Journal of Air Transport Management 67 (2018) 181–196

Contents lists available at ScienceDirect

Journal of Air Transport Management

journal homepage: www.elsevier.com/locate/jairtraman

Characterizing air traffic networks via large-scale aircraft tracking data: A comparison between China and the US networks

Pan Ren[a], Lishuai Li[b,*]

[a] Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong Special Administrative Region
[b] Department of Systems Engineering and Engineering Management, City University of Hong Kong, 83 Tat Chee Avenue, Hong Kong Special Administrative Region

ARTICLE INFO

Keywords:
Air traffic management
Air traffic network
Data mining
Data analytics

ABSTRACT

Air travel demand has continued to increase rapidly over the past decade, causing severe flight delays. To reduce such delays, Air Navigation Service Providers need to first understand the operational capacity and congestion risks associated with a network, and then develop strategies accordingly. However, limited studies have been conducted due to lack of data. New opportunities have arisen given the availability of large-scale aircraft tracking data and many other digitalized records of operations. In response, we develop a novel data-driven framework that characterizes the operational structure and dynamics of an air traffic network using actual tracking data. The framework includes several new statistical measures and data analytic techniques to summarize airspace availability, network structure, and utilization patterns. We then apply the framework to analyze the air traffic networks in China and the US. The results reveal distinctive characteristics of these two networks. Airspace availability for commercial flights is much more restricted in China than the US. The network in China has a clear structure with distinct utilization patterns, while the network in the US has a more flexible structure featuring complex dynamics. These operational differences indicate that China faces a greater chance of en-route congestion when compared with the US. The results also demonstrate that the data-driven approach is effective to identify the actual behavior and complexity of an air traffic network, which are not captured by existing methods.

1. Introduction

Air transport system capacity enhancements have failed to keep up with the increasing pace of demand growth all over the world, causing severe air traffic congestion and flight delays, which have had high economic costs and negative environmental effects (Ball et al., 2010). To reduce these delays, it is critical to understand the operational capacity, efficiency, and congestion risks associated with an air transport network, and to carry out strategic planning and tactical management interventions accordingly to mitigate these issues.

Despite extensive studies on air transport network, few research has been done to characterize the operational structure and dynamics of national-level air traffic networks, or compare how airspace is managed between different regions based on actual air traffic flows. A key limitation lies in the availability of operational data across different sources and regions. For example, aircraft tracks, air traffic control commands, fleet scheduling, these data were heavily regulated by national or regional agencies and airlines without proper information sharing among them. Air traffic service providers can rarely grasp a 'big picture' of the regional airspace.

New opportunities have arisen from the increasing availability of digitized air traffic data. For example, with the implementation of Automatic Dependent Surveillance – Broadcast (ADS-B), a satellite-based surveillance technology that tracks and broadcasts the location of each aircraft via satellite, it is now possible to track and analyze aircraft movement data on a global scale. With the appropriate analytical tools, post-event analysis can be carried out to examine the characteristics of the actual air traffic network.

Thus, the development of analytical tools to analyze large-scale multi-source operational data can significantly contribute to the improvement of air transport system. To support this effort, in this study, we aim to characterize the structure and dynamics of the air traffic network via a data-driven approach using large-scale aircraft tracking data. The analysis result will allow Air Navigation Service Providers (ANSP) to understand the current network better, identify deficiencies in Air Traffic Management (ATM) procedures, and provide recommendations for improving system capacity and efficiency.

Relevant literature exists in two groups. The first group of research focus on understanding the air transport system using network analysis, in which the air transport system is considered as a complex network

* Corresponding author.
E-mail addresses: panren2-c@my.cityu.edu.hk (P. Ren), Lishuai.li@cityu.edu.hk (L. Li).

https://doi.org/10.1016/j.jairtraman.2017.12.005
Received 22 September 2017; Received in revised form 8 December 2017; Accepted 16 December 2017
Available online 11 January 2018

composed of many interlinked subsystems. Some of these studies use complex network theory to analyze the topological characteristics of the system (Guimera et al., 2005; Vespignani, 2012; Cook et al., 2015; He at al., 2004; Li and Cai, 2004). In these studies, the air transport system is abstracted to a directed/undirected, weighted/unweighted network, where nodes are represented by airports and edges are direct flights linking two airports. The network is evaluated and optimized using the robustness metrics in network theory, which include betweenness, degree, centrality, and connectivity (DeLaurentis et al., 2008). Wei et al, (2014) designed a robust air transportation network by maximizing algebraic connectivity to reduce air traffic congestion. However, these studies only considered a static graph, meaning that the dynamics of the network were not explicitly considered. To address this limitation, other studies have focused on studying how delays propagate in the network. Empirical studies using flight schedule and delay data have analyzed the causes for primary delays and assessed how they spread over the network, causing reactionary delays (Beatty et al., 1999; Fricke and Schultz, 2009; Hansen, 2002; Abdelghany et al., 2004; Jetzki, 2009; Rebollo and Balakrishnan, 2014). Furthermore, Fleurquin et al. (2013) defined metrics that quantify the level of network congestion and the macro-scale behavior of delay dynamics. Péter and Szabó (2012) proposed an exact mathematical model for large-scale dynamic networks from the perspective of control theory using a class of positive systems, which could be used to achieve minimized delays. Another stream of studies used queueing theory to model and simulate flight delays within a network (Pyrgiotis et al., 2013; Hansen et al., 2009; Long and Hasan, 2009; Peterson et al., 1995a, 1995b; Xu, 2007; Shah et al., 2005); however, the performance of these models relies on how well the underlying conceptual network can represent real air traffic operations.

Recent studies suggest that representing air transport systems as airport networks is an oversimplified approach, as it discards important operational information of ATM (i.e., flow management). Some scholars propose that the air transport network can be regarded as aggregated multi-layers of airline networks (Zanin and Lillo, 2013; Du et al., 2016; Belkoura et al., 2016), airport networks, air navigation route networks (Sun et al., 2015), and air traffic management networks (Wang et al., 2017). Although the findings of these studies helped to better conceptualize the complexity of air transport systems, there is much work to be done to obtain a more comprehensive picture. One drawback associated with the current work on multi-layer air transport networks is that this endeavor requires a significant amount of detailed information from ANSP and airlines that can construct the different layers of the system, such as air route information, sector maps, letter of agreements (LOA), airline route networks, and so on; this makes it cumbersome to use across the air transport networks of different regions.

The second group of relevant research focus on to how to analyze and use large-scale air traffic operational data. Several studies have been carried out that use clustering techniques on actual tracking data for air traffic flow identification. Eckstein (2009) developed a flight track taxonomy to decompose a set of radar tracks according to their lateral, vertical, and conformance segments, for the purpose of evaluating procedural conformance of individual flights in the terminal airspace. Gariel et al, (2011) developed an analytical framework that uses density-based clustering to learn the typical patterns of traffic flows, which are then used as benchmarks to monitor the behavior of an aircraft in a given airspace. Other studies use hierarchical or spectral clustering to identify air traffic flows to and from an airport (Rehm, 2010; Enriquez, 2013). Nevertheless, these studies focused on flow identification on a relatively small scale – i.e., at the terminal area or by examining flows to and from an airport, without further analysis on ATM operations. The exception to this is the work of Conde Rocha Murca et al. (2016), as this research team stepped forward along the direction of characterizing large-scale ATM operations. The authors developed a data mining framework to characterize the air traffic flows in the transition/terminal airspace, and demonstrated how to use the results to assess the performance of tactical operations in the New York region on a daily basis.

In summary, further research is needed to study the air transport system, understand its network's structure and dynamics at different scales, and ensure that the analysis results are useful for ATM improvement. The availability of large-scale operational data and the advancements in data mining techniques have created unprecedented opportunities for such research. The aim of this investigation is thus to develop a data-driven framework to analyze an air traffic network using aircraft-tracking data, including route availability, network structure, and utilization patterns. The framework includes an innovative statistical measure – a modified Ripley's K-function – to assess air route availability, clustering techniques to identify major air routes, network analysis to quantify the actual air traffic network structure, and spatial–temporal analysis to reveal airspace utilization patterns. In order to demonstrate the proposed framework, we apply it to analyze and compare the air traffic networks in China and the US using one month of historical flight-tracking data.

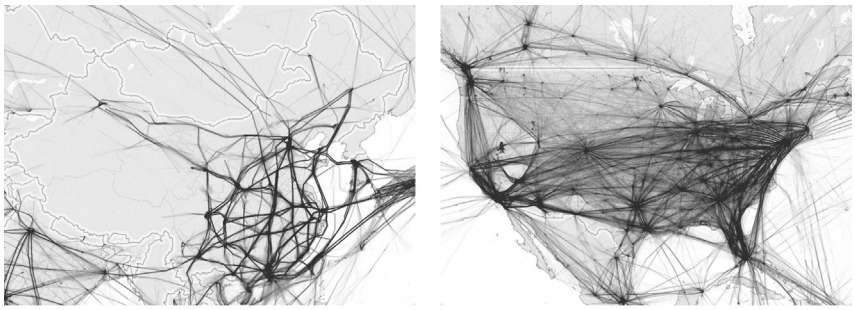
The remainder of this paper is organized as follows. Section 2 describes the aircraft-tracking data used in this study. Section 3 presents the proposed data-driven framework, including the algorithms and metrics used in each module of the framework. In Section 4, we conduct a comparative study of China and the US airspaces using the proposed framework. Finally, Section 5 summarizes our study and suggests some future research directions.

2. Dataset

Automatic Dependent Surveillance – Broadcast (ADS-B) is a key component of the US Next Generation Air Transportation System (NextGen) and it enables aircraft to track their positions and broadcast them via satellites (Gugliotta, 2009). ADS-B greatly enhances the safety of air travel by not only providing air traffic control with real-time, consistent, and visible position updates, but it also notes other aircraft equipped with ADS-B. Currently, 70% of all commercial passenger aircraft are equipped with an ADS-B transponder and the percentage is steadily increasing, as the ADS-B transponder will become mandatory for most aircraft around the world by 2020 (Flightradar24, 2015). Air traffic management decision makers can benefit significantly from such data when engaging in performance assessments, real-time monitoring, and strategic planning. Currently, several flight-tracking service providers operate a worldwide network of ADS-B receivers to collect and share live flight tracks, such as Flightrader24 and FlightAware.

The dataset used in this study features flight-tracking data collected from Flightradar24 every minute for 30 consecutive days, from November 1 to November 30, 2016, covering the airspace in China and the US. The geographic ranges of the airspace in China and the US are set as (17.37, 46.00, 92.86, 126.50) and (24.80, 49.20, −124.90, −60.40), respectively. The first two values are the boundary latitudes, while the last two values are the boundary longitudes. Fig. 1 shows all flight trajectories collected. For this research, we focus on air traffic networks consisting of the top 10 busiest airports in China and the US, as ranked by annual passenger traffic (CAAC, 2015a; ACI-NA, 2015). The information related to these airports is presented in Appendix A. In addition, we filter out airport pairs where the sample size is less than one flight per day. As a result, 40 airport pairs in China and 45 airport pairs in the US are left in our analysis. The flight trajectories of top 10 airports in China and the US are visualized in Fig. 2.

