Outlier Analysis of Airport Delay Distributions in US and China

Max Z. Li*†, Karthik Gopalakrishnan*†, Yanjun Wang‡† and Hamsa Balakrishnan†
†Department of Aeronautics and Astronautics,
Massachusetts Institute of Technology
Cambridge, MA, USA
Email: {maxli, karthikg, yanjun, hamsa}@mit.edu
‡College of Civil Aviation
Nanjing University of Aeronautics and Astronautics
Nanjing, 211106, China

***Abstract*—Outlier detection is a key component of several machine learning approaches. However, many existing techniques, especially for multi-dimensional signals, are not interpretable and do not explain why a specific classification was assigned to a particular data point. Another limitation is that most methods only consider the magnitude or intensity of the signal, and not its spatial distribution. We present a spectral approach to identify outliers based on the spatial distribution of a signal across the nodes of a graph without any explicit assumptions on the underlying probability distribution of the signal. By applying these techniques to airport delays, we not only identify outliers in the spatial distribution of delays, but also gain insights into the delay dynamics. Specifically, we compare spatial delay distributions in the US and China during the period 2012-17, and identify several interesting characteristics pertaining to critical airports for outlier detection. We characterize typical variabilities in the delay distributions, and the frequency of occurrence of outliers. Our results highlight the differences between the operational dynamics of the US and Chinese air transportation systems, and contribute to performance benchmarking between different airspace systems.**

***Index Terms*—US/China air transportation, flight delays, Graph Signal Processing, outlier analysis, aviation disruption**

I. Introduction

Air traffic delays lead to economic losses, environmental impacts, and customer dissatisfaction. They are a major cause of concern for airlines, airport operators, and air navigation service providers. In recent years, vast quantities of aviation data have been analyzed using state-of-the-art machine learning techniques in order to address the problems of predicting, mitigating, and recovering from flight delays and irregular operations. The focus of our work is to develop specialized outlier detection methods for multi-dimensional signals which frequently occur in air transportation. These and other related methods are an integral part of the statistical analysis and machine learning workflow in order to understand, assess, and manage airport delay patterns at a system-wide scale.

Our work characterizes and compares the *spatial delay patterns* in the US and mainland China airport networks. There are several factors that motivate our choice of these two regions. In terms of similarities, both networks cover vast geographical extents; for example, the longest intra-region flight for both networks is about 6 hours. Both networks carry a significant volume of passenger and cargo (777 million passengers in the US [1] and 126 million passengers in China during 2018), experience severe delays (19.3% flights delayed in the US [1] and 19.9% in China during 2018 [2]), and have limited strategic measures such as airport slot controls that mitigate long term demand-capacity imbalances. There are also significant differences in the rate of growth (passenger growth of 4.9% in the US [1] and 10.2% in China [3]), maturity of the infrastructure, seasonal effects and weather patterns, and air traffic management procedures. For example, China has a limited number of fixed airways that are regularly restricted due to military activities [4], [5]. These similarities and differences provide an interesting opportunity for comparisons, in particular using data-driven methods to study the impact of these factors on operational performance.

A. Motivation

Outlier detection is a well-studied problem in statistical analysis and machine learning. The primary objective of these methods is to identify whether a data point belongs to a common group or is an anomalous (or outlier) observation. This classification into outliers is particularly useful in two settings. First, pre-processing raw data to remove outliers is important for algorithms with excessive sensitivity to outliers in training data (e.g., regression analysis [6], AdaBoost [7]). Second, in safety-critical applications in which machine learning models interact with the physical world (e.g. self-driving cars, autonomous UAV navigation [8]) or face adversarial inputs, the statistical models may need to be accurate, or robust to outliers and edge cases. It is desirable for outlier detection methods to be interpretable, i.e., to be able to tell why a particular data point was classified as an outlier. All of this is part of a broader interest in the research community towards building explainable artificial intelligence systems in

*co-first and corresponding authors.

978-1-7281-5380-3/20/$31.00 ©2020 IEEE

order to improve acceptability and encourage wider adoption of these methods.

Outlier detection for signals on a graph possesses unique challenges. In terms of airport delays, this means that we are considering the delays at different airports in the network during a fixed time interval as one data point. Traditional outlier detection methods classify a data point as an outlier based on the magnitude of these delays, either by looking at individual airports or the total delay across the entire system. However, where the delays are spread, i.e. the *spatial distribution* of delays across various airports, is not considered in previous analyses. This means that an interesting and operationally anomalous class of delay days could not be explicitly identified. For example, consider two geographically proximate airports in the New York City region (e.g. LGA and JFK) that historically have highly correlated airport delays as they tend to be affected by the same weather systems. A situation where LGA and JFK experience very different delay magnitudes is thus unexpected, and may warrant further investigation. However, this requires firstly the identification of such scenario, and secondly an interpretation for why this scenario occurred. The motivation of our work is to formalize the notion of outliers based on the spatial distribution of a graph signal, and apply this for airport delay networks to obtain operational insights.

B. Literature review

We will review and discuss two areas of past research: outlier detection methodologies for multi-dimensional or graph-supported data, and prior aviation studies with international comparisons.

A classic approach to outlier detection in multi-dimensional data sets is to assume a statistical distribution for the data, and identify points that are at the tail(s) of the distribution [9], [10]. However, in general, estimating such a distribution is not trivial, especially in large-scale settings such as signals on graphs with a large number of nodes. Another popular approach is to cluster the data points, and classify points that do not belong to any cluster as an outlier [11].

However, a limited number of methods exist that are specific for graph-supported signals; one example include information-theoretic approaches [12], [13]. Graph signal processing (GSP) [14], and specifically the use of the Total Variation (TV) metric for graph signals have shown promise as a way to capture the spatial distribution of graph signals [15]. In this paper, we modify our previous approaches in [16], [17] to identify outliers in spatial delay distributions while relaxing symmetric and Gaussianity assumptions on the underlying distribution of the graph signals. Furthermore, no comparisons between spatial delay distributions in different air transportation networks were performed in [16] or any other prior work.

Comparative analysis of airport and airspace performance has received recent attention due to the rapid expansion of domestic and international aviation in several parts of the world [18]–[20]. Previous analyses are typically focused on the connectivity of airports in terms of non-stop flights, or the nature of historical flight trajectories [4], [21]. Studies have also analyzed the topology of various air transportation networks to identify key nodes [22]–[26] and edges [27], evaluate network stability [28], and characterize the role of airspace structure on flight conflicts [29]. For a comprehensive overview of this research direction, see [30]. Our work introduces a novel metric for characterizing spatial distributions of airport delays, and quantitatively comparing the US and Chinese airport networks from this perspective.

C. Contributions of work

In this work, we use a novel methodological framework for identifying spatial delay distributions in airport networks. We conduct a data-driven comparison of the operational characteristics of the US and China airspace networks, and identify outlier test cases that can be used to benchmark and evaluate machine learning applications. The main findings and contributions of our work are as follows:

1) *Identifying critical subsets of airports.* We find that the delay distributions at a small set of geographically proximate US East Coast airports are indicative of whether the system-wide delay distribution is expected or unexpected. By contrast, such an equivalent set of airports in China is spread over a large portion of the Chinese airport network (Section III-B).
2) *Variability in spatial delay distributions.* The baseline variability in terms of spatial distribution for delays in China is higher than in the US (Section III-C).
3) *Scale versus spatial delay distribution outliers.* We see that even though the baseline delay magnitudes are higher in China, the US experiences more outlier days in terms of delay magnitudes. On the other hand, China experiences more outliers in terms of spatial delay distributions, even though it has a higher baseline level of variability in its spatial delay distribution. In terms of temporal trends, we find significantly more outliers in winter months for both China and the US (Section IV).
4) *Framework for interpreting outliers.* Our final contribution is a two-step process for interpreting outliers that considers both the type of outlier and the operational implications. We demonstrate this framework by analyzing specific outlier days in China and the US. Our analysis not only identifies the specific airport delays behind the outlier classification, but also corroborates findings with operational factors (Section V).

