Empirical analysis of airport network and critical airports

Cong Wei a,b, Hu Minghua a,b, Dong Bin a,b, Wang Yanjun a,b,*, Feng Cheng c

a College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
b National Key Laboratory of Air Traffic Flow Management, Nanjing 211106, China
c The 28th Research Institute of China Electronics Technology Group Corporation, Nanjing 210000, China

Received 31 August 2015; revised 24 November 2015; accepted 25 December 2015
Available online 23 February 2016

KEYWORDS
Airport network; Critical airport; Spatial correlation; Spectral clustering; Power-law distribution; Power spectra

Abstract Air transport network, or airport network, is a complex network involving numerous airports. Effective management of the air transport system requires an in-depth understanding of the roles of airports in the network. Whereas knowledge on air transport network properties has been improved greatly, methods to find critical airports in the network are still lacking. In this paper, we present methods to investigate network properties and to identify critical airports in the network. A novel network model is proposed with airports as nodes and the correlations between traffic flow of airports as edges. Spectral clustering algorithm is developed to classify airports. Spatial distribution characteristics and intraclass correlation of different categories of airports are carefully analyzed. The analyses based on the fluctuation trend of distance-correlation and power spectrum of time series are performed to examine the self-organized criticality of the network. The results indicate that there is one category of airports which dominates the self-organized critical state of the network. Six airports in this category are found to be the most important ones in the Chinese air transport network. The flights delay occurred in these six airports can propagate to the other airports, having huge impact on the operation characteristics of the entire network. The methods proposed here taking traffic dynamics into account are capable of identifying critical airports in the whole air transport network.

© 2016 Chinese Society of Aeronautics and Astronautics. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

As the important components of air transport network, airports have a great impact on the operational performance of the air transport system. For instance, the European Organization for the Safety of Air Navigation (EUROCONTROL) holds a view that “understanding the variety of airports in Europe, their distribution, their traffic patterns, their aircraft mix, their strengths and their weakness is essential to understanding the strengths of the whole air traffic network”. Given
the geographic locations, demographics and economics, airports play different roles in the air transport system. Special states of an airport, such as flights delay and congestion, can propagate to the other airports, and may finally disseminate over the whole network. There is a “do less for more” effect of identifying critical airports in the network when managing the air transport system. However, there is no existing universal method or criteria to identify whether an airport is critical.

In the past decades, researchers and engineers have long sought the way to assess the performance of airport. In the field of performance assessment, key performance areas (KPAS) are determined based on airport operation characteristics at first. Then key performance indicators (KPIs) are selected according to the KPAS. Capacity, delay, and safety are the most common KPAS that have been widely investigated in the past. As one of the most important KPAS, capacity is normally evaluated based on mathematical models or simulation models. Airport’s efficiency can be reflected by capacities under certain time horizons and operating conditions. Parameters, such as punctuality, predictability, flight efficiency, and on-time performance, can be derived from flight delays. Blom et al. constructed multi-agent models to assess safety risk of airport operation. Safety level of airports can also be reflected through statistics such as rate of accidents/serious incidents/incidents. Integrated comprehensive transportation capability of an airport can be analyzed in terms of instrument flight rules (IFR) flights, average number of annual IFR movements, average number of annual passengers, passengers per IFR movement, annual IFR movements per runway, and so on. While some researchers apply data envelopment analysis to assess productivity, others decompose efficiency into production and efficiencies then apply mathematical methods or criteria to identify whether an airport is critical.

Research on complex networks has been advancing our knowledge of air transport network. The airport network structure has been widely studied both at the international levels and at domestic levels. Network properties have been examined, including degree distribution, average path length, diameter, clustering coefficient, in-out degree correlation, degree–degree correlation, clustering coefficient–degree correlation, betweenness, and rich-club coefficient. Empirical analyses show that air transport networks of Italy, Brazil, India, and the entire world are small-world networks. From complex network perspective, airports with higher degree or betweenness are the important nodes in the air transport network. Although topological structures of the network may reveal the importance of the airport, it is unable to identify the dominant airports due to the lack of consideration of traffic flow and the interactions among airports.

Here we propose a novel method to analyze airport network and to identify the most critical airports. Airport-pair correlations are calculated using traffic flow data. Airport network is then constructed with airports as nodes and airport-pair correlations as edges. All airports are categorized into groups by the proposed spectral clustering algorithm. Network topological structure is illustrated preliminarily. Distance-correlation and power spectra are used to demonstrate whether airport network is a self-organized critical system. We assume that the network criticality originates from a certain group of airports. Finally, the most critical airports are identified by analyzing the distribution of critical state (delay) transition intervals. In a critical system, there should be critical components influencing the whole system.

