on the region around their hubs, which reduces geographic overlap, compared to LCCs flying everywhere in Europe.
- **Node colors indicate overall competitive exposure and strategic positioning:** In Figure 6b, node colors represent the sum of all I_2 values associated with each airline, providing a measure of overall exposure to spatial competition. Airlines with lower summed values (i.e., U2, SK, UX) are embedded in dense competitive clusters with multiple rivals, whereas those with higher summed values (i.e., A3, LH, AY, OS) operate more distinct or dispersed networks from others. These results highlight not only the intensity of competition across European routes but also the strategic positioning of airlines within the spatial market landscape.

This application introduces a spatially grounded measure of airline competition by evaluating the substitutability of operating routes based on geographical proximity using an OT framework. Unlike traditional MMC indices, which typically rely on route overlap counts, the proposed approach captures both route partial substitutability and regional competition. In this way, the framework provides a continuous and nuanced measure of multimarket competition intensity, offering a more detailed perspective on airline spatial competition analysis.

7 Application 3: Comparing operational cost based on route length

This application shifts the focus from demand-oriented supply to distance-based operational characteristics, with particular attention to frequency deployment and the environmental implications of flights. By incorporating route length differences into the OT framework, we assess how costly it is for one airline to transition toward another’s frequency supply configuration by comparing their route length choices (short-, medium-, and long-haul routes).

7.1 Cost matrix analysis

The cost matrix d_3 is constructed using absolute differences in route length. Routes with similar distances incur lower costs, reflecting comparable aircraft requirements, fuel consumption profiles, and operating conditions. Figure 7 presents the empirical distribution of route length differences, where smaller values correspond to routes of more similar lengths. The distribution is noticeably right-skewed: the mean is 905.5, standard deviation 745.1, minimum 1, 25th percentile 334, median 721, 75th percentile 1,283, and maximum 4,955. This skew indicates that although many routes have similar lengths, a considerable number differ substantially, which can have a pronounced impact on operational costs and scheduling. The wide dispersion and heavy tail reflect the significant variability in route lengths across Europe.

7.2 Top 20 airlines market overview

Using the cost matrix based on route length absolute differences, we calculated the index I_3 for the top 20 airlines ranked by flight frequency. The results are shown in Figure 8, presented using the same plotting format as in the previous applications for consistency. Several patterns emerge:

1. **Separation between LCCs and FSCs:** We observe a separation between LCCs and FSCs in Figure 8b. This division is driven by their different cost structures and fleet strategies. LCCs typically operate homogeneous fleets optimized for routes of similar length. FSCs, by contrast, maintain more heterogeneous networks that include a wider range of route lengths, reflecting mixed route length strategies within domestic markets.

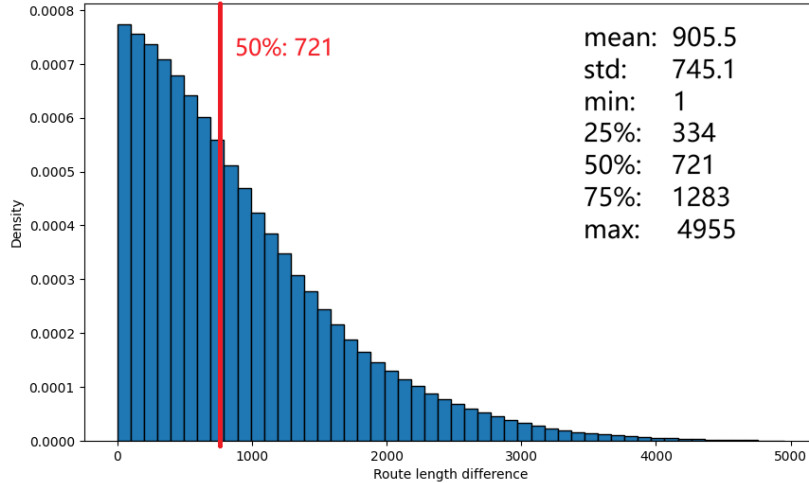


Figure 7: Histogram of route length differences between routes in 2024.