We choose to compare the air traffic network in China and the US because: 1) they are the busiest two regions in the world (ranked by passengers carried in 2015 (WBG, 2015)); 2) both are faced with significant flight delays (in 2015, 18.27% of flights were delayed and the average delay time relative to scheduled time was 11 min in the US (BTS, 2015), while 31.67% of flights were delayed and the average delay time was 21 min in China (CAAC, 2015b)); and 3) the airspaces



a) China b) USA

Fig. 1. All flight trajectories collected in China and the US.

are managed differently, which can result in distinctive air traffic network characteristics – i.e., 80% of China's national airspace is devoted to military use, while the US military only controls 20% of the US airspace, primarily over remote or ocean areas (Hsu, 2014).

3. Methodology

A new data-driven framework is proposed to characterize the operations of a national air traffic network based on ADS-B data. The framework consists of three parts in sequence: Part 1 quantifies airspace route availability; Part 2 analyzes air traffic network structure; and Part 3 reveals network utilization patterns. Fig. 3 provides an overview of the framework; it shows the sequence of the three parts, the analysis and results associated with each part, and demonstrates how they are interrelated. The details of each part are described in the following subsections.

3.1. Quantify airspace route availability

3.1.1. Filtering and resampling

The raw ADS-B data may contain incomplete flight tracks that do not link to an origin or a destination airport due to ADS-B receiver coverage limitations. Therefore, data filtering is performed to filter out the incomplete flight tracks based on the following rules: If the distance between the start point of a flight track and its associated origin airport is greater than 25 miles, or if the distance between the end point of a flight track and its associated destination airport is greater than 25 miles, the flight track is considered incomplete and is excluded from our analysis.

Then, resampling is performed to convert all raw flight tracks into time series with the same length of N data points. A raw flight track is represented by a vector, $F_i = \{x_{i1}, x_{i2}, x_{i3}, ..., x_{ij}, ..., x_{iN_i}\}$, where F_i represents Flight i, x_{ij} is the geographic coordinates of the j-th data point of Fi, and N_i is the original data length of F_i. N equals the average value of N_i across all flights within an origin and destination (OD) pair. The detailed resampling procedure is as follows:

- Step 1: Set the link between the origin airport and destination airport of F_i as the benchmark line, represented by OD.
- Step 2: Divide the OD line into (N–1) equal parts to get N benchmark points, including the start point (origin airport) and end point (destination airport).
- Step 3: Draw perpendicular lines of the OD line through each benchmark point, P_i.
- Step 4: Find the intersecting point between each perpendicular line and flight track F_i; these points constitute the resampled flight track, denoted as $F_i' = \{x_{i1}', x_{i2}', x_{i3}', ..., x_{ij}', ..., x_{iN}'\}$.

Fig. 4 illustrates the resampling method.

3.1.2. Modified Ripley's K-function

To measure air route availability in a specific airspace, we propose a new statistical measure, a modified Ripley's K-function. Air route availability is one of the factors that determine airspace capacity. An airspace with more space for commercial flights generally allows more routing choices and faces fewer opportunities for en-route congestion. For example, Fig. 5 shows the flight tracks between two airport pairs: 1) Guangzhou Baiyun International Airport (CAN) and Chengdu Shuangliu International Airport (CTU) in China and 2) Dallas/Fort Worth International Airport (DFW) and Chicago O'Hare International Airport (ORD) in the US. The distance and number of flights are similar in the two airport pairs; however, the air route availability is different. Flights between CAN and CTU only have one routing choice, with some deviations when cruising through the airspace; therefore, it has a higher chance of being congested, especially under convective weather impact.

Such differences in air route availability are visually apparent, yet challenging to quantify. Simple trajectory clustering to identify major routing choices will not work because airspaces of difference sizes cannot be compared fairly. In this paper, we measure air route availability via two approaches: 1) simple descriptive statistics of Euclidean distance between flight tracks by airport pair; and 2) a scale-free indicator, the modified Ripley's K function, to measure how dispersed/concentrated flight tracks are in a specified airspace.

A modified Euclidean distance between flight tracks by airport pair is calculated using the following equation:

$$D(i,j) = \frac{\sum_k dis(x'_{ik}, x'_{jk})}{N} \quad (1)$$

where x'_{ik} is the k-th geographic coordinate vector of flight F'_i, x'_{jk} is the k-th geographic coordinate vector of flight F'_j, N is the sample size of

A) The top 10 airports in China B) The top 10 airports in the US

Fig. 2. Flight trajectories of the top 10 airports in China and the US.

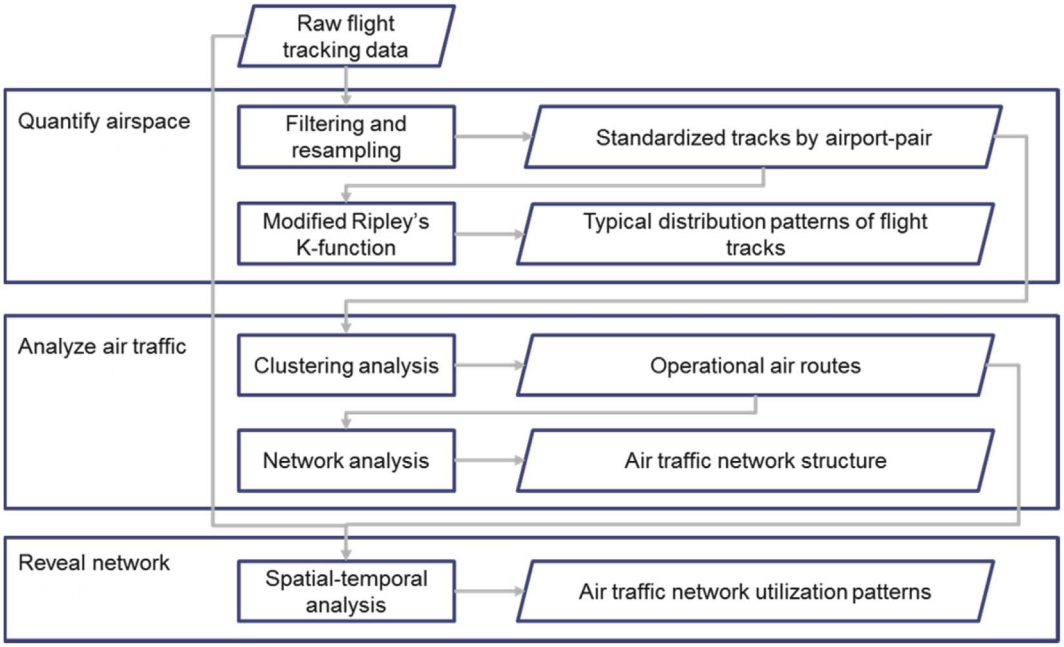


Fig. 3. Air traffic network characterization framework.

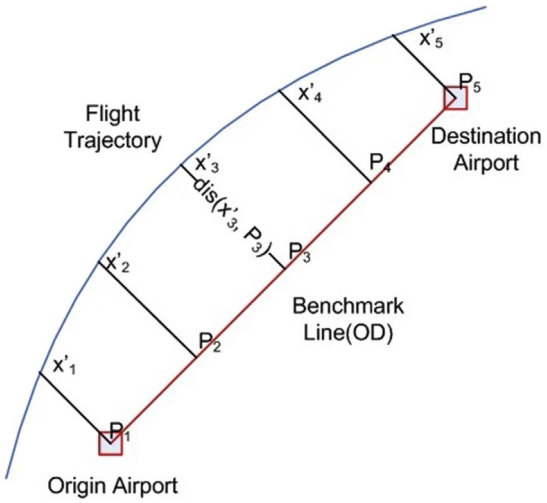


Fig. 4. Flight trajectory resampling method.

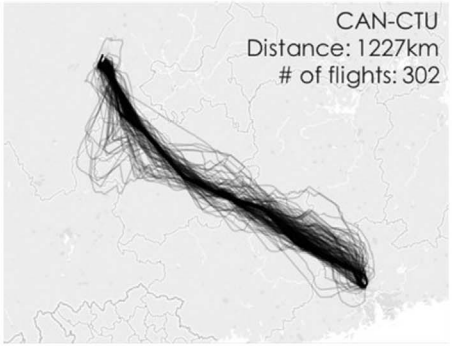
each resampled flight track, and $dis(x'_{ik}, x'_{jk})$ is the great-circle distance between trajectory points x'_{ik} and x'_{jk}. Flight F'_i and flight F'_j belong to the same airport pair. To compare different airport pairs fairly, the Euclidean distance between two flight tracks is standardized as standardized neighborhood distance (*SND*) by a factor of D_{OD}, which is the great circle distance between the origin and destination airport. The *SND* is calculated as

$$SND_{ij} = \frac{\sum_k dis(x'_{ik}, x'_{jk})}{N * D_{OD}} \tag{2}$$

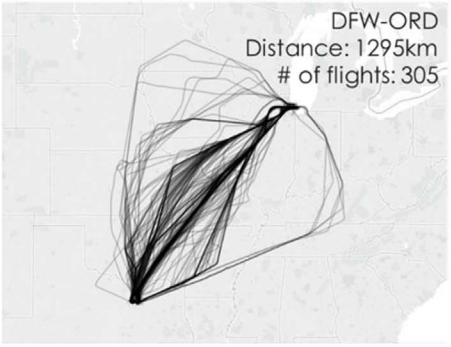
The Ripley's K-function is a widely used spatial analysis method that is employed to describe the distribution pattern of points over a given region of interest (Haase, 1995; Dixon, 2002). It allows researchers to determine whether the points appear to be dispersed, clustered, or randomly distributed throughout the study region. The function counts the number of neighboring points found within a given distance for each individual point; its formula (Haase, 1995) is as follows:

$$K(t) = \lambda^{-1} \sum_{i \neq j} I(d_{ij} < t)/n \tag{3}$$

where d_{ij} is the Euclidean distance between the i-th and j-th points in a dataset of n points, I equals 1 if $d_{ij} < t$ (the radius of the circle),



A) Guangzhou Baiyun International Airport (CAN) - Chengdu Shuangliu International Airport (CTU) in China



B) Dallas/Fort Worth International Airport (DFW) - Chicago O'Hare International Airport (ORD) in the US

Fig. 5. An example of air route availability within an airport pair.

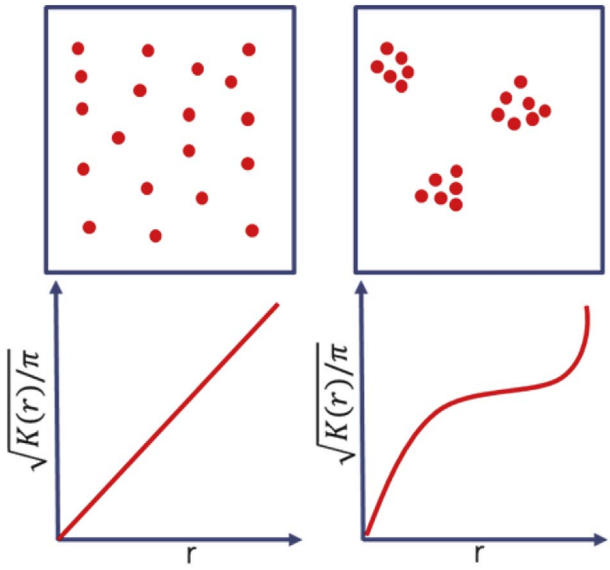


Fig. 6. Typical point distribution patterns and the corresponding K-function curves.

otherwise it equals to 0 and λ is the density of points (generally estimated as n/A, where A is the area of the region contains all points). Fig. 6 indicates two typical flight trajectory distribution patterns. The Ripley's K-function curve of randomly distributed points tends to be steadier than clustered points.