Finally, we emphasize that our methods are broadly applicable, and can be applied to detect outliers in any graph signal or multi-dimensional data set.

II. Data and methodology

A. Graph signal processing

Consider a set with M observations of a multi-dimensional signal $\mathcal{O}_M = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(M)}\}$, where $\mathbf{x}^{(k)} = \left[x_1^{(k)}, x_2^{(k)}, \ldots, x_N^{(k)}\right]^\mathsf{T} \in \mathbb{R}^{N\times 1}$ represents one such observation. This signal is assumed to be supported on the nodes of

a graph $G = (V, E)$, where $|V| = N$ is the set of nodes and E is the set of edges connecting the nodes. In this paper, the nodes are airports, and the edges are undirected and weighted by the sample Pearson correlation coefficient of the signal between the nodes. The edge weights are represented using an adjacency matrix $A \in \mathbb{S}^{N\times N}$, where the element a_{ij} is given by

$$a_{ij} = \frac{\sum_{k=1}^{M}\left(x_i^{(k)} - \hat{\mu}_i\right)\left(x_j^{(k)} - \hat{\mu}_j\right)}{\sqrt{\sum_{k=1}^{M}\left(x_i^{(k)} - \hat{\mu}_i\right)^2}\sqrt{\sum_{k=1}^{M}\left(x_j^{(k)} - \hat{\mu}_j\right)^2}}, \quad (1)$$

where $\hat{\mu}_i = \frac{1}{M}\sum_{k=1}^{M} x_i^{(k)}$ is the empirical mean of the signal at node i. Note that since the graph is weighted by the correlation coeffecient and is thus undirected, the adjacency matrix is symmetric, i.e. $A = A^\intercal$. When denoting a generic graph signal and not any particular element in $\mathcal{O}_M$, we drop the superscript (k) for brevity and simply denote it as $\mathbf{x}$.

The *combinatorial graph Laplacian* is defined as $\mathcal{L} = D - A$, where D is a diagonal matrix with $d_{ii} = \sum_j a_{ij}$. Since $\mathcal{L}$ is a row stochastic matrix, we have that 0 is an eigenvalue, and its multiplicity is the number of connected components of the graph. Assuming a fully-connected graph, we arrange the eigenvalues in ascending order without loss of generality as $0 = \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_N$, and denote their corresponding orthonormal eigenvectors as $\{v_1, v_2, \ldots, v_N\}$. This set of eigenvectors forms a basis for graph signals in $\mathbb{R}^{N\times 1}$; we will refer to this set as the set of *eigenvector modes*, or simply *modes* for the remainder of the paper. Any graph signal $\mathbf{x}$ can be written as a linear combination of these orthogonal eigenvector modes.

We now define some standard terminology from GSP, a graph-supported analogue of discrete signal processing:

Definition 1. The *Graph Fourier Transform (GFT)* of a signal $\mathbf{x} \in \mathbb{R}^{N\times 1}$ supported on a graph with eigenvector modes $\{v_1, v_2, \ldots, v_N\}$ is defined as $(\alpha_1, \alpha_2, \ldots, \alpha_N)$, where $\alpha_i = \mathbf{x}^\intercal v_i$. It represents the component of that signal along each of the eigenvector modes.

Definition 2. The *spectral energy* of a signal $\mathbf{x} \in \mathbb{R}^{N\times 1}$ with GFT $(\alpha_1, \alpha_2, \ldots, \alpha_N)$ is defined as $\left(\alpha_1^2, \alpha_2^2, \ldots, \alpha_N^2\right)$. This implies that the *total energy* can be equivalently written as the sum of the spectral energy components, i.e. $\|\mathbf{x}\|_2^2 = \sum_{i=1}^{N} \alpha_i^2$.

Definition 3. The *Total Variation* $\mathrm{TV} : \mathbb{R}^{N\times 1} \times \mathbb{R}^{N\times N} \to \mathbb{R}$ metric of a graph signal $\mathbf{x}$ with respect to a graph Laplacian $\mathcal{L} \in \mathbb{R}^{N\times N}$ associated with an adjacency matrix $A = [a_{ij}] \in \mathbb{R}^{N\times N}$ is defined as

$$\mathrm{TV}(\mathbf{x}, \mathcal{L}) = \mathbf{x}^\intercal \mathcal{L}\mathbf{x} = \frac{1}{2}\sum_{i\neq j} a_{ij}(x_i - x_j)^2. \quad (2)$$

For notational brevity, we will write $\mathrm{TV}(\mathbf{x}, \mathcal{L})$ as $\mathrm{TV}(\mathbf{x})$ when the graph Laplacian is unambiguous.

The TV measures the smoothness of the graph signal $\mathbf{x}$; the higher the TV, the greater the difference in the graph signal magnitudes across adjacent nodes. Graph signals with a low TV are said to be *smooth*. The TV is small when the signals are relatively the same at adjacent nodes, or the edge weight connecting nodes with differing signal values are small. Note that if the correlation between any two nodes is non-negative, then the TV is also non-negative for any signal $\mathbf{x}$.

Proposition 1. *Suppose the GFT of* $\mathbf{x}$ *is* $(\alpha_1, ..., \alpha_N)$. *Then, the following two statements are equivalent:*
(i) $\mathrm{TV}(\mathbf{x}) = \mathbf{x}^\intercal \mathcal{L}\mathbf{x}$, *and (ii)* $\mathrm{TV}(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i^2 \lambda_i$.

Proof. The proof follows directly from the definition of the graph Laplacian and the GFT. □

Proposition 1 relates the TV of the signal to its GFT and spectral energy distribution. Eigenvectors corresponding to lower eigenvalues (recall our ascending ordering of eigenvalues, so we can equivalently talk about eigenvalue magnitudes and their index) are *smoother*, meaning they have a higher percentage of their spectral energies in the lower modes, and consequently a smaller TV. On the other hand, the higher eigenvector modes represent more energetic signal distributions, meaning they contribute more to the spectral energy and result in higher TV. We will refer to these as *high-energy* or *energetic* eigenvector modes. The eigenvalues of the graph Laplacian are a direct measure of the smoothness of the signal and determine the contribution of that particular mode to the TV. Thus, if a signal typically has a greater fraction of its spectral energy at modes corresponding to a higher index, this indicates that the signal has higher TV.

B. Delay data and preprocessing

Airport delay data is obtained for the US from the Aviation System Performance Metrics (ASPM) database maintained by the Federal Aviation Administration (FAA), and for China from the Operations Monitoring Center of the Civil Aviation Administration of China (CAAC). We analyze data from 2012-17, and restrict the airport network to 30 airports in both cases. The US airports are chosen based on the FAA Core 30 list, and for China based on the traffic volume (see Table III and Figure 10 in the Appendix for a list and map of all 30 airports for both networks). We eliminate cancelled and diverted flights from both data sets and construct a graph signal of airport delays for each day in the 5-year period for both countries. The graph signal at each airport is equal to the total delay at that airport. The total delay for an airport is defined as the sum of the mean inbound and outbound delays in minutes seen at every hourly interval for the day, where a day is defined as a 24-hour period. After the data preprocessing, we obtain $M = 2,192$ graph signals $\mathbf{x} \in \mathbb{R}_{\geq 0}^{30\times 1}$ for both networks.