The rest of this paper is organized as follows. In Section 2, the original data and processing principles are presented. In Section 3, spectral clustering algorithm is developed and applied to classify airports into different categories. Section 4 analyzes the relationships between airport-pair correlation and spatial distance. In Section 5, 1/f spectra is used to probe the long-range memory effects of traffic flow time series. Section 6 gives the transition intervals distributions for the highest delay contribution rate of critical airports or critical airport-pairs. Conclusions are presented in Section 7.

2. Data

Empirical flights data operated within mainland China from January 1st, 2011 to June 30th, 2011 was obtained from the Operation Center of Air Traffic Management Bureau. After data cleaning, there are about two million pieces of record of flight data left for analysis. Each piece of record contains key information of a flight, including departure airport, arrival airport, scheduled time of arrival/departure, and actual time of arrival/departure.

Since the main focus of this study is to identify critical airports in the network, the following airports are out of scope of analysis:

1. International airports (including Hong Kong, Macao, and Taiwan).
2. Airports without traffic flow.
3. Airports with arriving traffic flow or departure traffic flow only.

In total, 161 airports are examined in the presented work.

3. Airport network topology

Clustering algorithms have been used in various airport networks to discover potential community structures or to group airports with the maximum similarity. To analyze airport network topology, airports with similar traffic patterns should be identified according to one or more characteristics. The spectral clustering algorithm is widely used which derives from spectral graph theory. It is assumed that each sample is regarded as node \( V \). We assign weighted value \( W \) to the edge \( E \) according to the similarity among samples. Therefore, there is an undirected weighted graph \( G = (V, E) \). Clustering is then transformed to the optimal partitioning of the graph.

Here we develop a spectral clustering algorithm to study airport network based on the correlations between time series of traffic flow of airports. We can uncover spatial topology of the correlation among airports through the clustering results and compare correlation characteristics of different categories of airports. We assume that the sample set contains \( N \) airports, and let \( f_i(t) \) represent the traffic flow value of the \( i \)th airport at the \( t \)th slot. \( T \) is number of time slots every day. The spectral clustering algorithm is depicted in the following.
Step 1. Calculate correlation $\rho_{ij}$ between each airport-pair’s traffic flow time series $f_i$ and $f_j$ based on pearson correlation coefficient as,

$$
\rho_{ij} = \frac{\sum (f_i(t) - \bar{f}_i)(f_j(t) - \bar{f}_j)}{\sqrt{\sum (f_i(t) - \bar{f}_i)^2} \sqrt{\sum (f_j(t) - \bar{f}_j)^2}}
$$

where $\rho_{ij} \in [-1, 1]$. The flow time series of each airport lasts six months and has a temporal resolution of 15 min. 

Step 2. Form the similarity matrix $S \in \mathbb{R}^{n \times n}$, where $S_{ij} = \exp[-\sin^2(\arccos(\rho_{ij})/2)/\sigma^2]$, $\sigma$ is a scaling parameter (which is usually taken as $\sigma = 1$)\(^{31,32}\). 

Step 3. Define $D$ to be the diagonal matrix: $D_{ii} = \sum_{j=1}^n S_{ij}$, so that the Laplacian matrix can be defined as $L = D^{-1/2}SD^{-1/2}$. 

Step 4. Calculate the eigenvectors $\{v_0, v_1, v_2, \ldots, v_{n-1}\}$ and corresponding eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n-1}$ of $L$. 

Step 5. In matrix perturbation theory, the gap between the $k$th and $(k+1)$th eigenvalues is defined as eigengap.\(^{33}\) Here we form eigengap series $\{g_1, g_2, \ldots, g_{n-1}\}$ based on obtained eigenvectors. Then we find the first maximum value of the eigengap series in the sequence. The subscript of the value is the number of clusters: $K$. The larger the maximum value of eigengap is, the more stable the subspace formed by selected $k$ number of eigenvectors is.$^{33}$

Step 6. Select $k$ largest eigenvectors $\{v_1, v_2, \ldots, v_k\}$ corresponding to eigenvalues and construct new matrix $V' = [v_1, v_2, \ldots, v_k]$ with eigenvectors. Form $V'$ by normalizing each row of matrix $V$: 

$$
V'_y = \sqrt{\frac{V_y}{\sum y^2}}
$$

Step 7. Each row of $V'$ is viewed as a point. Then k-means clustering algorithm is performed on these points. 

All airports in the investigated dataset are clustered into 6 categories. Clustering results are presented in Table 1, in which airports are given by International Civil Aviation Organization (ICAO) codes. 