Here, we extend Ripley's K function to describe the spatial distribution patterns of flight trajectories in the specified airspace. The formula is modified as:

$$K(r) = \frac{\sum_{i=1}^{N_m} \sum_{j=1}^{N_m} h_r(SND_{ij})}{N_m^2} \quad (4)$$

where N_m is the number of flight tracks in airport pair m, r is a given neighborhood distance, and $h_r(SND_{ij})$ equals 1 if $SND_{ij} < r$, and 0 otherwise.

The new Ripley's K-function computes the percentage of neighboring flight tracks within a given distance of each flight track. Fig. 7 indicates three typical distribution patterns of flight tracks between two airports, which are: Class 1–distributed in several clusters, Class 2–concentrated in one cluster, and Class 3–dispersed.

We propose a method to classify these three typical patterns based on the properties of K-function curves. First, to identify Class 1 cases, where the trajectories are distributed in several clusters, we identify the inflection points of the K-function curve, which are the points at which the curve changes from being convex to concave, or vice versa, indicating the number of neighboring flight track stops or starts increasing with distance r. Two criteria are applied successively to identify Class 1 cases. First, there must be at least two inflection points. Second, the distance between two consecutive inflection points, which is calculated as SND^*D_{OD}, must be greater than 10 nm (width of an airway) to ensure that the gap between clusters of flight tracks is significant enough. Fig. 8 shows the inflections of K-function curves for three cases. In the left panel, there are three inflection points, and the distances between all consecutive points are greater than 10 nm; therefore, the distribution of flight tracks in this airport pair would be identified as Class 1. In the center panel, even though there are six inflection points, the distances between all consecutive pairs are less than 10 nm. In the right panel, there is only one inflection point. For the central and right panels, these cases cannot be classified into Class 1. Additional information needs to be verified in order to group them into either Class 2 or 3.

Then, to distinguish between Classes 2 and 3, we check the value of r when K(r) equals 0.9 ($r_{k(r)=0.9}$) and compare it with a threshold value. $r_{k(r)=0.9}$ represents the SND that covers 90% of flight tracks in an airport pair. When $r_{k(r)=0.9}$ is less than the threshold value, we consider that the flight tracks are concentrated in a single cluster and are classified as Class 2; otherwise, flight tracks are dispersed and classified as Class 3. The threshold value is determined by sensitivity analysis. We

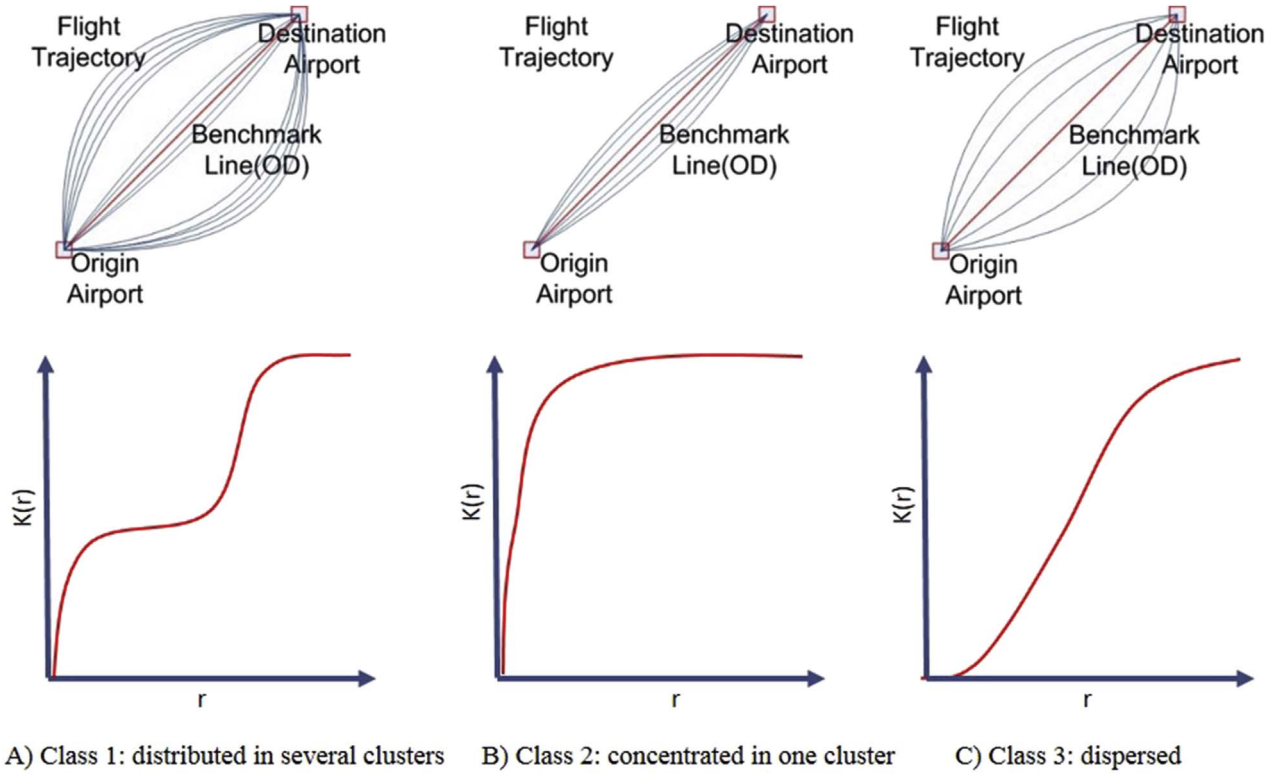


Fig. 7. Typical distribution patterns of flight tracks and the corresponding K-function curves.

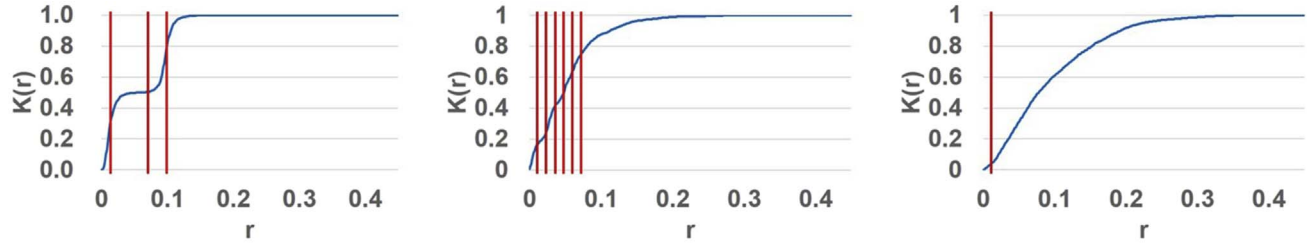


Fig. 8. Use of inflection points on K function curves to identify Class 1 cases.

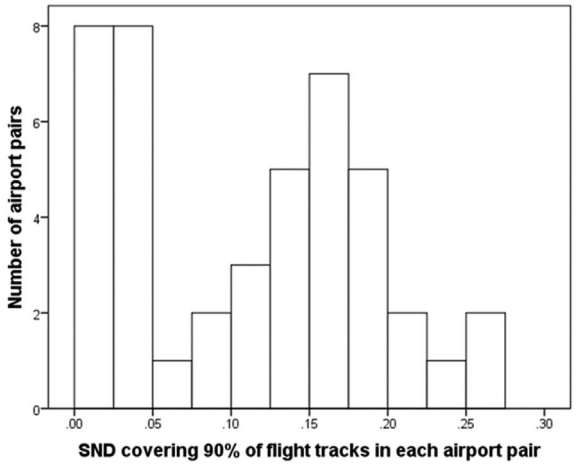


Fig. 9. Distribution of the SND covering 90% of flight tracks in an airport pair for all airports pairs included in the study in both China and the US, excluding Class 1 airport pairs.

summarize the distribution of "the SND covering 90% flight tracks in an airport pair" of all airports pairs included in the study in both China and the US, excluding Class 1 airport pairs; this distribution is shown in Fig. 9. It can be observed that there is a significant gap around 0.05; as such, we set the threshold value as 0.05.

3.2. Analyze air traffic network structure

The second part of the framework focuses on characterizing network structure based on actual air traffic flows. Trajectory clustering and air route merging is first performed to identify the network structure of national-level air traffic flows. Network analysis is then conducted to evaluate the features of real network operations, including centrality, transport efficiency, connectivity, and robustness.

3.2.1. Trajectory clustering

Cluster analysis is performed on flight tracks by airport pair to better understand the air routes used by actual flights, without any prior knowledge of airspace structure. These air routes are then used to construct an operational air traffic network. Cluster analysis is a commonly used data-mining technique that can identify common groups of observations in a dataset. Many clustering algorithms have been developed for different purposes. Here, we use a clustering algorithm called Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al., 1996). The basic idea of DBSCAN is to progressively find density-connected points to form a cluster. If at least Minpt points lie within a ball of radius centered at a point, a cluster is then created. Two unique features of DBSCAN make it suitable for this problem: 1) it can automatically determine the number of clusters; and 2) it can handle data with outliers. In this problem, the number of air routes may not be the same across different airport pairs, and this number is unknown without prior knowledge of airspace structures. Using DBSCAN, this number can be automatically determined based on data distribution patterns. Abnormal flight tracks can emerge due to vectoring or under other special conditions, which can be treated as outliers in DBSCAN. Several studies have shown that DBSCAN is an effective technique for identifying the norms of operations in air transportation systems (Gariel et al., 2011; Li et al., 2015; Conde Rocha Murca et al., 2016).

Two key input parameters, Minpt and Epsilon, may result in changes to the clustering solution. Epsilon is the maximum radius of neighborhood distance between points and Minpt is the minimum number of points in an Epsilon neighborhood. For our dataset, the clustering result is not sensitive to Minpts when it is between 3 and 10, so we set Minpt as 5 for all cases. The value of Epsilon is set differently for each airport pair to find the best result that matches the distribution pattern. We compute the k-nearest neighbor distances – where k equals Minpt – plot these k-distances in a descending order for all data points, and find the first "valley" of the sorted k-distance graph (Ester et al., 1996). This "valley" point corresponds to a threshold point where a sharp change in the gradient occurs along the k-distance curve, representing a change in density distribution amongst data points. The k-distance value of the threshold point is used as the Epsilon value for DBSCAN (Ester et al., 1996).

3.2.2. Air route merging

We should note that the air routes generated via trajectory clustering cannot be directly used to construct the air traffic network, as the same airspace resource used by different airport pairs could be identified as multiple air routes, particularly since the clustering is performed by airport pair. For example, as shown in Fig. 10A, the air route depicted in red linking Chengdu Shuangliu International Airport (CTU) and Beijing Capital International Airport (PEK) shares the same en-route airspace resource with two other airport pairs: the air route in blue that links CTU and Xi'an Xianyang International Airport (XIY) (blue line) and the one in green that links XIY and PEK. Fig. 10B shows another example: the air route between Hong Kong International Airport (HKG) and PEK (blue line) shares the same en-route airspace resource with the air route between Shenzhen Bao'an International Airport (SZX) and PEK (red line).