The edge weights a_{ij} in both graphs are computed as the correlation coefficient between the delay signals at airports via Equation 1; we use the same graph abstraction in prior work [16]. We note that all the correlations are strictly positive, thus both graphs have only one connected component. Hence, the graph Laplacian will have only one zero eigenvalue, with all real eigenvectors and eigenvalues. We refer readers to [17] for a discussion on graphs with mixed and negative correlations.

C. Outlier detection methodology

We now formalize and define our notions of outliers for graph signals. We define a non-parametric approach based on the interquartile range (IQR) to define the outlier bounds. We explicitly account for skewed distributions by adjusting the bounds based on the *medcouple* statistic [31], [32]. This turns out to be critical in our application, where it would not be appropriate to assume symmetry in the TD or TV distributions since both are non-negative quantities. Given a univariate sample set $\{y_1, y_2, \ldots, y_K\}$, we define the value of its first quartile as Q_1, the median of the sample set as Q_2, and the third quartile value as Q_3. The IQR is the difference between the third and first quartile, i.e. $IQR = Q_3 - Q_1$. The medcouple statistic is defined as follows:

$$MC = \underset{y_i \leq Q_2 \leq y_j}{\text{med}} h(y_i, y_j), \tag{3}$$

note that we take a median conditioned on $y_i \leq Q_2 \leq y_j$, where for all $y_i \neq y_j$, the kernel function $h(\cdot, \cdot)$ is given by:

$$h(y_i, y_j) = \frac{(y_j - Q_2) - (Q_2 - y_i)}{y_j - y_i}. \tag{4}$$

The medcouple statistic is a robust measure of skewness [31], and will be used to adjust the IQR in order to form outlier bounds for asymmetric distributions [32]. In our analysis, all distributions are skewed to the right, i.e. $MC > 0$. This is expected due to the non-negativity of airport delays. Thus, we only utilize the adjusted box-plot formulas for the right-skewed case. We refer readers to [32] for the left-skewed case.

For simplicity, we restrict ourselves to the 1-norm of graph signals when we talk about the TD at an airport. Furthermore, we assume non-negative signals at each node, which is reasonable since airport delays are always a non-negative quantity. Thus, $\|\mathbf{x}\| = \|\mathbf{x}\|_1 = \sum_i x_i$. We will use this definition of the norm of a graph signal vector for the rest of the paper.

The first notion of a graph signal outlier that we will define identifies *outliers in scale*, meaning that $\mathbf{x}$ is an outlier based solely on the magnitude of the TD, and not the distribution. This notion of an outlier is the simplest, most commonly used, and most intuitive. This definition states that the graph signal is an outlier in scale if the 1-norm of the graph signal lies outside a skew-adjusted empirically determined central region:

Definition 4. A data point $\mathbf{x}$ is classified as an *outlier in scale (OIS)* if

$$\|\mathbf{x}\| \notin \left[\underline{\Omega},\, \overline{\Omega}\right],$$

where the lower bound $\underline{\Omega} \coloneqq Q_1 - 1.5e^{-4\times MC} IQR$, the upper bound $\overline{\Omega} \coloneqq Q_3 + 1.5e^{3\times MC} IQR$, and the IQR and $MC \geq 0$ are defined on the set of 1-norms $\boldsymbol{\mathcal{V}}_{\|\mathbf{x}\|} = \left\{\mathcal{V}_{\|\mathbf{x}\|,1} = \left\|\mathbf{x}^{(1)}\right\|, \ldots, \mathcal{V}_{\|\mathbf{x}\|,M} = \left\|\mathbf{x}^{(M)}\right\|\right\}$.

The idea behind identifying outliers based on the spatial distribution of the graph signal relies on the edge weights being the historical correlations between the airport delays. If the delays at airport i and j have been historically correlated, then $a_{ij} \to 1$ and it is *expected* that on any given day, the delays at both airports will be high, or the delays at both airports will be low. This would mean that the contribution of this airport pair to the TV, $a_{ij}(x_i - x_j)^2$, will be small as the difference term would be small. On the other hand, if the delays are *unexpected*, meaning that the delays at one airport is abnormally higher than the other, or vice versa, the contribution to the TV is large. Now consider the other extreme, where the historical correlation between an airport pair is small, i.e. $a_{ij} \to 0$. This means that there is no implicit expectation regarding relative delay magnitudes between these two airports, and that one would expect them to vary in an uncorrelated manner with respect to each other. Thus, even though the squared difference between their delay signals, i.e. $(x_i - x_j)^2$ could be high, their contribution to the TV is always small due to a_{ij} being closed to 0.

This relationship between a_{ij} and $(x_i - x_j)^2$ highlights the utility of TV as a metric to identify *unexpected* spatial distributions of delays across the entire network. When the TV is low, this implies that the delays are following *expected* patterns consistent with historical correlations. On the other hand, a high TV implies that various pairwise airports that are historically highly correlated are experiencing delays inconsistent with past correlations.

We formalize our first definition that characterizes the spatial distribution of airport delays based on the TV metric. Analogous to the previous definition for OIS, we classify a graph signal as an outlier if its TV lies outside an empirically determined region based on $\mathcal{O}_M$:

Definition 5. A data point $\mathbf{x}$ is classified as a *weak outlier in distribution (weak OID)* if

$$\mathrm{TV}(\mathbf{x}) \notin \left[\underline{\Gamma},\, \overline{\Gamma}\right]$$

where the the lower bound $\underline{\Gamma} \coloneqq Q_1 - 1.5e^{-4\times MC} IQR$, the upper bound $\overline{\Gamma} \coloneqq Q_3 + 1.5e^{3\times MC} IQR$, and the IQR and $MC \geq 0$ are defined on the set $\boldsymbol{\mathcal{V}}_{\mathrm{TV}(\mathbf{x})} = \left\{\mathcal{V}_{\mathrm{TV}(\mathbf{x}),1} = \mathbf{x}^{(1)\intercal}\mathcal{L}\mathbf{x}^{(1)}, \ldots, \mathcal{V}_{\mathrm{TV}(\mathbf{x}),M} = \mathbf{x}^{(M)\intercal}\mathcal{L}\mathbf{x}^{(M)}\right\}$.

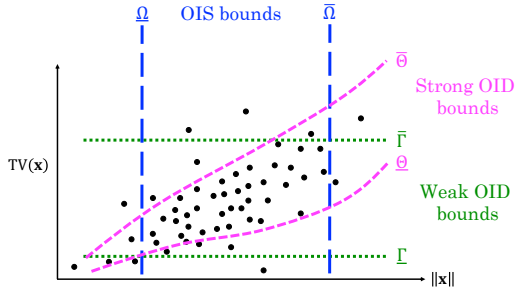


Fig. 1. Depiction of the various outlier detection bounds.