Fig. 1 shows the spatial distribution of 6 categories of airports respectively. While the purple circles indicate the airports in the category, the black circles represent other airports in the network. The number of airports in each category varies from 18 to 34. Compared with partitioning clustering results, spatial distributions of spectral clustering results are overlapping. Multiple regions are involved in the distribution of each category. The overall range is large though most airports are concentrated in certain regions. The busiest airports in China mainland are classified into category $C_6$ (for instance, Beijing Capital international airport-ZBAA, Shanghai Pudong international airport-ZSPD, Shanghai Hongqiao international airport-ZSSS, Shenzhen Bao’an international airport-ZGSZ, Guangzhou Baiyun international airport-ZGGG, Chengdu Shuangliu international airport-ZUUU, etc.). Most of these airports are located in the southeastern coastal regions. Few airports are found to be in the central and northern parts of China. In western China, only urumchi diwopu international airport is in $C_6$. Compared with other categories, airports in $C_4$ and $C_5$ are scattered nationwide. Airports of $C_6$ are with the highest intraclass correlation, much higher than other categories. Airports of $C_4$ are of the lowest intraclass correlation. All the five categories of airports show similar intraclass correlation fluctuation, except that $C_6$ shows comparatively wider intraclass correlation fluctuation (Fig. 2).

Clustering results indicate that air transport network has the following three features: (1) Since spectral clustering is based on the correlations between flow time series, airports in the same category have similar traffic features. (2) Traffic features are influenced by geographical positions. For example, most of the cities in $C_6$ are capital cities of the provinces or municipalities. (3) Similarities among airports are not affected by the geometrical distances. Therefore, airport network is characterized by a complex multi-core topology.

### Table 1

<table>
<thead>
<tr>
<th>Category</th>
<th>Airport</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$</td>
<td>ZBDD, ZBNY, ZBOW, ZBSH, ZGKL, ZGMX, ZGWZ, ZGHZ, ZHES, ZHNY, ZHHK, ZLIC, ZLLL, ZLZW, ZPDL, ZPJH, ZPMS, ZSLG, ZSLO, ZSSH, ZSWH, ZSXX, ZUJJ, ZULB, ZUMY, ZUQJ, ZUTR, ZUYB, ZWCM, ZHYE, ZYL</td>
</tr>
<tr>
<td>$C_2$</td>
<td>ZBCZ, ZBUL, ZBYC, ZGSD, ZHXF, ZLYA, ZPBS, ZPWS, ZPZT, ZSAQ, ZSCN, ZSFY, ZSGS, ZSJ, ZSLQ, ZSRL, ZUNZ, ZYCY, ZYJM, ZYIZ</td>
</tr>
<tr>
<td>$C_3$</td>
<td>ZBWX, ZGBS, ZGJC, ZLXN, ZLYL, ZPDQ, ZPLC, ZPLJ, ZPSM, ZSDD, ZSGW, ZSY, ZSW, ZSZS, ZUBD, ZUKD, ZUTC, ZUXC</td>
</tr>
<tr>
<td>$C_4$</td>
<td>ZBSD, ZBER, ZBDH, ZGBH, ZGCD, ZGDY, ZGOW, ZLGY, ZSNG, ZSJ, ZSJ, ZSJ, ZSNT, ZSTX, ZSWF, ZSWX, ZSYN, ZUXD, ZULZ, ZUNP, ZUY1, ZUZH, ZWAT, ZWBL, ZWKC, ZWKN, ZWNL, ZWTC, ZWYN, ZYBS, ZDYQ, ZYJX, ZYM, ZYQQ, ZYJJ</td>
</tr>
<tr>
<td>$C_5$</td>
<td>ZBCF, ZBLA, ZBMZ, ZBSN, ZBUH, ZGJZ, ZHL, ZHIC, ZLHD, ZLGM, ZLYS, ZSCG, ZSQZ, ZUGU, ZULS, ZUNC, ZUWX, ZWAK, ZWHM, ZWKL, ZWSH, ZWTN, ZYDD, ZYMH</td>
</tr>
<tr>
<td>$C_6$</td>
<td>ZBAA, ZBHH, ZBSJ, ZBTJ, ZBYN, ZGG, ZGHA, ZGNN, ZGSZ, ZHCC, ZHH, ZJSY, ZLY, ZPPP, ZSAM, ZSFZ, ZSHC, ZSJN, ZSNB, ZSNJ, ZSOF, ZSPD, ZSQD, ZSS, ZSWZ, ZSYT, ZUCK, ZUGY, ZUUU, ZWWW, ZYCC, ZYHB, ZYTX</td>
</tr>
</tbody>
</table>

### 4. Spatial correlation

In order to further investigate spatial fluctuation of the correlations between airport-pairs, fluctuation trends of airport-pair correlation along with distance change are analyzed. Here we calculate the distance $r$ between any two airports and the corresponding pearson correlation $C(r)$. $r$ are the correlation coefficients $\rho_{ij}$ of the airport-pairs whose distance between the airports is $r$. Fig. 3 shows the relationship between $C(r)$ and $r$. Error bars indicate the standard deviations of $\rho_{ij}$ of the airports with the same $r$. 