To ensure the reality of air route structure reflected in the network that we construct from the ADS-B data, we develop a procedure to merge air routes using the same airspace resource after the trajectory clustering. For each air route identified from trajectory clustering, we evaluate whether it should be merged with any other route. The merge criteria are: If the Euclidean distance* between one air route and

* The Euclidean distance is calculated as $\sum_k dis(x_{ik}, x_{jk})/N$, where x_{ik} is the k-th point

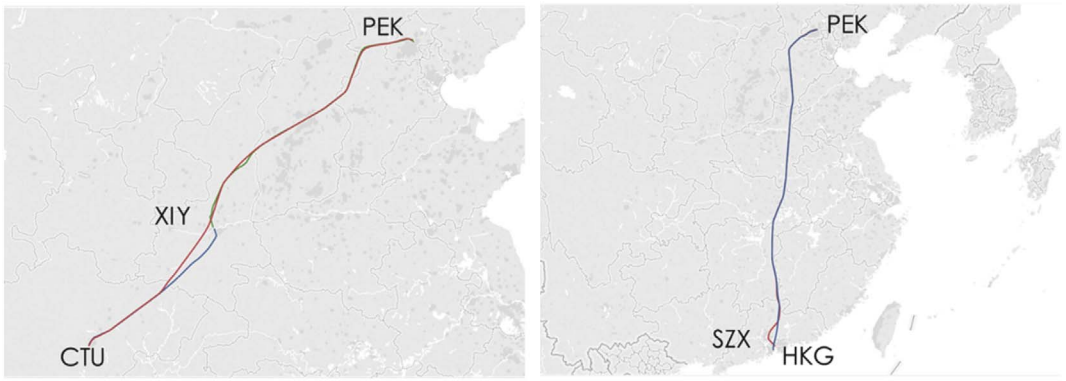


Fig. 10. Examples of air routes sharing the same airspace resource.

another is less than 25 nm, this air route is deleted and the traffic load on it is merged with the other. The pseudo code of the air route merge procedure is shown as follows:

```
Air route merging procedure
1   Input: Air route set AR, Air route traffic load set T
2   Start
3   for each air route ar_i in air route set AR
4       if ar_i has been visited
5           continue next air route
6       else
7           Mark ar_i as visited
8           for each other air route ar_j in AR
9               if (distance(ar_i, ar_j)<25nm)
10                  lr=Max(ar_i , ar_j)    # The longer air route
11                  sr=Min(ar_i , ar_j)    # The shorter air route
12                  T_sr= T_sr + T_lr    # Combine traffic load on sr and lr to sr
13                  Split lr into 3 segments by using airports of sr as split-point
14                  Delete lr from air route set AR
15                  Delete the segment sharing the same airspace resource with sr
16                  for each other segment
17                      if (length(segment)>100nm) # Enough to be a new route
18                          Add it to air route set AR as a new air route
19                          T_arnew= T_ari    # Assign traffic load on lr to new air route
20                      end
21                  end
22              end
23              break
24          end
25      end
26  end
27  End
28  Output: Updated air route set AR', Updated traffic load set T'
```

Image 1

3.2.3. Network structure analysis

After the air routes are identified by cluster analysis and merged using the merging procedure, we construct an undirected weighted network and apply network analysis to assess its structural features. We denote the air traffic network as $G(V, E)$, where node set V represents airports and edge set E represents the links between two airports. The adjacency matrix of the air traffic network consisting of $N_{airport}$ airports is defined by a $N_{airport}$* $N_{airport}$ binary matrix $A = \{a_{ij}\}$, whose element a_{ij} equals 1 when there is at least one air route (after merging procedure) connecting airport i to airport j, and 0 otherwise (i, j = 1,2,…,$N_{airport}$). Weights of edges are represented by weight matrix $W = \{w_{ij}\}$, where w_{ij} is defined by the number of air routes (after the merging procedure) linking airport i and airport j.

Several metrics are available for the network structure assessment. Here, we choose three commonly used metrics to evaluate the heterogeneity among nodes, the efficiency of transport on a network, and the resilience of a network.

1) Weighted degree centrality

The weighted degree centrality is a measure of the strength of vertices in a network. The node degree is extended to the strength degree for a weighted network, which is the sum of the weights of all links attached to node i (Barrat et al., 2004). The strength degree of airport i in the air traffic network is defined as:

$$s_i = \sum_{j=1}^{N_{airport}} a_{ij} w_{ij} \tag{5}$$

The degree centrality of the air traffic network is calculated as:

$$S = \frac{1}{N_{airport}} \sum_{i=1}^{N_{airport}} s_i \tag{6}$$

And the standard deviation of weighted degree centrality is calculated as:

$$S_{SD} = \sqrt{\frac{1}{N_{airport}} \sum_{i=1}^{N_{airport}} (s_i - \bar{s})} \tag{7}$$

2) Characteristic path length

The characteristic path length measures the efficiency of information or mass transporting on a network, which is defined as the average of the shortest path lengths between any two nodes of a network. Since a higher weight w_{ij} implies more available air routes between airport i and airport j in the air traffic network, the path length between airport i and airport j is therefore shorter. Consequently, the shortest path length is defined as the smallest sum of the inverse weights of the links throughout all possible paths from node i to node j (Newman, 2001), which is calculated as:

$$Sd_{ij} = \min_{\gamma(i,j)\in\Gamma(i,j)} \sum_{m,n\in\gamma(i,j)} \frac{1}{w_{mn}} \tag{8}$$

where $\gamma(i, j)$ is a path from node i to node j, $\Gamma(i, j)$ is the class of all possible paths from i to j, and m and n are the points along the path $\gamma(i, j)$. The characteristic path length of the air traffic network is defined as:

$$P = \frac{2}{N_{airport}(N_{airport} - 1)} \sum_{i,j} Sd_{ij} \tag{9}$$

When there are more air routes linking two airports, the shortest path length is smaller, more routing choices are available, and more traffic throughput per unit of time is possible; therefore, a network with a smaller characteristic path length facilitates more efficient and rapid flight movements.

3) Algebraic connectivity

Algebraic connectivity is calculated as the second smallest eigenvalue of the Laplacian matrix for a weighted network (Fiedler, 1989). The Laplacian matrix is defined as $L = \{l_{ij}\} = D\text{-}W$, where D is the degree matrix, W is the weight matrix, and l_{ij} is the Laplacian distance between node i and node j, which is calculated as:

$$l_{ij} = \begin{cases} -w_{ij} & if \ \ i \neq j \\ \sum_{i=1}^{n} w_{ij} & if \ \ i = j \end{cases} \tag{10}$$

Since the second smallest eigenvalue of L serves as a lower bound of node connectivity and edge connectivity, the algebraic connectivity is very important when measuring the robustness of a network to node and its edge failures; in this sense, the larger the algebraic connectivity, the more difficult it is to cut a graph into independent components, which means that this network is increasingly more robust.

3.3. Reveal network utilization patterns

The last part of the framework aims to reveal the utilization pattern of the air traffic network based on actual traffic flow data. The constructed air traffic network from previous steps represents the static structure of a national air traffic network. From an operational perspective, we would also like to understand the utilization dynamics of this air traffic network to answer questions such as, “Which are the busy or rarely used air routes?”, “Are there any typical patterns in selecting which air routes to use during different times of the day or under light/heavy traffic demand?”, etc.

To reveal such network utilization patterns, we perform a spatial–temporal analysis by constructing a Network Traffic Load Matrix, which is denoted as NTLM = $\{n_{ij}\}$; elements n_{ij} are equal to the number of flights traveling along air route j in hour i, if this air route is used, otherwise it equals 0. Each row of NTLM reflects the distribution of traffic load across air routes in a given hour, while each column gives the traffic load on an air route across the hours included in the analysis. Cluster analysis is performed on both the rows and columns to identify the typical utilization patterns and groups of air routes that share similar usage patterns.

4. Results

4.1. Data filtering

Following a data filtering, about 90% of the flights in the 40 airport pairs in China are recognized as complete flights, and 72% of the flights in the 45 airport pairs in the US are recognized as complete flights. Table 1 shows the detailed numbers.

4.2. Results of airspace route availability

4.2.1. Statistics on Euclidean distance between flight tracks by airport pair

Within each airport pair, we calculate the modified Euclidean distance of all pairs of trajectories following Equation (1), and record four statistical measures: the maximum value, minimum value, mean value, and standard deviation. The maximum value reflects the distance between the two furthest points on two flight trajectories, while the minimum value measures the distance between the two nearest trajectories. To summarize the entire air traffic network, we calculate the average values of these four measures over all airport pairs, and the results are shown in Table 2; for example, column 3 shows the mean value of pairwise distance between trajectories, which implies that the average Euclidean distance between a pair of tracks connecting any particular airport pair in China (the US) is equal to 25.29 nm (54.60 nm).

Table 1
Data filtering results: number of flights included in the study.

	China (40 airport pairs)	United States (45 airport pairs)
Number of original flights	42,815	48,797
Number of complete flights	38,420	35,058
Percentage of complete flights	90%	72%

Table 2
Statistics on the pairwise distance between trajectories within each airport pair over the China and the US networks (nm).

	Max	Min	Mean	Std
China	99.56	0.43	25.29	18.64
United States	269.86	0.19	54.60	41.13

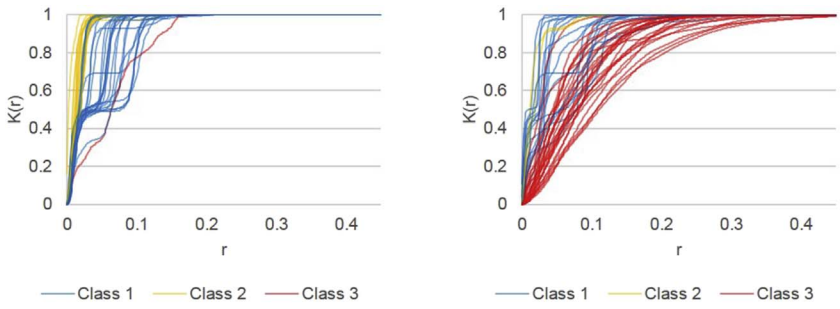


Fig. 11. Modified Ripley's K-function curves of the flight track distributions in each airport pair.

As shown in Table 2, and except for the minimum value, the US has larger values than China on the remaining three measures. Both the maximum and mean values of the pairwise distance reflect how much airspace is available for flight tracks between two airports. The results demonstrate that airspace availability is much larger in the US than in China. The standard deviation is also larger in the US, which means that the variation of distances between flight tracks is larger in the US than in China. The minimum value is lower in the US than in China, which shows that it is rare to have two flights flying on the exact same track in China. The potential reasons for this could be the implementation of a less advanced navigation system and/or different air traffic procedures. The key finding from this result is that flight tracks within each airport pair are much more concentrated in China than in the US, indicating a more restricted airspace for commercial flights in China.

4.2.2. Distribution patterns of flight tracks based on a modified Ripley's K function

Given the standardized neighborhood distance of flight tracks, we obtain the modified Ripley's K-function curve of flight track distribution in each airport pair in China and the US, as shown in Fig. 11. These K curves are classified based on the criteria proposed in Section 3.1.2., whereby the different cases were categorized into three classes: Class 1 (in blue): distributed in several clusters; Class 2 (in yellow): concentrated in one cluster; and Class 3 (in red): dispersed. Fig. 12 shows examples of real flight tracks of three distribution types and the corresponding K curves.