A concern with this characterization of outliers in distribution is that it does not account for the inherent quadratic relationship between the TV and the TD (see Definition 3). This means that as the TD increases, the TV also increases, and the notion of what is expected or unexpected in terms of TV changes based on the TD. It for this reason that we refer to the previous definition of outliers in distribution as *weak*, and we propose a more practically relevant, tighter definition for *strong outliers in distribution*:

Definition 6. A data point $\mathbf{x}$ is classified as a *strong outlier in distribution (strong OID)* if

$$\mathrm{TV}(\mathbf{x}) \notin \left[\underline{\Theta},\, \overline{\Theta}\right]$$

where the lower bound $\underline{\Theta} \coloneqq Q_1 - 1.5e^{-4\times MC} IQR$, the upper bound $\overline{\Theta} \coloneqq Q_3 + 1.5e^{3\times MC} IQR$, and the IQR and $MC \geq 0$ are defined on the (sub)set $\boldsymbol{\mathcal{V}}_{\mathrm{TV}(\mathbf{x})|\|\mathbf{x}\|} \subseteq \boldsymbol{\mathcal{V}}_{\mathrm{TV}(\mathbf{x})}$.

By conditioning on the magnitude of the TD, the outlier bounds for strong OID eliminate the effect of scale, and identifies outliers solely based on the relative spatial distribution of delays on the graph. Figure 1 summarizes the three different outlier bounds from Definitions 4, 5, and 6.

Algorithm 1 Computing strong OID bounds

Input: Minimum bin size n; TD set $\boldsymbol{\mathcal{V}}_{\|\mathbf{x}\|}$; TV set $\boldsymbol{\mathcal{V}}_{\mathrm{TV}(\mathbf{x})}$
Output: Outlier bound bins $\widetilde{\mathbf{LB}}_{\|\mathbf{x}\|}$; Upper outlier bound $\overline{\Theta}$; Lower outlier bound $\underline{\Theta}$

$\widetilde{\boldsymbol{\mathcal{V}}}_{\|\mathbf{x}\|} \leftarrow$ Sort $\boldsymbol{\mathcal{V}}_{\|\mathbf{x}\|}$ s.t. $\mathcal{V}_{\|\mathbf{x}\|,i} \leq \mathcal{V}_{\|\mathbf{x}\|,j}, \forall i < j$ and $i, j \in \{1, ..., M\} \times \{1, ..., M\}$
$\mathbf{LB}_{\|\mathbf{x}\|} \leftarrow \left\{ \widetilde{\mathcal{V}}_{\|\mathbf{x}\|,i} \,\middle|\, i \in \left\{1, ..., \min\left\{M, 1 + \left\lfloor \frac{M}{n} \right\rfloor + n\right\}\right\}\right\}$
$\Delta_{\max} \leftarrow \max_i \left\{ \widetilde{\mathcal{V}}_{\|\mathbf{x}\|,i+1} - \widetilde{\mathcal{V}}_{\|\mathbf{x}\|,i} \right\}$
for *Lower bound index* $i_{\mathbf{LB}} = 1 : \left|\mathbf{LB}_{\|\mathbf{x}\|}\right|$ **do**
  if $i_{\mathbf{LB}} < \left|\mathbf{LB}_{\|\mathbf{x}\|}\right|$ **then**
    $\ell_b \leftarrow \mathbf{LB}_{\|\mathbf{x}\|,i_{\mathbf{LB}}}$; $\ell_u \leftarrow \mathbf{LB}_{\|\mathbf{x}\|,i_{\mathbf{LB}}+1}$
    $\boldsymbol{\mathcal{V}}^{i_{\mathbf{LB}}}_{\mathrm{TV}(\mathbf{x})} \leftarrow \left\{ \mathcal{V}_{\mathrm{TV}(\mathbf{x}),i} \,\middle|\, \ell_b \leq \mathcal{V}_{\mathrm{TV}(\mathbf{x}),i} < \ell_u \right\}$
    $\widetilde{\mathbf{LB}}_{\|\mathbf{x}\|,i_{\mathbf{LB}}} \leftarrow \frac{1}{2}\left(\mathbf{LB}_{\|\mathbf{x}\|,i_{\mathbf{LB}}} + \mathbf{LB}_{\|\mathbf{x}\|,i_{\mathbf{LB}}+1}\right)$
  else
    $\ell_b \leftarrow \mathbf{LB}_{\|\mathbf{x}\|,i_{\mathbf{LB}}}$
    $\boldsymbol{\mathcal{V}}^{i_{\mathbf{LB}}}_{\mathrm{TV}(\mathbf{x})} \leftarrow \left\{ \mathcal{V}_{\mathrm{TV}(\mathbf{x}),i} \,\middle|\, \ell_b \leq \mathcal{V}_{\mathrm{TV}(\mathbf{x}),i} \right\}$
    $\widetilde{\mathbf{LB}}_{\|\mathbf{x}\|,i_{\mathbf{LB}}} \leftarrow \frac{1}{2}\left(\mathbf{LB}_{\|\mathbf{x}\|,i_{\mathbf{LB}}} + \max\left\{\boldsymbol{\mathcal{V}}_{\|\mathbf{x}\|}\right\}\right)$
  end
  $\left\{Q_1^{i_{\mathbf{LB}}}, Q_3^{i_{\mathbf{LB}}}\right\} \leftarrow \left\{Q_1 \text{ of } \boldsymbol{\mathcal{V}}^{i_{\mathbf{LB}}}_{\mathrm{TV}(\mathbf{x})}, Q_3 \text{ of } \boldsymbol{\mathcal{V}}^{i_{\mathbf{LB}}}_{\mathrm{TV}(\mathbf{x})}\right\}$
  $IQR^{i_{\mathbf{LB}}} \leftarrow Q_3^{i_{\mathbf{LB}}} - Q_1^{i_{\mathbf{LB}}}$
  $MC^{i_{\mathbf{LB}}} \leftarrow$ compute medcouple for $\boldsymbol{\mathcal{V}}^{i_{\mathbf{LB}}}_{\mathrm{TV}(\mathbf{x})}$
  if $MC^{i_{\mathbf{LB}}} \geq 0$ **then**
    $\overline{\theta}^{i_{\mathbf{LB}}} \leftarrow Q_3^{i_{\mathbf{LB}}} + 1.5\left(IQR^{i_{\mathbf{LB}}}\right)\exp\left(3 \times MC^{i_{\mathbf{LB}}}\right)$
    $\underline{\theta}^{i_{\mathbf{LB}}} \leftarrow Q_1^{i_{\mathbf{LB}}} - 1.5\left(IQR^{i_{\mathbf{LB}}}\right)\exp\left(-4 \times MC^{i_{\mathbf{LB}}}\right)$
  else
    $\overline{\theta}^{i_{\mathbf{LB}}} \leftarrow Q_3^{i_{\mathbf{LB}}} + 1.5\left(IQR^{i_{\mathbf{LB}}}\right)\exp\left(4 \times MC^{i_{\mathbf{LB}}}\right)$
    $\underline{\theta}^{i_{\mathbf{LB}}} \leftarrow Q_1^{i_{\mathbf{LB}}} - 1.5\left(IQR^{i_{\mathbf{LB}}}\right)\exp\left(-3 \times MC^{i_{\mathbf{LB}}}\right)$
  end
end
$\widetilde{\mathbf{LB}}_{\|\mathbf{x}\|} \leftarrow \left\{ \widetilde{\mathbf{LB}}_{\|\mathbf{x}\|,1}, ..., \widetilde{\mathbf{LB}}_{\|\mathbf{x}\|,\left|\mathbf{LB}_{\|\mathbf{x}\|}\right|} \right\}$
$\widehat{\overline{\Theta}} \leftarrow \left\{ \overline{\theta}^1, ..., \overline{\theta}^{\left|\mathbf{LB}_{\|\mathbf{x}\|}\right|} \right\}$; $\widehat{\underline{\Theta}} \leftarrow \left\{ \underline{\theta}^1, ..., \underline{\theta}^{\left|\mathbf{LB}_{\|\mathbf{x}\|}\right|} \right\}$