2. **Isolated NT–WF cluster:** Figure 8b reveals that an isolated cluster emerges for Binter Canarias (NT) and Widerøe (WF). This pattern is confirmed in Figure 8a, where NT and WF exhibit large I_3 values, represented by red and blue bubbles indicating index values above the 25th percentile relative to other airlines, while showing small I_3 values (green bubbles, within the lowest 25%) with each other. This can be attributed to their highly specific operational environments, characterized by regional networks and limited route length variability. Their route networks are markedly different from those of both mainline carriers and LCCs operating across Europe, leading to high transition costs when compared with other airlines.

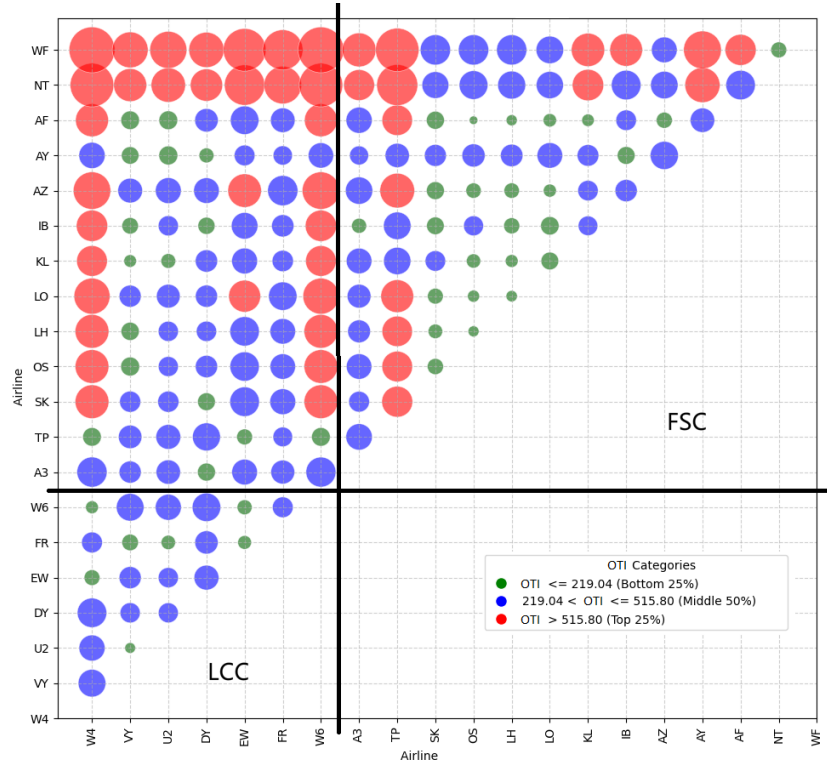
8 Discussion

Our study demonstrates that OT offers a coherent and flexible framework for comparing airline flight supply across multiple dimensions. By incorporating domain-specific cost structures into the OT formulation, the proposed indices capture features of airline pairwise relationships that are not observable using traditional market-level or airline-level measures.