The classification results of K-function curves for China and the US are listed in Table 3. Flight tracks in most airport pairs in China are identified as clustered, while flight tracks in the US are more dispersed.

Both the Euclidean distance analysis and modified Ripley's K-function analysis demonstrate that airspace route availability is much more limited in China than in the US – flights can only fly through limited air corridors in China, while airlines in the US have more freedom to determine their flight routes.

Table 3
Classification results of K-function curves.

	Class 1: distributed in several clusters	Class 2: concentrated in one cluster	Class 3: dispersed
China	25	14	1
United States	16	2	27

4.3. Results of air traffic network structure

4.3.1. Trajectory clustering

Fig. 13 shows the trajectory clustering results; the centroids of the operational air routes linking the top 10 airports in China and the US are displayed as black lines. In this case, 100 and 179 air routes are identified in China and the US, respectively, which means that, on average, there are about two air routes per airport pair in China and four per airport pair in the US for the top 10 airports.

The number of clusters and the percent of outliers identified for each airport pair are shown in Figs. 14 and 15. It is evident that more clusters are identified in the US, which means that there more operational air routes are available in the US. Moreover, the percent of outliers is also higher in the US, which indicates that flights do not follow rigid air corridor structures in the US.

4.3.2. Air route merging

After the air route merging procedure, the number of available air routes linking the top 10 airports in China had reduced from 100 to 66, while the number of available air routes linking the top 10 airports in the US had only reduced by 6, which indicates that in China, there are more air routes linking different airport pairs that share the same

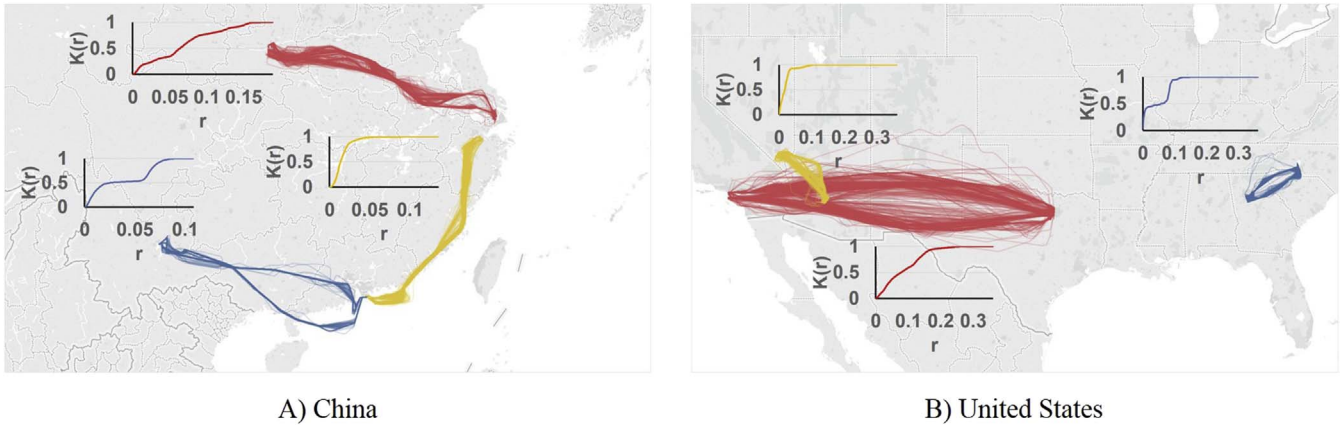
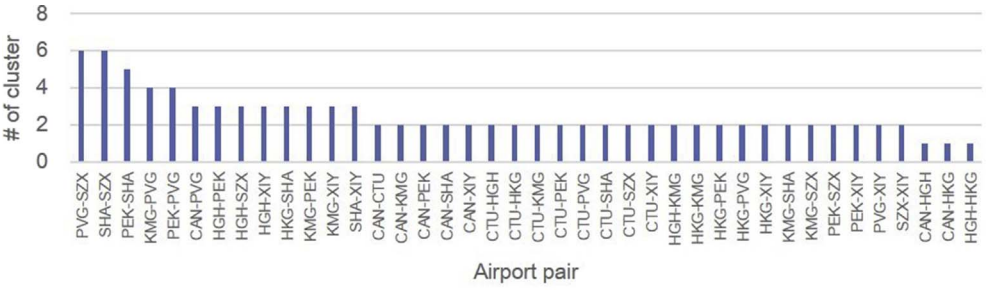


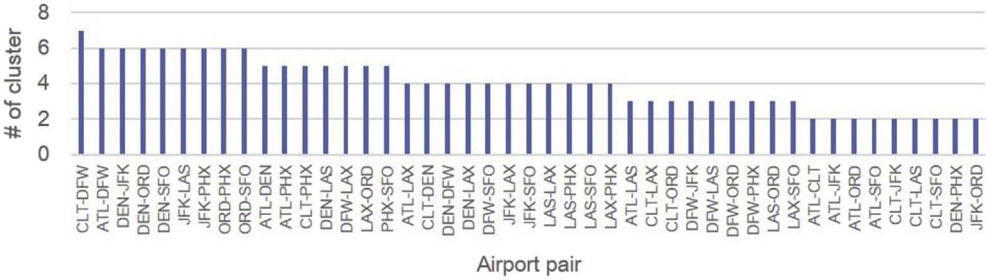
Fig. 12. Examples of real flight tracks of three distribution types and the corresponding K curves.

A) China B) United States

Fig. 13. Centroid of operational air routes linking the top 10 airports in China and the US.



A) China



B) United States

Fig. 14. Number of clusters identified in each airport pair.

airspace when compared with the US. The results also indicate that the traditional airport network cannot adequately represent the air transport system in China. The before-and-after comparisons of air routes are shown in Figs. 16 and 17.

4.3.3. Network structure

We quantify the air traffic network structure using three network metrics. The weighted degree centrality is a measure of the strength of airports in the network; the characteristic path length measures the efficiency of air transportation on the network; and the algebraic connectivity reflects the resilience of the air traffic network. In addition, we compare the network metrics of our network based on actual traffic with the traditional airport-based network, which is constructed by simply linking the top 10 airports, without considering operational air routes and traffic. The results are listed in Table 4.

The results show that, when considering the actual air routes, the difference in the air traffic network structure between China and the US is much more significant when compared with the traditional network based on airports alone. The air traffic network can better reveal the reality of airspace structures.

The weighted degree centrality of the air traffic network is larger in the US, which indicates that the node strength is higher. There are about 35 air routes connected to each airport on average, which is about three times that of China. It is much easier for airports in the U.S. air traffic network to reach another airport.

The characteristic path length of the US air traffic network is smaller than in China. Airports in the US are much “closer” to one another due to the greater availability of routes connecting the airports. From air service providers' points of view, the US network is more powerful in transporting airplanes.

Moreover, the algebraic connectivity of the US is larger. The failure rate of links between airport pairs are much lower because aircraft have more air route choices when cruising the US airspace. The network is therefore more robust and tends to have fewer failures when an abnormality occurs, such as server weather in certain locations, a temporary close-down of an airport, etc.

4.4. Results of network utilization patterns

Based on 30 days' flight-tracking data, the NTLM is a 720×66 matrix for China and a 720×173 matrix for the US, respectively. Fig. 18 shows the heatmaps for the NTLMs of China and the US. The rows depict the 720 h of the month, the columns show the air routes, and the color intensity indicates the traffic load on a particular air route in an hour. Details of the air route index are listed in Appendix B. Visually, China's network China tends to follow a few major patterns, while the network utilization dynamics of the US are much more complicated.

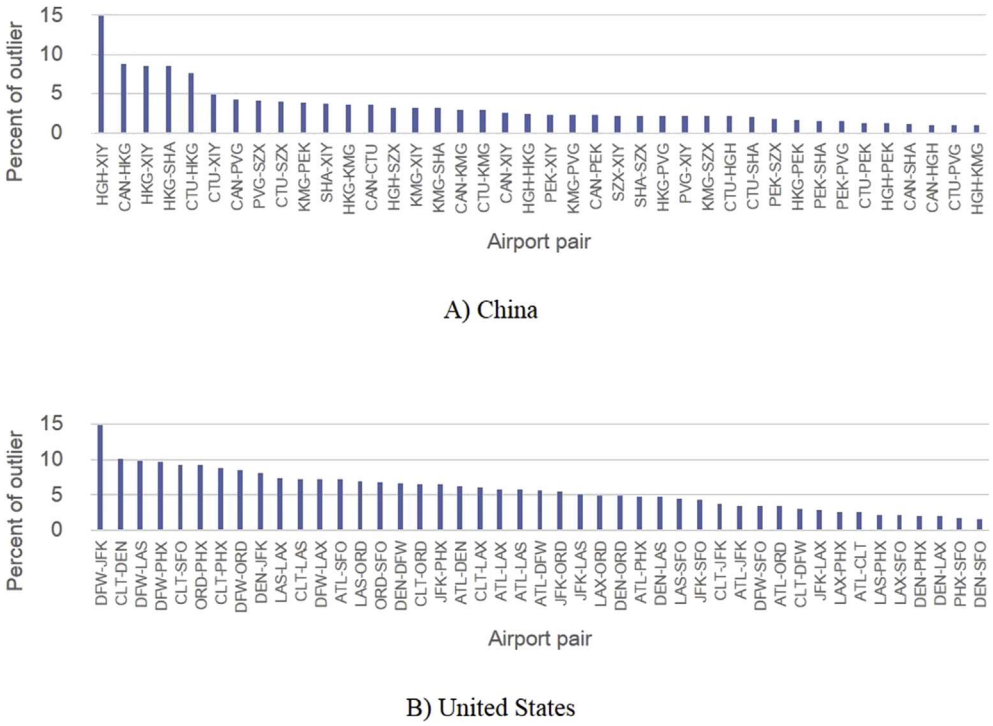


Fig. 15. Percent of outliers in each airport pair.

4.4.1. Identification of network utilization patterns

To identify the different network utilization patterns in the China and US networks, K-means clustering is performed on the rows of each of the two NTLMs. We use the within-sum-of-squares (WSS) metric and implement the elbow method to find a suitable value of K (Kodinariya and Makwana, 2013). As a result, four clusters and seven clusters were identified in China and the US, respectively. Tables 5 and 6 summarize the network utilization pattern represented by each cluster.

In China, there are two major network utilization patterns: Pattern 1 is "nighttime operations with light traffic load" and Pattern 2 is "daytime operations with heavy traffic load". In addition, two small clusters representing two other patterns were identified. Pattern 3 corresponds to "morning peak hour operations with heavy traffic loads on air routes of PEK-Pearl River Delta Region, CTU-Pearl River Delta Region, and PEK-XIY" and Pattern 4 relates to "special late-night operations with heavy traffic loads on the air routes of PEK-Pearl River Delta Region, Yangtze River Delta Region-Pearl River Delta Region, and PEK-XIY". Fig. 19 show the details of the four network utilization patterns in China, including the histograms of the utilization patterns over each hour of the day, the average number of flights per hour on each route, and the detailed network flows. Below is a detailed description of each pattern according to the information shown in Fig. 19.