In practice, due to a finite number of data points in $\mathcal{O}_M$, it is not possible to condition exactly on $\|\mathbf{x}\|$. Hence, we compute a relaxed, discretized bound instead. Algorithm 1 describes the steps required to compute the discretized bounds $\left\{\widehat{\overline{\Theta}}, \widehat{\underline{\Theta}}\right\}$ for strong outliers in distribution. The idea behind Algorithm 1 is to condition the TV on a discrete interval around $\|\mathbf{x}\|$, rather than a particular value for the 1-norm. To obtain robust estimates, we dynamically vary the interval widths to ensure an equal number of data points n in each interval, with the possible exception of the last bin. The TV of each data observation within the (sub)set that falls into a particular interval is used to compute the IQR and MC, then the outlier bounds for that particular interval. Finally, the output from Algorithm 1 is linearly interpolated for points between the discrete bins, and extrapolated at the edges till the boundary data points. This process is illustrated in Figure 2. The discretized estimates of the strong OID bounds converge to the actual bounds as the number of data points increases and the maximum bin width $\Delta_{\max}$ (see Algorithm 1) decreases, i.e.

$$\lim_{\substack{|\mathcal{O}_M| \to \infty \\ \Delta_{\max} \to 0}} \left\{\widehat{\overline{\Theta}}, \widehat{\underline{\Theta}}\right\} = \left\{\overline{\Theta}, \underline{\Theta}\right\}.$$

We emphasize that our approach does not assume any underlying distributions for the graph signals, the TV, or the conditional distribution of the TV. In addition, we also explicitly consider the skew of these empirical distributions, and the practical challenges associated with estimating robust bounds from a possibly small, finite data set $\mathcal{O}_M$. In the next two sections, we present our findings obtained by using GSP spectral decomposition techniques, as well as our outlier detection methods, to the US and China airport delay networks.

III. Spectral analysis of the networks

In this section, we analyze the US and China airport delay correlation networks, the eigendecomposition of their graph Laplacians, and the average spectral energy distributions. We will highlight how the eigenvectors of the graph Laplacian complement common operational knowledge about airport delay patterns in the US and China. Furthermore, the associated eigenvalues and spectral energies enable a comparison of the spatial variance of delays observed in the Chinese airport network versus the US.

A. US and China correlation networks

We plot in Figure 3 the correlation coefficients for our network of 30 Chinese and US airports, computed from the 5-year data described in Section II-B. The resultant correlation network for China shows a much larger subset of airport exhibiting high pairwise correlations in terms of their total delay time series. This is in contrast with the US, where there are two distinct airport subsets with correlation coefficients higher than the rest of the network – East Coast airports, and to a lesser extent, West Coast airports. Typically, these high correlations are due to geographic proximity which leads to common weather impacts, combined with operational factors such as traffic flow and airline hub characteristics. For example, airport pairs such as ORD-MDW (US) and SZX-CAN (China) do not have any traffic flows, but are collocated in the same metroplex area, thus resulting in high correlations. On the other hand, the delay correlations at airport pairs such as SHA-SZX and BOS-LGA are influenced more by the high volume of shared traffic flows rather than geographic factors.

These correlation networks indicate that, compared to the US, the Chinese network has a larger and more geographically diverse set of airports whose delays are closely coupled. We

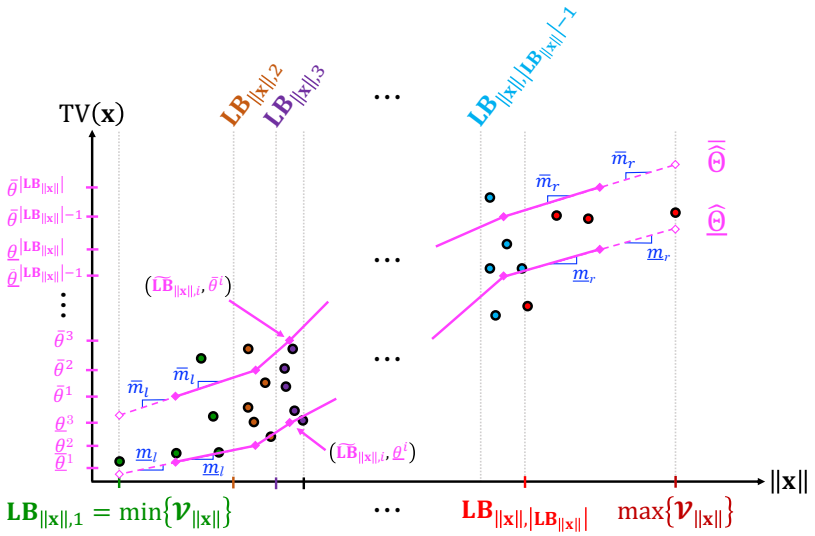


Fig. 2. Illustration of the end-point interpolation scheme used to extend the strong OID bounds $\left\{\widetilde{\mathbf{LB}}_{\|\mathbf{x}\|}, \left\{\overline{\widehat{\Theta}}, \underline{\widehat{\Theta}}\right\}\right\}$ retrieved from Algorithm 1.

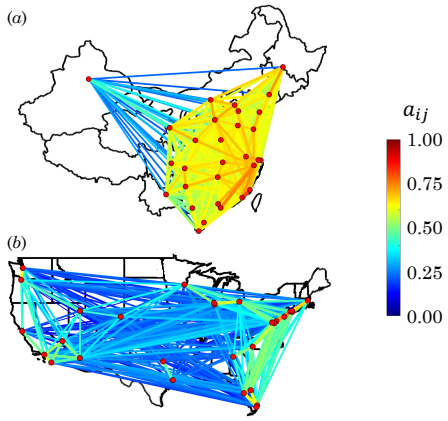


Fig. 3. Airport delay correlations shown with geographical context for the (a) Chinese and (b) US airspace. Higher correlations are also emphasized with wider lines. Note that HNL is not shown in (b) for simplicity.

explore this further in the next subsection by interpreting the eigendecomposition of the graph Laplacians corresponding to these correlation networks. The resultant eigenvector modes help us in identifying *specific groups of airports* which contribute to unexpected delay distributions.

B. US and China eigenvector modes

Recall that the eigenvector modes allows for the decomposition of any airport delay signal vector into linear combinations of these modes, where higher-indexed modes (corresponding to larger eigenvalues) are more "energetic" and result in higher TV. We order the set of 30 eigenvalues for China and the US airport network in ascending order according to their magnitude; thus, Figure 4 provides a geographic depiction of the top 5 most energetic eigenvector modes for both networks.

Eigenvector modes have interesting operational interpretations based on the sign of each of the 30 components in v_i. In Figure 4 we represent positive components in blue and negative components in red; the important characteristic is the *difference* in signs between two or more airports, not so much the sign of an airport itself, since any scalar multiple of the eigenvector is also an eigenvector. For example, in mode v_{27} for China, the signs on NKG and TYN are the opposite of TAO and FOC. This mode portrays a scenario where airport delays at NKG and TYN are decreasing, and airport delays at TAO and FOC are increasing, *or vice versa*. This point regarding comparing different signs, and not the sign of one particular airport, is crucial to keep in mind for the rest of this paper.