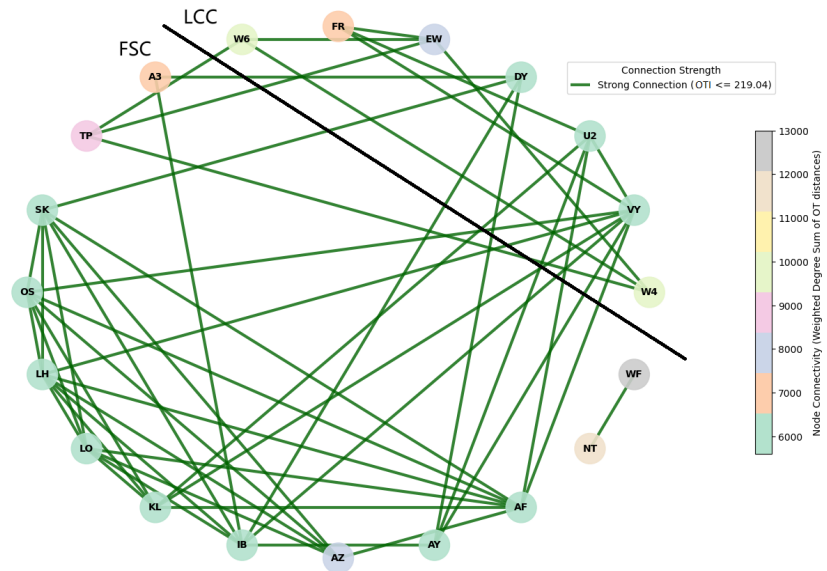
First, the seasonal-demand-based index (I_1) shows that airlines can exhibit strong strategic similarity even when serving different routes, provided that these routes share comparable seasonal demand profiles. The contribution is twofold. On the one hand, we use the weekly changes in flight capacity at the route level as a proxy indicator for the evolution of market demand over time. This representation is entirely data-driven and does not rely on ex ante behavioral assumptions. The resulting measure of route similarity provides insights for network design decisions. More broadly, our study introduces a novel perspective on measuring route similarity that complements existing route clustering approaches based on network topology or operational characteristics (Sun and Wandelt, 2014; Bombelli et al., 2017). On the other hand, the clustering of certain low-cost carriers (LCCs) with full-service carriers (FSCs) indicates a convergence in capacity allocation strategies, consistent with the gradual hybridization of airline business models observed in recent years.

Second, the geography-based index (I_2) provides a continuous measure of multimarket competition intensity by accounting for spatial substitutability between routes. Unlike traditional MMC indices, which treat markets as either overlapping or not, I_2 reflects gradually varying competitive pressure across neighboring route markets. The denser competitive clusters among LCCs confirm their concentration on high-demand profitable routes, whereas FSCs display more geographically diversified networks, reducing direct spatial competition.

Third, the route-length-based index (I_3) captures operational and cost-related differences in frequency deployment strategies. The clear separation between LCCs and FSCs reflects distinct fleet structures and cost models. The isolation of regional carriers such as NT and WF further illustrates how regional network structures translate into higher transition costs compared with other main European carriers. Overall, the findings indicate that OT-based measures provide a richer, multidimensional



(a) I_3 indices of airlines represented by bubble size and categorical colors.



(b) Network representation of airline pairs within the lowest 25% of I_3 values.

Figure 8: Illustration of I_3 values among the top 20 airlines, based on route length differences.

analysis of airline supply, capturing both strategic similarity and competitive pressure.

Despite these strengths, the proposed approach has several limitations. The choice of cost parameters (e.g., the spatial decay parameter λ) affects the results of I_2 , and therefore requires careful calibration. In addition, the framework relies on observed supply, which may not fully reflect latent passenger preferences. Future research could address these limitations by incorporating regional GDP, price information, weekly passenger flow data, or utility-based demand models to refine the cost structure and better approximate underlying demand. In addition, the current framework treats routes independently and therefore does not explicitly account for network effects arising from connecting flights. Extending the approach to incorporate network connectivity, such as transfer opportunities, hub structures, or itinerary-based competition, would allow a more comprehensive representation.

At the same time, the OT framework provides notable methodological advantages. It enables asymmetric comparisons, accommodates heterogeneous cost structures, and remains robust to differences in network size. Building on these properties, a natural extension is the development of asymmetric indices to investigate imbalances in competitive pressure between airlines. Moreover, the OT index can be computed at different seasonal aggregation levels, such as monthly or quarterly, allowing seasonal patterns in airline supply to be examined. Finally, by comparing an airline's OT indices with those of its key competitors over time, the framework can be used to trace its strategic evolution within the competitive landscape.

9 Conclusion

This study shows that optimal transport (OT) provides a rigorous and flexible framework for comparing airline route-supply distributions under multiple cost formulations on a common geometric space, which here is the space of European domestic routes. We address different questions and construct cost matrices accordingly, while preserving methodological consistency. By representing airline supply as probability distributions and defining relevant cost structures, the proposed indices capture seasonal, spatial, and operational dimensions of strategic similarity and competition.