Pattern 1 "nighttime operations with light traffic load": Traffic activity is primarily observed during the nighttime and early morning (0–8am). The traffic load across all air routes is relatively light. The air traffic network is basically in an "off" mode.

Pattern 2 "daytime operations with heavy traffic load": Traffic activity surges between 10 a.m. and 10 p.m. Most of the air routes are heavily loaded with traffic. The entire network can be considered to be in an "on" mode.

Pattern 3 "morning peak hour operations with heavy traffic loads on air routes of PEK-Pearl River Delta Region, CTU-Pearl River Delta Region, and PEK-XIY": heavy traffic load on a few routes during the morning (6 a.m. and 10 a.m.). This can be viewed as a special operating mode of the system.

Pattern 4 "special late-night operations with heavy traffic loads on the air routes of PEK-Pearl River Delta Region, Yangtze River Delta Region-Pearl River Delta Region, and PEK-XIY": heavy traffic on some routes between 0 a.m. and 2 a.m. This could be night cargo traffic or passenger traffic that absorbs the delays accumulated during the daytime. It can be viewed as another special operating mode of the system.

When compared with China, it is evident that network utilization is much more complex in the US. Although seven patterns corresponding to seven clusters are identified, the operational characteristics of the seven patterns are not that distinctive, as shown in Table 6. The diversity of different operational patterns over the entire national air traffic network

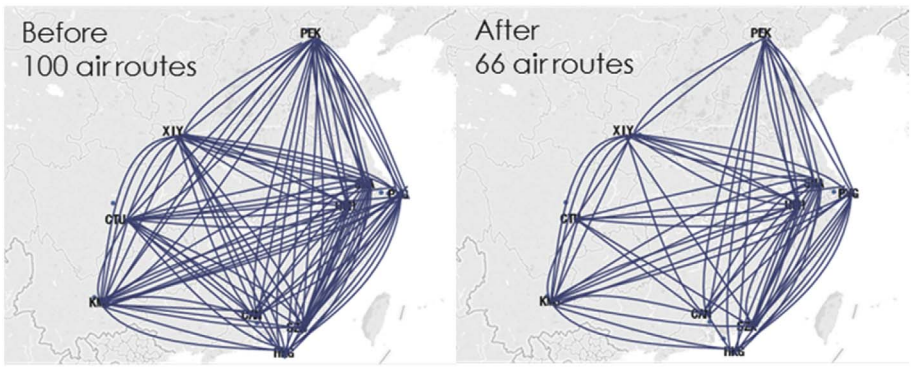


Fig. 16. Available air routes in China before and after air route merging.

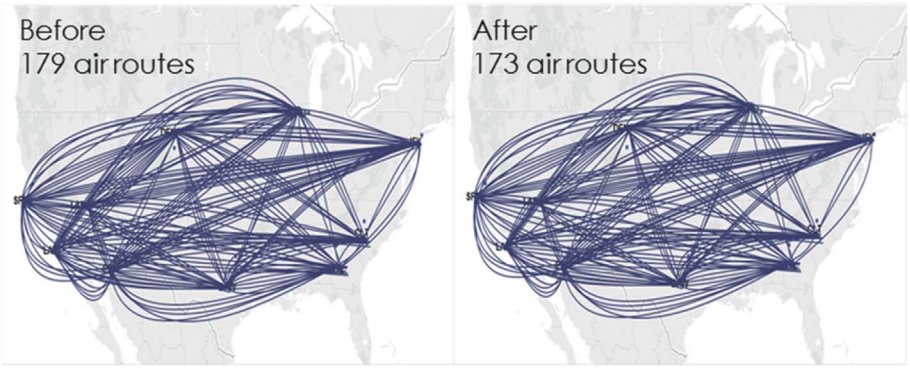


Fig. 17. Available air routes in the US before and after air route merging.

Table 4
Network structure assessment results.

	Network based on airports alone		Network based on actual air traffic	
	China	United States	China	United States
Weighted degree centrality (average)	8	9	13.20	34.60
Weighted degree centrality (standard deviation)	0.94	0	3.01	3.53
Characteristic path length	1.11	1	0.63	0.27
Algebraic connectivity	7	10	8.84	28.26

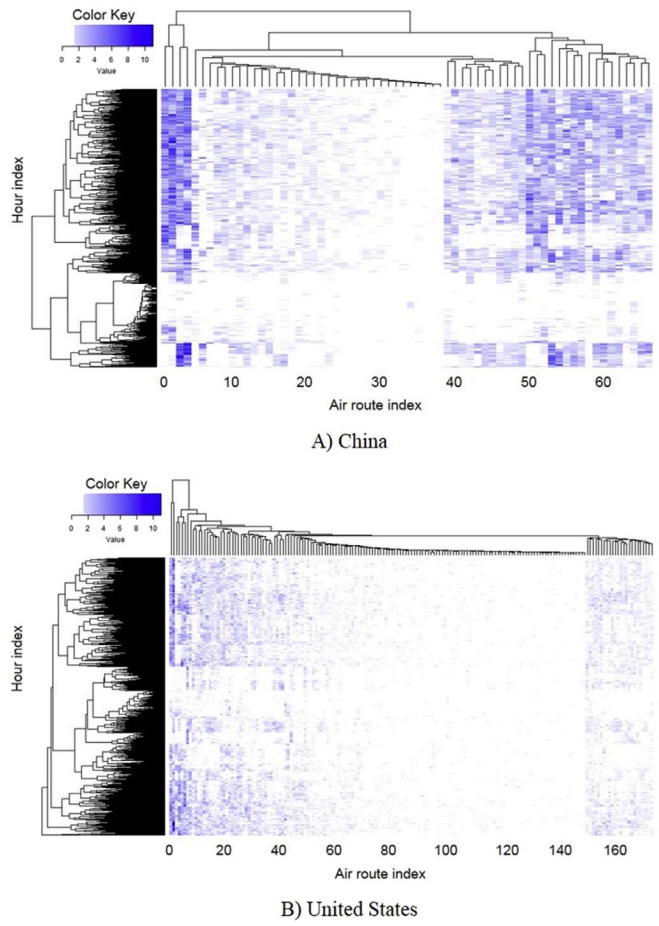


Fig. 18. Heatmaps of the network traffic load matrix.

can be hardly captured by clustering analysis on a month's worth of data. We pick three relatively distinctive patterns and show them in Fig. 20. The three patterns are Pattern 1 “nighttime operations”, Pattern 3 “daytime operations”, and Pattern 6 “evening peak hour operations”.

Overall, the results of the network utilization analysis reveal that both the structure and dynamics of the air traffic network are relatively simple in China. Fewer routes and fewer utilization patterns were identified, meaning that the network in China was either “full” or “empty” for most of the time. By contrast, the air traffic network in the US is much more complicated, as it demonstrates variable structure and utilization patterns, depending on many factors, such as weather, traffic demand, and capacity allocation.

5. Conclusion

The availability of large-scale aircraft tracking data and many other digitalized records of operations have risen new opportunities to characterize an air traffic network in terms of its actual behavior and complexity. In this paper, we proposed a novel framework to quantify the operational capacity and utilization patterns of an air traffic network in national airspace using ADS-B data. Several novel statistical measures and data analytic techniques are integrated in this framework. A case study is implemented based on the framework to analyze and compare the air traffic networks in China and the US. The results reveal that, airspace availability for commercial flights is much more restricted in China, and the network structure is less well connected and less robust, indicating that the operational airspace capacity of China is lower than that of the US. Moreover, air traffic network utilization is less flexible and less diverse in China. China's air traffic network maintains the same structure and runs with either “full” or “little” traffic loads throughout different hours of a day, while much more diverse operation patterns are identified in the US.

This work can be directly applied in practice to help ANSPs compare their operations worldwide and inform the development of appropriate strategies to improve the capacity and efficiency of an air traffic system. For example, strategies to reduce flight delays in China should consider en-route congestion, if airspace availability cannot be changed in the short term due to social–political constraints. From the perspective of academic research, this work present a new way of constructing an air traffic network model. Compared with the traditional graph theory approach, the proposed data-driven approach extracts features of the network from operational data, rather than assumptions, which makes the finding more meaningful for real-world ATM operations.

Based upon this work, future work can focus on developing ATM strategies to reduce flight delays utilizing the network features characterized through the data-driven approach. We can apply this work to other network models, such as the one proposed by Péter and Szabó (2012). Moreover, we can built new network flight delay simulation models to explicitly consider the en-route congestions. Lastly, decision support systems can be developed to evaluate and recommend ATM improvement strategies.

Table 5
Network utilization patterns in China.

	Descriptive	# of data points	Avg. # of active routes	Avg. # of flights
Pattern 1	Nighttime operations, light traffic load	172	6	8
Pattern 2	Daytime operations, heavy traffic load	377	36	78
Pattern 3	Morning peak hour operations, heavy traffic load on certain routes	91	28	63
Pattern 4	Special late-night operations, heavy traffic load on certain routes	80	22	45

Table 6
Network utilization patterns in the US.

	Descriptive	# of data points	Avg. # of active routes	Avg. # of flights
Pattern 1	Nighttime operations, light traffic load	139	12	17
Pattern 2	Morning peak hour operations, medium traffic load on certain routes	127	22	29
Pattern 3	Daytime operations, medium traffic load on most routes	108	31	44
Pattern 4	Noon peak hour operations, medium traffic load on most routes	81	40	62
Pattern 5	Special day operations, heavy traffic load on certain routes	78	38	61
Pattern 6	Evening peak hour operations, heavy traffic load	73	52	83
Pattern 7	Late-night operations, heavy traffic load in western and middle US	114	40	65