Visually, the most energetic eigenvector modes of China are significantly different from those of the US (see the top and bottom rows in Figure 4) in terms of the geographic extent of the highlighted airports. Most of the airports in China are widely distributed across the eastern part of the country. With the exception of PEK in mode v_{29} and SXZ in v_{30}, none of the airports in these high-energy modes are ranked in the top five airports in terms of traffic volume. Furthermore, there is only one high energy (and thus, high TV) mode in China (v_{28}) where geographically proximate airports (XMN and FOC) are experiencing opposite delay trends, whereas this scenario occurs for all US eigenvector modes.

It is interesting that our data-driven delay analysis identifies the same airports from another study that identified critical airports for system resilience [26]. Note that such a list of airports is different from those arising out of studies on operational dynamics or traditional network measures such as degree and betweenness centrality. In particular, FOC has low traffic volumes (ranked 27$^{\text{th}}$ in 2018), and is not considered to be a central airport in terms of its connectivity. However, it is determined to be critical in a simulation-based study [26] that quantify the resilience of the system. Analogously, our GSP analysis based on empirical data also identified FOC in four of the five high-energy modes as a crucial airport in determining

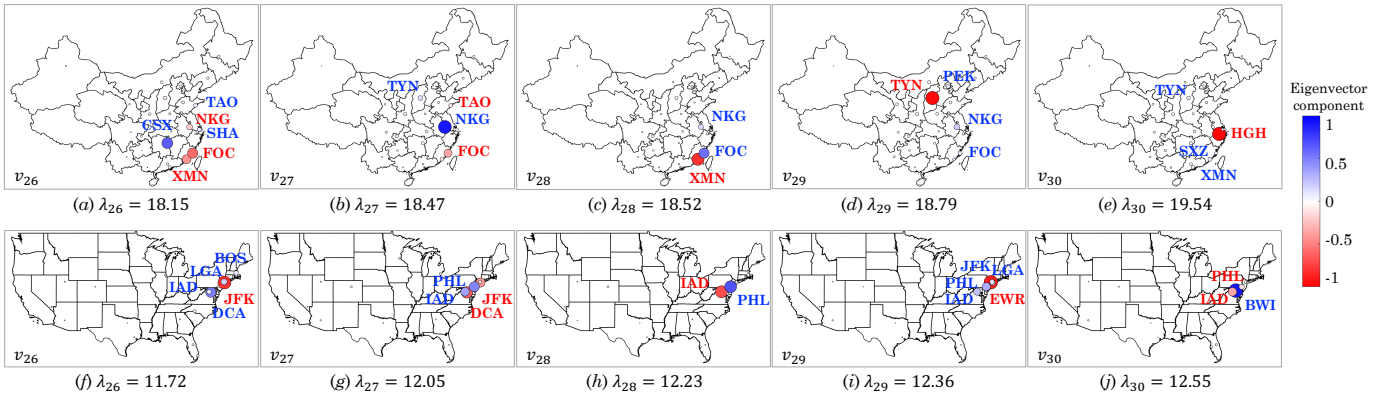


Fig. 4. Top 5 most energetic eigenvector modes (i.e. v_{26} through v_{30}) of the graph Laplacian for China (*a*)-(*e*) and the US (*f*)-(*j*).

whether the system-wide delay distribution is expected or not.

Furthermore, the high-energy US modes involve airports within the same multi-airport system having opposing delay trends; for example, LGA and JFK in New York City have opposing delay trends in v_{26}. This is not the case in China; for example, the two Shanghai airports (PVG and SHA) never have opposing delay trends in the high-energy modes.

For the US, we see that the most energetic eigenvector modes correspond to cliques of East Coast airports with delays trending opposite to each other. These include airports serving the same metropolitan area having opposing delay trends; examples of this include modes v_{27} (IAD versus DCA), v_{29} (JFK and LGA versus EWR), and v_{30} (IAD versus BWI). Even if one of these modes has a high spectral energy on a given day, we expect a higher TV for that day due to the large eigenvalues associated with these modes (Proposition 1). Going one step further, we note that for the US, checking whether the delay distribution of the system is expected is approximately equivalent to checking if the delay distributions in the East Coast are expected. Similarly, it is likely that the system is an outlier in distribution if the East Coast airports experience an unexpected distribution of delays.

While it may be sufficient to primarily monitor a geographically localized subset of airports in the US to analyze the spatial delay distributions, the same cannot be said for the China airport network. We see a lack of geographic consistency in the airports highlighted by the five most energetic eigenvector modes for China. In particular, for China we see airports as far north as TYN and PEK appearing in some of the modes, along with airports in the southeast such as XMN and FOC. Although this is expected given the correlation networks for the two countries (Figure 3), the eigenvector modes provide specific cliques of airport that cannot be identified through a simple ranking of correlation coefficients, since these are inherently limited to pairwise interactions.

C. US and China spectral energies

The eigenvalue coresponding to an eigenvector mode is a measure of the "frequency" of this mode, whereas its spectral energy – specifically, the percent contribution of its spectral energy to the spectrum for an average day, averaged over the entire 10-year data set – is a measure of its *contribution*, or *impact* within the network. We plot these two quantities in Figure 5 for both the US as well as the Chinese network. We observe a clear distinction in the magnitude of the eigenvalues between the two countries. The average magnitude of US eigenvalues is 9.17, and they are lower than all but one eigenvalue from the China airport network. The average magnitude for the Chinese eigenvalues is 16.93, indicating that the average TV of delay signals in China is significantly higher than the US.

The differences in terms of the spectral energy distribution is less pronounced between China and the US. More than 80% of the average spectral energy is contained in the constant mode for both countries (80.92% for the US and 87.87% for China) due to the fact that typical days in both networks do not experience significant disruptions and/or unexpected spatial distributions of delay. Hence, we remove this constant mode for a more nuanced comparison between the more operationally interesting modes v_2 through v_{30} in Figure 5. We also note that in the China airport network, there appears to be a couple of modes that dominate the average eigendecomposition, whereas the distribution is more even in the US.

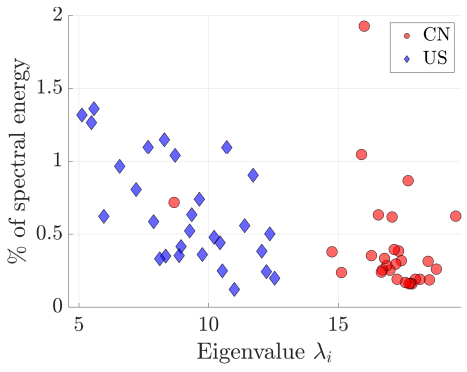


Fig. 5. Percentage of spectral energy per eigenvector mode versus associated eigenvalue for the airport delay graph of US and China (CN).

IV. Outlier analysis

A. Identifying outliers using total variation and delay

For each day with an associated airport delay graph signal $\mathbf{x}$, we can compute the TD and TV, and visualize it as a point $(\|\mathbf{x}\|, \mathrm{TV}(\mathbf{x}))$ on a *TV versus TD plot*. Using the skew-adjusted IQR method for detecting outliers detailed in Section II, we evaluate the outlier in scale (OIS), weak outlier in distribution (OID), and strong OID bounds for the China and US airport networks. We present the TV-TD plots for China and the US in Figure 6, with OIS, weak OID, and strong OID bounds demarcated.