As shown in Fig. 3, spatial correlation is divided into two ranges: (1) When the distance $r < 1585$ km, the correlation $C(r)$ between airport-pairs does not decay while increasing distance $r$ (the slope is $-0.04$); (2) When $r \geq 1585$ km, there is a power-law relationship between $C(r)$ and $r$: $C(r) = r^{-\beta}$, $\beta = 0.66$. It is suggested that there exists power-law correlation at the level of the airport network. Long-range correlations usually emerge in systems with local interactions in the vicinity of a critical point.31 The interactions between critical points and the effect of critical points on other points will disseminate in the entire network. Critical points are determined by the distance-correlation fluctuation trends of 6 categories of airports.

The distance-correlation trends of each category of airports are plotted in Fig. 4, while the exponents of six categories lie within $[0.17, 1.00]$. In previous section, correlation results show that intraclass correlation between airport-pairs of $C_6$ ($\approx 0.5$) is much higher than the airports in other categories (Fig. 2). The correlation, however, does not decay when increasing distance. In contrast, airports in $C_4$ exhibit power law decay with exponent higher than that of the entire network (exponent: $1 > 0.66$), while they have the lowest intraclass correlation. Except that $C_5$ is with the smallest exponent, intraclass correlations in other classes decay faster than those in the entire network. According to the traffic characteristics and correlation of $C_6$, it can be concluded that airports of $C_6$ should be critical points in the network. They share similar traffic features, and their strong mutual effect do not weaken when increasing distance. On the contrary, $C_4$ contains small regional airports with light traffic flow. These airports are influenced by critical ones in the network and propagate the influence to other relevant airports. Therefore, the correlation decays as distance increases. The decay may be faster due to the absence of critical points in these categories.

5. Temporal correlation

Characteristics of flow time series can be explored with the power spectra from another perspective. Power spectra analysis is a frequency domain analysis method based on Fourier
transform. Power Spectra answers the question “which frequencies contain the signal’s power?”. In power spectra analysis, the total energy of a time series is decomposed into components at various frequencies. The prominent period hidden in the time series can be judged according to variance contributions of various frequency waves. The corresponding frequency of the maximum value of the power spectra is the prominent frequency of the time series.\(^{34,35}\)

The data of each airport lasts NUM days, with \(T\) time slots per day. Let \(yf_i\) be the traffic flow time series of the \(i\)th airport with the length of \(T/\text{NUM}\). The classical spectral estimation method is described as follows.

**Step 1.** Calculate the average traffic flow time series of all airports in the network.

\[
F(l) = \frac{1}{N} \sum_{i=1}^{N} yf_i(l) \quad l = 1, 2, \ldots T \times \text{NUM}
\]

**Step 2.** Suppose that the average flow time series is a set of stationary random signals. Then, we calculate the autocorrelation function of the time series as

\[
R(m) = \frac{1}{T \times \text{NUM} - |m|} \sum_{l=1}^{T \times \text{NUM} - |m|} F(l)F(l+m) \quad |m| \leq T \times \text{NUM}
\]

**Step 3.** Based on Wiener-Khinchin theorem, estimate the power spectra by taking the Fourier transform of autocorrelation function.

\[
P(e^{j\omega}) = \sum_{m=-\infty}^{\infty} R(m)e^{-j\omega m}
\]

Power spectra of the average flow time series of entire network shows a clear power-law scaling behavior with scaling exponent \(\alpha = -0.90\) which is approximately equivalent to \(1/f\). Research has shown that \(1/f\) spectra is an important phenomenon of self-organized criticality. Besides, \(1/f\) spectra indicates that long-range memory effects of time series transmit within a wide range of time scales. The results obtained here clearly indicate that air transport system is a critical system.