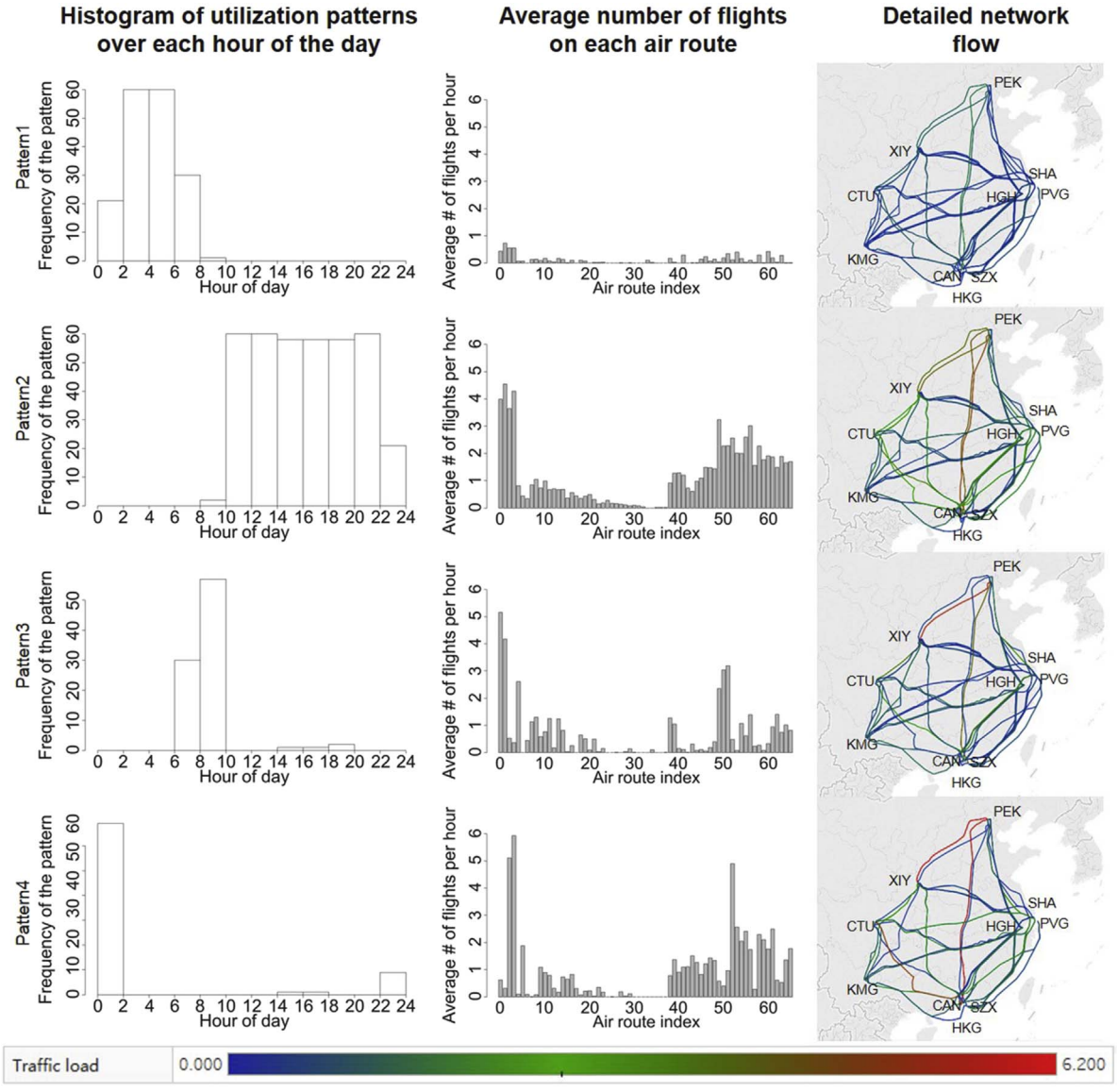


Fig. 19. Details of network utilization patterns in China.

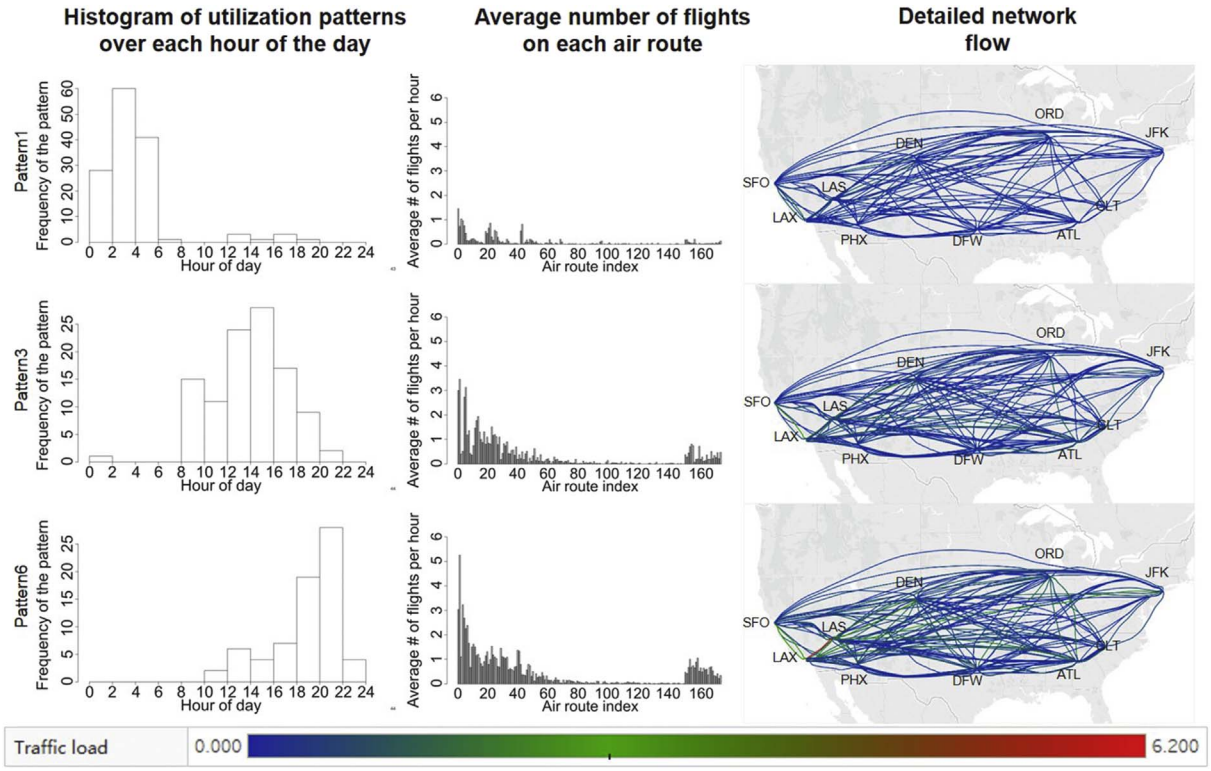


Fig. 20. Details of network utilization patterns in the US.

Acknowledgments

The work was supported by the Hong Kong Research Grant Council General Research Fund Grant (Project No. 11209717), Early Career Scheme (Project No. 21202716), and the National Natural Science Foundation of China Young Scientists Fund (Project No. 71601166). The authors would like to thank Professor Amedeo Odoni and Professor R. John Hansman at Massachusetts Institute of Technology (MIT) for the insightful discussions and constructive comments that helped to improve this study.

Appendix A

Table 7
Information of the top 10 airports in China and the US.

China				US			
IATA	City	Latitude	Longitude	IATA	City	Latitude	Longitude
PEK	Beijing	40.0801	116.585	ATL	Atlanta	33.6366	−84.428
HKG	Hongkong	22.3288	114.194	ORD	Chicago	41.9785	−87.904
CAN	Guangzhou	23.3924	113.299	LAX	Los Angeles	33.9425	−118.407
PVG	Shanghai	31.1434	121.805	DFW	Dallas-Fort Worth	32.896	−97.038
CTU	Chengdu	30.5785	103.947	DEN	Denver	39.8616	−104.672
SZX	Shenzhen	22.6393	113.811	JFK	New York	40.6398	−73.778
KMG	Kunming	24.9924	102.744	SFO	San Francisco	37.6189	−122.375
HGH	Hangzhou	30.2295	120.434	CLT	Charlotte	35.2140	−80.943
SHA	Shanghai	31.1979	121.336	LAS	Las Vegas	36.0801	−115.152
XIY	Xi'an	34.4471	108.752	PHX	Phoenix	33.4342	−112.012

Appendix B

Table 8
Air route index of China.

Index	Air route	Index	Air route	Index	Air route	Index	Air route	Index	Air route
1	PEK-XIY2	2	HKG-PEK2	3	PEK-XIY1	4	HKG-PEK1	5	PEK-PVG3
6	PEK-SHA1	7	PEK-PVG4	8	KMG-SHA1	9	KMG-XIY2	10	CTU-KMG1

11	HGH-SZX2	12	CTU-KMG2	13	HGH-XIY1	14	CTU-HGH2	15	HGH-HKG1
16	HGH-KMG1	17	HGH-SZX1	18	HGH-PEK1	19	HGH-KMG2	20	HGH-XIY2
21	HGH-SZX3	22	KMG-XIY3	23	PVG-SZX1	24	PVG-SZX6	25	HKG-SHA1
26	HKG-KMG1	27	KMG-PVG1	28	HKG-KMG2	24	HGH-XIY3	30	CAN-HKG1
31	HKG-SHA2	32	PVG-SZX4	33	SHA-XIY3	28	CAN-PVG3	35	KMG-PVG2
36	SHA-SZX6	37	SHA-SZX5	38	PVG-SZX3	32	HGH-PEK2	40	CAN-KMG2
41	SHA-SZX3	42	PVG-SZX2	43	KMG-SHA2	36	CTU-HGH1	45	CTU-PVG1
46	SHA-SZX1	47	SHA-SZX2	48	PVG-XIY1	40	HGH-PEK3	50	CTU-SZX2
51	CAN-SHA1	52	CTU-XIY2	53	CAN-CTU1	44	PEK-SHA2	55	SZX-XIY2
56	CTU-XIY1	57	CAN-HGH1	58	PEK-SHA3	59	CAN-SHA2	60	CTU-SHA2
61	HKG-PVG2	62	CAN-XIY1	63	SHA-XIY1	64	HKG-PVG1	65	KMG-SZX1
66	KMG-XIY1								

Table 9
Air route index of the US.

Index	Air route	Index	Air route	Index	Air route	Index	Air route	Index	Air route
1	LAX-SFO2	2	LAS-LAX2	3	DEN-SFO1	4	LAX-SFO3	5	DEN-LAS2
6	LAS-LAX3	7	JFK-LAX1	8	DEN-LAS1	9	JFK-SFO1	10	JFK-LAS7
11	ATL-LAS1	12	ATL-LAX1	13	DEN-PHX1	14	ATL-ORD2	15	ATL-CLT2
16	ATL-DFW2	17	ATL-DFW3	18	ATL-CLT1	19	LAX-PHX3	20	LAS-SFO1
21	LAS-ORD1	22	LAX-ORD1	23	LAX-PHX4	24	LAS-PHX2	25	DEN-PHX2
26	DFW-ORD1	27	DEN-ORD2	28	ATL-DEN3	24	LAS-SFO2	30	DFW-SFO1
31	DEN-DFW2	32	DEN-LAX3	33	LAS-PHX1	28	DEN-DFW1	35	CLT-JFK1
36	CLT-ORD1	37	CLT-JFK2	38	DFW-LAX4	32	ATL-DEN1	40	DFW-ORD2
41	ATL-ORD1	42	LAX-ORD2	43	LAS-LAX1	36	JFK-LAS5	45	DFW-LAX2
46	DFW-LAX3	47	PHX-SFO4	48	DEN-SFO4	40	ORD-SFO2	50	DFW-LAX1
51	DEN-SFO2	52	ORD-PHX3	53	DFW-PHX1	44	DEN-ORD3	55	ATL-JFK2
56	DEN-ORD1	57	JFK-ORD1	58	ATL-LAX2	59	DEN-LAS3	60	ATL-CLT3
61	LAS-SFO4	62	ORD-SFO1	63	LAS-PHX4	64	ATL-DFW1	65	DEN-SFO3
66	CLT-ORD2	67	ATL-LAS2	68	DEN-JFK5	69	DEN-JFK3	70	ATL-PHX3
71	ATL-DFW7	72	JFK-PHX1	73	ATL-CLT4	74	ATL-DFW5	75	DFW-SFO3
76	DFW-JFK1	77	CLT-DFW4	78	DFW-JFK2	79	PHX-SFO5	80	ORD-SFO4
81	JFK-PHX5	82	JFK-PHX4	83	DFW-JFK3	84	ATL-DFW4	85	DEN-ORD5
86	CLT-LAX2	87	LAS-SFO5	88	PHX-SFO2	89	JFK-SFO2	90	ATL-DEN2
91	ORD-SFO5	92	LAS-ORD2	93	DEN-DFW3	94	LAX-SFO1	95	LAX-PHX1
96	DFW-ORD3	97	ORD-SFO6	98	DEN-ORD4	99	LAS-PHX3	100	LAX-PHX2
101	CLT-LAX3	102	JFK-LAS6	103	DEN-LAS4	104	LAS-ORD3	105	CLT-LAS2
106	ORD-PHX2	107	JFK-LAX3	108	ATL-DEN4	109	ATL-LAS3	110	ATL-PHX2
111	DFW-SFO2	112	JFK-PHX3	113	ORD-PHX4	114	ATL-DEN5	115	JFK-LAS1
116	JFK-LAS4	117	CLT-DFW6	118	LAS-SFO3	119	DEN-ORD6	120	DEN-LAS5
121	DEN-JFK1	122	CLT-DEN4	123	CLT-DEN5	124	CLT-DEN3	125	DEN-JFK2
126	JFK-SFO4	127	JFK-PHX6	128	LAX-ORD4	129	DFW-LAX5	130	DEN-JFK6
131	DEN-SFO5	132	ATL-DFW6	133	JFK-LAX4	134	CLT-DEN2	135	CLT-DFW5
136	ATL-LAX4	137	CLT-PHX5	138	CLT-PHX4	139	ORD-PHX5	140	ATL-PHX6
141	DFW-LAS3	142	ORD-PHX6	143	LAS-LAX4	144	DEN-JFK7	145	LAX-ORD5
146	ATL-SFO2	147	DEN-SFO6	148	CLT-DFW7	149	JFK-LAS3	150	CLT-DEN1
151	JFK-ORD2	152	DFW-PHX2	153	ORD-PHX1	154	DFW-LAS1	155	CLT-PHX1
156	LAS-LAX5	157	DEN-JFK31	158	ORD-SFO3	159	ATL-PHX5	160	CLT-ORD3
161	JFK-LAS2	162	CLT-LAX1	163	CLT-SFO1	164	CLT-LAS1	165	JFK-PHX2
166	ATL-SFO1	167	DFW-LAS2	168	ATL-PHX1	169	CLT-DFW2	170	DEN-DFW4
171	CLT-DFW3	172	ATL-PHX4	173	DFW-PHX3				