The index I_1 captures similarities in supply relative to seasonal demand patterns, I_2 measures competition through spatial substitutability across routes, and I_3 reflects operational differences in route length and frequency deployment. Applied to European domestic airline markets, these indices draw a competition map among airlines, identify meaningful similarity airline clusters, and reveal substantial heterogeneity in operational strategies. Relative to traditional MMC or concentration measures, the OT-based approach offers an interpretable continuous measure, by naturally accommodating partial substitutability between markets and heterogeneous route conditions.

Beyond the supply of air passenger transport, this framework can be extended to other transport systems, such as rail, road, maritime, and freight transport, where the supply is spatially distributed and where transport costs can be defined consistently. In this sense, OT constitutes a general methodological contribution for analyzing competition, network structural changes, and strategic similarity in transport markets.

Future research could further enrich the framework by incorporating pricing information, passenger flows, environmental cost components, or interactions across transport modes, thereby deepening its relevance for competition analysis and sustainability assessment.

Funding

This research was possible thanks to the financial support of the Région Occitanie/Pyrénées-Méditerranée, and the École nationale de l'aviation civile (ENAC).

References

Aguirregabiria, V. and Ho, C.-Y. (2010). A dynamic game of airline network competition: Hub-and-spoke networks and entry deterrence. *International Journal of Industrial Organization*, 28(4):377–382.

- Aguirregabiria, V. and Ho, C.-Y. (2012). A dynamic oligopoly game of the us airline industry: Estimation and policy experiments. *Journal of Econometrics*, 168(1):156–173.
- Alderighi, M., Cento, A., Nijkamp, P., and Rietveld, P. (2007). Assessment of new hub-and-spoke and point-to-point airline network configurations. *Transport Reviews*, 27(5):529–549.
- Allard, R. F. and Moura, F. (2016). The incorporation of passenger connectivity and intermodal considerations in intercity transport planning. *Transport Reviews*, 36(2):251–277.
- Baumgarten, P., Malina, R., and Lange, A. (2014). The impact of hubbing concentration on flight delays within airline networks: An empirical analysis of the us domestic market. *Transportation Research Part E: Logistics and Transportation Review*, 66:103–114.
- Bombelli, A., Segarra Torne, A., Trumbauer, E., and Mease, K. D. (2017). Automated route clustering for air traffic modeling. In *AIAA Modeling and Simulation Technologies Conference*, page 1318.
- Brander, J. A. and Zhang, A. (1990). Market conduct in the airline industry: an empirical investigation. *The RAND Journal of Economics*, pages 567–583.
- Brander, J. A. and Zhang, A. (1993). Dynamic oligopoly behaviour in the airline industry. *International journal of Industrial organization*, 11(3):407–435.
- Ciliberto, F., Murry, C., and Tamer, E. (2021). Market structure and competition in airline markets. *Journal of Political Economy*, 129(11):2995–3038.
- Daft, J. and Albers, S. (2015). An empirical analysis of airline business model convergence. *Journal of Air Transport Management*, 46:3–11.
- Du, W.-B., Zhou, X.-L., Lordan, O., Wang, Z., Zhao, C., and Zhu, Y.-B. (2016). Analysis of the chinese airline network as multi-layer networks. *Transportation Research Part E: Logistics and Transportation Review*, 89:108–116.
- Galichon, A. (2016). *Optimal transport methods in economics*. Princeton University Press.
- Goetz, A. R. (2002). Deregulation, competition, and antitrust implications in the us airline industry. *Journal of Transport Geography*, 10(1):1–19.
- Graham, B. and Vowles, T. M. (2006). Carriers within carriers: A strategic response to low-cost airline competition. *Transport Reviews*, 26(1):105–126.
- Grauberger, W. and Kimms, A. (2016). Airline revenue management games with simultaneous price and quantity competition. *Computers & Operations Research*, 75:64–75.
- Hess, S. (2016). Modelling air travel choice behaviour. In *Airport Competition*, pages 151–175. Routledge.
- International, A. C. (2023). European regional airports.
- Kantorovich, L. V. (1942). On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, volume 37, pages 199–201.
- Kouwenhoven, M. (2008). The role of accessibility in passengers’ choice of airports.
- Kwieciński, D. (2017). Measures of competitive intensity—analysis based on literature review. *Central European Management Journal*, 25(1):53–77.
- Leite, D. and De Bacco, C. (2024). Similarity and economy of scale in urban transportation networks and optimal transport-based infrastructures. *Nature Communications*, 15(1):7981.
- Li, Z.-C., Lam, W. H., Wong, S., and Fu, X. (2010). Optimal route allocation in a liberalizing airline market. *Transportation Research Part B: Methodological*, 44(7):886–902.
- Lieshout, R. (2012). Measuring the size of an airport’s catchment area. *Journal of Transport Geography*, 25:27–34.