Note that the observed delay within the China airport network is higher than the US, resulting in a significantly higher TV for China due to the quadratic-dominated relationship between TD and TV. However, since the outlier bounds are trained with respect to each country's data, the outlier statistics can be compared across the two countries. We summarize the outlier statistics in Table I. We observe that in terms of the TD, which is the simplest measure of the severity of a disruption, the number of OIS in the US is higher than in China. This indicates that even though the TD may be higher on average for China compared to the US, there are more days in the US where the TD was *unexpectedly* high or low. The statistics for OID highlight the importance of using the strong OID in lieu of the weak OID definition. While the US has more weak OID than China, the reverse is true when the strong OID bounds are used, which takes into account the fact that TV grows quadratically with TD. The weak OID is thus not a very reliable metric to identify spatial distribution outliers, and its conclusions may even be contrary to ones obtained from the strong OID definition. We conclude that the Chinese airport network not only incurs more delays than the US, but the delays also tend to be spatially distributed at unexpected sets of airports. Given the lack of geographic consistency in the higher-energy eigenvalue modes for China, this indicates that the unexpected spatial delay distributions are also likely to be geographically dispersed throughout the country.

Outlier Type	China	US
OIS	19	**34**
Weak OID	16	**30**
OIS & Weak OID	5	**8**
Strong OID	**103**	73

TABLE I
Number of outlier days out of 2,192 days (2012-2017) for China and the US, categorized by outlier type.

B. Monthly distribution of outliers

We plot the total number of strong OID days in each month for China and the US in Figure 7 to analyze seasonal and temporal patterns. While one might presume that there are more unexpected spatial delay distributions in the summer due to disruptions such as thunderstorms, we found that May, June, and July actually contain the least number of strong OID days for both countries. The temporal distribution of outliers from

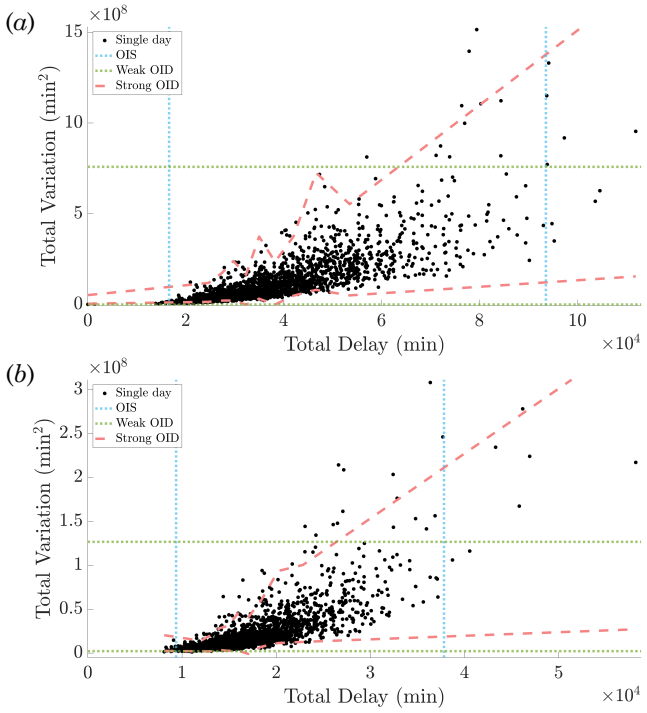


Fig. 6. TV versus TD for all days in 2012-2017 for China (*a*) and the US (*b*), with outlier in scale (OIS) and distribution (OID) bounds marked.

Figure 7 suggests that unexpected spatial delay distributions are much more common in the winter, at least from a system-wide perspective for both countries.

Nor'easter snowstorms and significant cancellations during the months of December through March contribute to a higher occurrence of strong OID days in the winter for the US. Similarly, for China, meteorological factors such as fog and snowstorms lead to significant unexpected delay distributions in the winter months. Additionally, trends in consumer preference also appear to play a major role. The large number of outliers in October are particularly clustered around the first week, which is a week-long national holiday in China (Golden Week). A surge in aviation demand during this travel season may be a contributing factor for the occurrence of a large number of strong OID days. Finally, we would like to point out that seasonal effects are more pronounced in China than the US, indicating more volatile, unpredictable, and weather-sensitive operations in China. This may be a consequence of the high-growth phase of the Chinese aviation market. Our analysis would help policymakers and system managers to identify these specific instances of unexpected patterns, and direct their efforts towards reducing their occurrence.

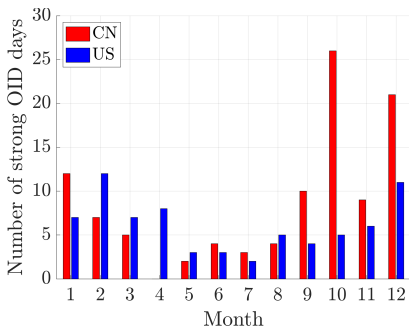


Fig. 7. TV versus TD for all days in 2012-2017 for China (*a*) and the US (*b*), with outlier in scale (OIS) and distribution (OID) bounds marked.

V. Interpretability of Outliers

A. Framework for interpreting the outliers

We emphasized that one advantage of our methods is the ability to provide interpretations at multiple stages of the outlier classification process. In this section, we present a framework for outlier interpretability, and then illustrate its applicability by providing interpretations for why a particular set of days were classified as strong OID.

Our interpretation framework can be thought of as a workflow that we visualize in Figure 8. Note the dual layer of interpretations available through this framework: The first layer of interpretability distinguishes between different *types* of outliers, i.e. strong OID versus OIS. We do not interpret weak OID days, as we previously demonstrated the tighter performance of strong OID bounds. The second layer of interpretability allows for *operational* insights to be derived from GFT and spectral modes. In particular, if the input day $\mathbf{x} \in \mathbb{R}_{\geq 0}^{N \times 1}$ is determined to be a strong OID via the appropriate bounds, then we can compute the spectrum of the signal $\{\alpha_1^2, \ldots, \alpha_N^2\}$. Using the spectral energy percentage contained in each eigenvector mode, we can retrieve the dominant eigenvector modes by selecting those with significant energy contributions. We can then identify subsets of airports implicated in a particular strong OID day, and use those airports as the basis for an in-depth operational analysis.

B. Outlier interpretation examples

We apply our framework to interpret some strong OID days within our data set and illustrate the operational insights that can be gained from this process. Specifically, we identify disruptive events that occurred during a particular day by focusing our attention on prominent airports within activated eigenvector modes and cross-referencing with factors ranging from meteorological events, airport outage, consumer behavior, and airline operational practices.

We select 6 strong OID days from the China airport network and present them in Table II along with their TD, TV, and the fraction of spectral energy $\alpha_i^2 / \sum_i \alpha_i^2$ explained by the dominant modes. We plot the relevant modes in Figure 9.