In the following, we further investigate the power spectra of 6 categories of airports. Power exponent of \(C_6\) is \(-0.9\), while the exponents of the other categories lie within \([-0.81, -0.66]\), which are shown in Fig. 5. Therefore, the average traffic time series of \(C_6\) can be interpreted as \(1/f\) spectra suggesting self-organized criticality. It can be concluded that the airport network is with self-organized criticality because of the 6th category of airports \(C_6\) which dominates the overall characteristics of the entire airport network. This is in accordance with previous results on the evolution of the spatial correlation. The airports in \(C_6\), as key nodes of the airport network, have a great impact on the entire network.

6. Distribution of state transition intervals

Even for airports in \(C_6\), they play different roles in the network. As a key indicator of airport performance analysis, delay often occurs in special or extreme operational conditions, such as military activities, severe weather, and major security concerns. Therefore, delay can be, to a greater extent, considered as an indicator of the critical operation state of airport. To uncover the core airports in \(C_6\), we turn to the study.
of state transition based on delay. The brief idea is as follows. Firstly, we calculate the number of delayed flights for each airport in $C_6$ with temporal resolution of 15 min. Then, we calculate contribution rate of airport as the rate of number of delayed flights of each airport to the total number of delayed flights of the network in each time slot. The airports are sorted in descending order according to their contribution rates. Finally, the airport with the highest contribution rate can be found in each slot, and then there will be a new time series formed by the highest airports, for instance, $\{\text{ZBAA, ZGGG, ZUUU, ZSPD, ZBAA, ZGSZ}\}$. When one airport contributes most to the delay at network level, it is regarded as the core of the network. According to the time series, the intervals between the same airports are obtained to estimate the distribution of critical state transition. As shown in Fig. 6, the state transition intervals of these 6 airports do exhibit power law decay, while the others do not (‘$\text{Pr}(\cdot)$’ in the Fig. 6 represents the probability). The exponents of these 6 airports lie within $[0.75, 3.11]$.

7. Conclusions and future work

(1) A novel aviation network is constructed with airports as nodes and the correlations between traffic flow of airport-pairs as edges. All airports are classified into 6 categories by the proposed spectral clustering algorithm. Among the 6 categories, airports in $C_6$ are of the highest intraclass correlation, and they have quite similar features in terms of traffic characteristics and geographical locations.
It is found that the spatial correlations among airports exhibit power law decay when the distance between airports is larger than 1585 km. Besides, 1/f spectra is observed in average traffic flow time series of airport network. Spatial and temporal correlations analyses show that air transport system is a critical system with self-organized criticality. After further comparison of spatial and temporal correlation among 6 categories of airports, a group of critical airports have been identified from $C_6$. They are the primary source of self-organized criticality.

Airport’s delay is analyzed with the assumption that the critical state of the airport network can be reflected by airport’s delay. The distributions of state transition intervals for airports in $C_6$ are estimated based on the highest delay contribution rate. There are 6 airports with state transition intervals showing power-law distributions. Beijing international airport (ZBAA) is with the highest scaling exponent. It can be concluded that these 6 airports are the most critical ones in the network. The frequently occurred delay in these airports will propagate to other airports. Thus, the performance of these airports has a great impact on the performance of the entire network.

The identification of critical airports in the critical airport network has potential implications for the management of air transport system, for instance, to improve operation efficiency of the system. We have presented the method to identify critical airports in the network. Further efforts will be given to the comparative study of identified critical airports, to profoundly analyze their similarities and differences. The presented method here will be performed with the data of other countries to validate whether their aviation networks are of self-organized criticality.

Acknowledgements

The authors thank anonymous reviewers for their critical and constructive suggestions and comments on the manuscript. This work was co-supported by the National Natural Science Foundation of China (No. 61304190), the Fundamental Research Funds for the Central Universities of China (No. NJ20150030), and the Natural Science Foundation of Jiangsu Province of China (No. BK20130818).

References


Cong Wei is a Ph.D. student in College of Civil Aviation, Nanjing University of Aeronautics and Astronautics (NUAA). He received his B.S. degree from the same university in 2011. His area of research includes air traffic performance evaluation and air traffic complexity theory.

Wang Yanjun is an associate professor in College of Civil Aviation, Nanjing University of Aeronautics and Astronautics (NUAA). He is the director of the Office of International Affairs, deputy director of Airspace Assessment Lab of the National Key Laboratory of Air Traffic Flow Management. He received the B.S. degree in Air Traffic Management and Dispatch and the M.S. degree in Transportation Engineering from NUAA in 2004 and 2007, respectively. He was a doctoral researcher at EUROCONTROL Experimental Center from 2009 to 2012, and received his Ph.D. from Télécom ParisTech (ENST), France in 2012. His current research interests cover human factors, complex systems modeling and analysis, and data science.