- Lijesen, M. G., Nijkamp, P., and Rietveld, P. (2002). Measuring competition in civil aviation. *Journal of Air Transport Management*, 8(3):189–197.
- Liu, Q. (2026). Enhancing airline multimarket competition analysis with a pairwise competitive intensity index based on weighted jaccard index. *Journal of Air Transport Management*, 131:102931.
- Lohmann, G. and Koo, T. T. (2013). The airline business model spectrum. *Journal of Air Transport Management*, 31:7–9.
- Lordan, O., Sallan, J. M., Escorihuela, N., and Gonzalez-Prieto, D. (2016). Robustness of airline route networks. *Physica A: Statistical Mechanics and its Applications*, 445:18–26.
- Lordan, O., Sallan, J. M., and Simo, P. (2014). Study of the topology and robustness of airline route networks from the complex network approach: a survey and research agenda. *Journal of Transport Geography*, 37:112–120.
- Ma, W., Wang, Q., Yang, H., and Zhang, Y. (2020). Is multimarket contact an antitrust concern? a case of china’s airline market. *Transportation Research Part A: Policy and Practice*, 132:515–526.
- Magdalina, A. and Bouzaima, M. (2021). An empirical investigation of european airline business models: Classification and hybridisation. *Journal of Air Transport Management*, 93:102059.
- Mazzeo, M. J. (2003). Competition and service quality in the us airline industry. *Review of industrial Organization*, 22(4):275–296.
- Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704.
- Montesuma, E. F., Mboula, F. M. N., and Souloumiac, A. (2024). Recent advances in optimal transport for machine learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Oliveira, M. V. and Oliveira, A. V. (2018). What drives effective competition in the airline industry? an empirical model of city-pair market concentration. *Transport Policy*, 63:165–175.
- Oum, T. H., Zhang, A., and Zhang, Y. (2005). Inter-firm rivalry and firm-specific price elasticities in deregulated airline markets. *Transport Economics*, pages 239–262.
- Peyré, G., Cuturi, M., et al. (2019). Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607.
- Prince, J. T. and Simon, D. H. (2009). Multimarket contact and service quality: Evidence from on-time performance in the us airline industry. *Academy of Management Journal*, 52(2):336–354.
- Rozenberg, R., Szabo, S., and Šebešćáková, I. (2014). Comparison of fsc and lcc and their market share in aviation. *International Review of Aerospace Engineering (IREASE)*, 7(5):149–154.
- Schipper, Y., Rietveld, P., and Nijkamp, P. (2003). Airline deregulation and external costs: a welfare analysis. *Transportation Research Part B: Methodological*, 37(8):699–718.
- Sun, X. and Wandelt, S. (2014). Network similarity analysis of air navigation route systems. *Transportation Research Part E: Logistics and Transportation Review*, 70:416–434.
- Sun, X., Zheng, C., Wandelt, S., and Zhang, A. (2024). Airline competition: A comprehensive review of recent research. *Journal of the Air Transport Research Society*, 2:100013.
- Verma, T., Araújo, N. A., and Herrmann, H. J. (2014). Revealing the structure of the world airline network. *Scientific reports*, 4(1):5638.
- Villani, C. et al. (2009). *Optimal transport: old and new*, volume 338. Springer. page 30-31.
- Wang, W., Fu, X., Wang, K., Sun, X., Wandelt, S., Wang, J., and Wu, X. (2025). Low-cost carrier development under airline-within-airline strategy: Bibliometric analysis and systematic literature review. *Research in Transportation Business & Management*, page 101305.