Appendix C. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jairtraman.2017.12.005.

References

Abdelghany, K.F., Shah, S.S., Raina, S., Abdelghany, A.F., 2004. A model for projecting flight delays during irregular operation conditions. J. Air Transport. Manag. 10 (6), 385–394.

Airports Council International - North America (ACI-NA), 2015. Airport traffic reports. http://www.aci-na.org/content/airport-traffic-reports, Accessed date: 1 August 2017.

Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A., Zou, B., 2010. Total

delay impact study: a comprehensive assessment of the costs and impacts of flight delay in the United States. Technical Report. The Institute for Transportation Studies at the University of California, Berkeley.
Barrat, A., Barthelemy, M., Pastor-Satorras, R., Vespignani, A., 2004. The architecture of complex weighted networks. Proc. Nat. Acad. Sci. U. S. A. 101 (11), 3747–3752.
Beatty, R., Hsu, R., Berry, L., Rome, J., 1999. Preliminary evaluation of flight delay propagation through an airline schedule. Air Traffic Contr. Q 7 (4), 259–270.
Belkoura, S., Cook, A., Peña, J.M., Zanin, M., 2016. On the multi-dimensionality and sampling of air transport networks. Transport. Res. E Logist. Transport. Rev. 94, 95–109.
Bureau of Transportation Statistics (BTS), 2015. Airline on-time statistics and delay causes:on-time arrival performance, national (January-December, 2015). http://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp?display=data&pn=1, Accessed date: 7 June 2017.
Civil Aviation Administration of China (CAAC), 2015a. Civil aviation airport production statistical report. http://www.caac.gov.cn/XXGK/XXGK/TJSJ/201603/t20160331_30105.html, Accessed date: 1 August 2017.
Civil Aviation Administration of China (CAAC), 2015b. Civil aviation industry development statistical report. http://www.caac.gov.cn/XXGK/XXGK/TJSJ/201605/P020160531575434538041.pdf, Accessed date: 7 June 2017.
Conde Rocha Murca, M., DeLaura, R., Hansman, R.J., Jordan, R., Reynolds, T., Balakrishnan, H., 2016. Trajectory clustering and classification for characterization of air traffic flows. In: 16th AIAA Aviation Technology, Integration, and Operations Conference, (June), pp. 1–15.
Cook, A., Blom, H.A., Lillo, F., Mantegna, R.N., Micciche, S., Rivas, D., ... Zanin, M., 2015. Applying complexity science to air traffic management. J. Air Transport. Manag. 42, 149–158.
DeLaurentis, D., Han, E.-P., Kotegawa, T., 2008. Network-theoretic approach for analyzing connectivity in air transportation networks. J. Aircraft 45 (5), 1669–1679.
Dixon, P.M., 2002. Ripley's K Function. Encyclopedia of Environmetrics.
Du, W.B., Zhou, X.L., Lordan, O., Wang, Z., Zhao, C., Zhu, Y.B., 2016. Analysis of the Chinese airline network as multi-layer networks. Transport. Res. E Logist. Transport. Rev. 89, 108–116.
Eckstein, A., 2009, May. Automated flight track taxonomy for measuring benefits from performance based navigation. In: Integrated Communications, Navigation and Surveillance Conference, 2009. ICNS'09. IEEE, pp. 1–12.
Enriquez, M., 2013, June. Identifying temporally persistent flows in the terminal airspace via spectral clustering. In: Tenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2013).
Ester, M., Kriegel, H.P., Sander, J., Xu, X., 1996, August. A density-based algorithm for discovering clusters in large spatial databases with noise. KDD 96 (34), 226–231.
Fiedler, M., 1989. Laplacian of graphs and algebraic connectivity. Banach Cent. Publ. 25 (1), 57–70.
Fleurquin, P., Ramasco, J.J., Eguiluz, V.M., 2013. Systemic delay propagation in the US airport network. Sci. Rep. 3, 1159.
Flightradar24, 2015. How it works. https://www.flightradar24.com/how-it-works, Accessed date: 7 June 2017.
Fricke, H., Schultz, M., 2009, June. Delay impacts onto turnaround performance. In: ATM Seminar.
Gariel, M., Srivastava, A.N., Feron, E., 2011. Trajectory clustering and an application to airspace monitoring. IEEE Trans. Intell. Transport. Syst. 12 (4), 1511–1524.
Gugliotta, Guy, November 16, 2009. An Air-traffic Upgrade to Improve Travel by Plane. The New York Times Retrieved 2009-11-17.
Guimera, R., Mossa, S., Turtschi, A., Amaral, L.N., 2005. The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. Proc. Natl. Acad. Sci. Unit. States Am. 102 (22), 7794–7799.
Haase, P., 1995. Spatial pattern analysis in ecology based on Ripley's K-function: introduction and methods of edge correction. J. Veg. Sci. 6 (4), 575–582.
Hansen, M., 2002. Micro-level analysis of airport delay externalities using deterministic queuing models: a case study. J. Air Transport. Manag. 8 (2), 73–87.
Hansen, M., Nikoleris, T., Lovell, D., Vlachou, K., Odoni, A., 2009, June. Use of queuing models to estimate delay savings from 4D trajectory precision. In: Eighth USA/Europe Air Traffic Management Research and Development Seminar.
He, Y., Zhu, X., He, D.R., 2004. Statistics and developing model of Chinese skyway network. Int. J. Mod. Phys. B 18 (17n19), 2595–2598.
Hsu, K., 2014. China's Airspace Management Challenge. U.S.-China Economic and Security Review Commission Staff Report, November, 2014. . www.uscc.gov/sites/default/files/Research/China%27s%20Airspace%20Management%20Challenge.pdf, Accessed date: 7 June 2017.
Jetzki, M., 2009. The Propagation of Air Transport Delays in Europe. PhD diss. Department of Airport and Air Transportation Research RWTH, Aachen University.
Kodinariya, T.M., Makwana, P.R., 2013. Review on determining number of cluster in K-Means clustering. Int. J. 1 (6), 90–95.
Li, L., Das, S., Hansman, R.J., Palacios, R., Srivastava, A.N., Sep 2015. Analysis of flight data using clustering techniques for detecting abnormal operations. J. Aero. Inf. Syst. 12 (9), 587–598. http://arc.aiaa.org/doi/abs/10.2514/1.I010329.
Li, W., Cai, X., 2004. Statistical analysis of airport network of China. Phys. Rev. E 69 (4), 046106.
Long, D., Hasan, S., 2009, September. Improved prediction of flight delays using the LMINET2 system-wide simulation model. In: 9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO), Hilton Head, SC.
Newman, M.E., 2001. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Phys. Rev. E 64 (1), 016132.
Peterson, M.D., Bertsimas, D.J., Odoni, A.R., 1995a. Models and algorithms for transient queueing congestion at airports. Manag. Sci. 41 (8), 1279–1295.
Peterson, M.D., Bertsimas, D.J., Odoni, A.R., 1995b. Decomposition algorithms for analyzing transient phenomena in multiclass queueing networks in air transportation. Oper. Res. 43 (6), 995–1011.
Pyrgiotis, N., Malone, K.M., Odoni, A., 2013. Modelling delay propagation within an airport network. Transport. Res. C Emerg. Technol. 27, 60–75.
Péter, T., Szabó, K., 2012. A new network model for the analysis of air traffic networks. Period. Polytech. Transp. Eng. 40 (1), 39–44.
Rebollo, J.J., Balakrishnan, H., 2014. Characterization and prediction of air traffic delays. Transport. Res. C Emerg. Technol. 44, 231–241.
Rehm, F., 2010. Clustering of flight tracks. In: Proceedings of the AIAA Infotech@Aerospace.
Shah, A.P., Pritchett, A.R., Feigh, K.M., Kalarev, S.A., Jadhav, A., Corker, K.M., ... Bea, R.C., 2005, June. Analyzing air traffic management systems using agent-based modeling and simulation. In: Proceedings of the 6th USA/Europe Seminar on Air Traffic Management Research and Development.
Sun, X., Wandelt, S., Linke, F., 2015. Temporal evolution analysis of the European air transportation system: air navigation route network and airport network. Transport. B Transport Dynamics 3 (2), 153–168.
Vespignani, A., 2012. Modelling dynamical processes in complex socio-technical systems. Nat. Phys. 8 (1), 32.
Wang, Y., Xu, X., Hu, M., Zhan, J., 2017. The structure and dynamics of the multilayer air transport system. In: The 12th USA/Europe Air Traffic Management R&D Seminar 25-30 June, 2017, Seattle, USA.
Wei, P., Spiers, G., Sun, D., 2014. Algebraic connectivity maximization for air transportation networks. IEEE Trans. Intell. Transport. Syst. 15 (2), 685–698.
World Bank Group (WBG), 2015. Air transport, passengers carried. http://data.worldbank.org/indicator/IS.AIR.PSGR?year_high_desc=true, Accessed date: 7 June 2017.
Xu, N., 2007. Method for Deriving Multi-factor Models for Predicting Airport Delays. George Mason University.
Zanin, M., Lillo, F., 2013. Modelling the air transport with complex networks: a short review. Eur. Phys. J. Spec. Top. 215 (1), 5–21.