Williams, K. R. (2022). The welfare effects of dynamic pricing: Evidence from airline markets. *Econometrica*, 90(2):831–858.

Yuan, Z. and Barwick, P. J. (2024). Network competition in the airline industry: An empirical framework. Technical report, National Bureau of Economic Research.

Zhang, Y. and Round, D. K. (2009). Policy implications of the effects of concentration and multimarket contact in china’s airline market. *Review of Industrial Organization*, 34(4):307–326.

Zou, L., Yu, C., and Dresner, M. (2012). Multimarket contact, alliance membership, and prices in international airline markets. *Transportation Research Part E: Logistics and Transportation Review*, 48(2):555–565.

A Proof

Proposition 1. Let $D(\cdot, \cdot)$ denote the Haversine distance on the sphere, which is a metric. For two unordered OD pairs $r = \{x_1, x_2\}$ and $s = \{y_1, y_2\}$, define

$$d_2(r, s) = \min\{D(x_1, y_1) + D(x_2, y_2), D(x_1, y_2) + D(x_2, y_1)\}.$$

Then d_2 is a metric on the space of unordered OD pairs.

Proof. We verify the four metric axioms.

Non-negativity. Since $D \geq 0$, both sums $D(x_1, y_1) + D(x_2, y_2)$ and $D(x_1, y_2) + D(x_2, y_1)$ are non-negative. Their minimum is therefore non-negative:

$$d_2(r, s) \geq 0.$$

Identity of indiscernibles. If $r = s$, then up to permutation $x_1 = y_1$ and $x_2 = y_2$, hence

$$d_2(r, s) = D(x_1, x_1) + D(x_2, x_2) = 0.$$

Conversely, if $d_2(r, s) = 0$, then one of the two sums is zero. Since D is a metric, this implies that the corresponding matched points coincide. Therefore, $r = s$ as unordered pairs.

Symmetry. By symmetry of D ,

$$D(x_i, y_j) = D(y_j, x_i),$$

and the two candidate sums defining $d_2(r, s)$ are identical to those defining $d_2(s, r)$. Hence

$$d_2(r, s) = d_2(s, r).$$

Triangle inequality. Let $r = \{x_1, x_2\}$, $s = \{y_1, y_2\}$, and $t = \{z_1, z_2\}$. Let σ and τ be permutations in \mathfrak{S}_2 achieving the minima in $d_2(r, s)$ and $d_2(s, t)$, respectively. Then

$$d_2(r, s) = D(x_1, y_{\sigma(1)}) + D(x_2, y_{\sigma(2)}),$$

$$d_2(s, t) = D(y_1, z_{\tau(1)}) + D(y_2, z_{\tau(2)}).$$

By the triangle inequality for D ,

$$D(x_i, z_{\tau(\sigma(i))}) \leq D(x_i, y_{\sigma(i)}) + D(y_{\sigma(i)}, z_{\tau(\sigma(i))}) \quad \text{for } i = 1, 2.$$

Summing over $i = 1, 2$ yields

$$D(x_1, z_{\tau(\sigma(1))}) + D(x_2, z_{\tau(\sigma(2))}) \leq d_2(r, s) + d_2(s, t).$$

Since $d_2(r, t)$ is the minimum over all such matchings,

$$d_2(r, t) \leq D(x_1, z_{\tau(\sigma(1))}) + D(x_2, z_{\tau(\sigma(2))}) \leq d_2(r, s) + d_2(s, t).$$

Conclusion. All metric axioms are satisfied; hence d_2 is a distance on the space of unordered OD pairs. \square