- *6/10/2017 and 9/25/2017:* The dominant eigenvector mode v_8 shown in Figure 9(*c*) indicates that these two days were classified as outliers because the two Shanghai airports SHA and PVG had significantly higher delays than airports in the north, specifically HET, SHE, and CGO. This is operationally interpretable as heavy rain and thunderstorms affected operations on both days in the Shanghai area. In particular, September 25 involved more than 100 cancellations at PVG, and June 10 resulted in a 50% capacity reduction at SHA and PVG.
- *12/29/2012:* The dominant mode v_{23} indicates that this day was classified as an outlier because of two geographically proximate airports (TAO and TNA) having substantially different delay magnitudes. Heavy snowfall at TAO (resulting in 145 cancelled flights) with relatively no noticeable impact at TNA provides an operational interpretation for this unexpected delay distribution.
- *1/18/2012:* The eigenvector mode v_{24} highlights unexpected spatial delay distributions where delays at TAO and CKG move in opposite directions to XIY. Severe ice accumulation and fog at XIY forced a major airline, China Eastern, to cancel 186 flights. Interestingly, the delays were contained at XIY and did not spread to the other China Eastern focus city of TAO. This is operationally unexpected, as delays typically propagate within an airline's sub-network during disruptions at a major hub.
- *6/21/2012:* Eigenvector mode v_6 indicates that airports in the north – HET, HRB, SHE, DLC – had delay magnitudes opposite to airports in the south (KMG, CAN, SZX). This resulted in an unexpected delay distribution attributable to geographically-localized disruptions. Specifically, adverse weather south of the Yangtze river resulted in flood emergencies, disrupting airport and airspace operations for all the southern airports. The northern airports remained largely unaffected, and experienced low delays. Again, this is unexpected given the typical tendency of delays to propagate and spread.
- *1/28/2012:* The activation of eigenvector mode v_3 indicates that the delays at HAK were significantly higher (or lower) than other airports in China, resulting in the classification of this day as a strong OID. Upon closer inspection, we identify two operational factors that may have contributed to such a delay pattern. First, January 28, 2012 was the last day of the Spring Festival holidays in China, resulting in higher-than-normal scheduled flights at HAK (61,698 passengers on January 28 versus 36,142

Date	**TD** ($\times 10^4$ min)	**TV** ($\times 10^8$ min^2)	**EV Mode % spec. energy**
9/25/2017	5.70	8.12	8 (12%)
1/28/2012	5.88	6.94	3 (23%)
12/29/2012	7.64	11.00	23 (10%)
6/10/2017	7.79	14.00	8 (15%)
1/18/2012	7.95	15.10	24 (18%)
6/21/2012	8.03	11.10	6 (10%)

TABLE II
THE SIX STRONG OID DAYS IN THE CHINESE NETWORK WITH THE HIGHEST TD. THE DATE IS GIVEN IN MONTH/DATE/YEAR FORMAT.

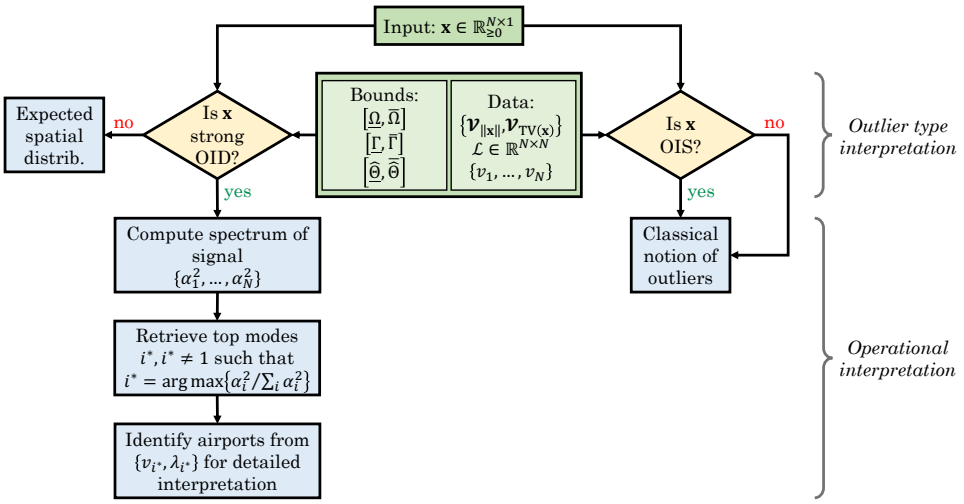
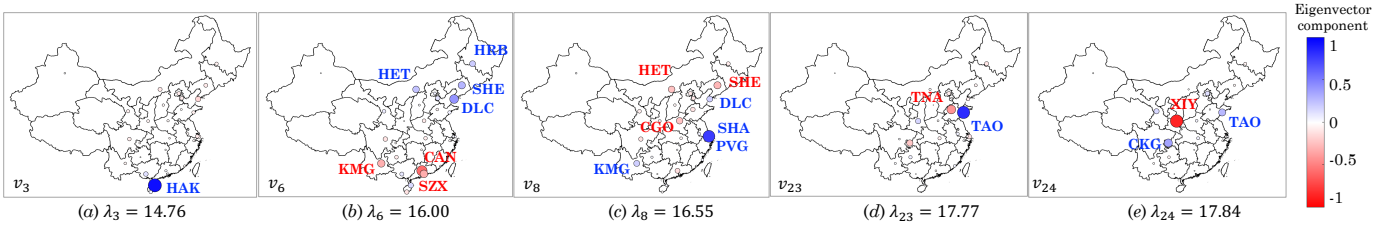


Fig. 8. Flowchart depiction of our framework for the two-step process of interpreting strong OID days in an air transportation setting.



(a) $\lambda_3 = 14.76$ (b) $\lambda_6 = 16.00$ (c) $\lambda_8 = 16.55$ (d) $\lambda_{23} = 17.77$ (e) $\lambda_{24} = 17.84$

Fig. 9. Eigenvector modes observed during the six strong OID days in the Chinese airspace.

on January 22). Second, heavy fog on January 28 exacerbated the already-strained demand-capacity imbalance and led to severe flight delays out of HAK.

Days classified as strong OID in the US can also be interpreted by following the same workflow as detailed in Figure 8. We previously carried out a detailed study where US outliers were analyzed and interpreted [16]. Some of the operationally relevant factors that helped interpret the outlier classification included airport- and airline-specific outages, airline hub locations, traffic flows, geographic proximity, and specific weather disruptions such as nor'easter snowstorms, Atlantic hurricanes, and thunderstorms. As an example of conclusions drawn through this interpretation framework, we found that nor'easters in the US tend to activate extremely high-energy modes, much more so than other disruption types.

VI. Conclusions

We proposed a novel technique to identify outliers in graph signals without any prior assumptions on the underlying distribution of the signals. We demonstrated the applicability of our methods for outlier detection in air traffic delays by comparing the spatial distribution of delays in US and Chinese airport networks. In summary, we (1) identified critical subsets of airports in the US and Chinese airport networks that should be monitored for unexpected spatial delay distributions; (2) observed higher baseline variability in spatial delay distributions in China as compared to the US; (3) compared OIS, weak OID, and strong OID outlier statistics between China and the US as well as examined temporal trends; (4) demonstrated the theoretical as well as operational interpretability of our outlier identification results.

We anticipate several interesting directions for future work. Some are more theoretical, such as deriving analytical outlier bounds under specific distributions for the graph-supported data (e.g. the Gaussian case was addressed in [17]), using a directed graph for the underlying structure, or inferring causal trends from the time-series of nodal signals. Another interesting and applications-oriented research direction is the predictive aspect. For example, can we predict the spatial distribution of airport delays for the rest of the day, or deduce whether a day would be an outlier (OIS, weak OID, or strong OID) based on available partial information, e.g. you have data up to 10 am for this day. Such studies may be augmented by performing a spectral analysis at the temporal resolution of an hour, rather than the daily timescale we used in this work. Finally, our ongoing work related to airline-specific sub-networks has shown that outlier analysis conditioned on individual airlines (e.g. airline-specific graph Laplacians, eigenvector modes, etc.) produces results that differ significantly from network-wide findings and may further improve predictive models.

Acknowledgment

This work was partially supported by NSF under CPS Award No. 1739505, the National Natural Science Foundation

of China (Grant Nos. 61773203, U1833126), and an NSF Graduate Research Fellowship (Max Z. Li).

REFERENCES

[1] "Bureau of Transportation Statistics," https://www.transtats.bts.gov, accessed: 2019-10-27.

[2] "Statistical communique of the civil aviation administration of china on the operations in the civil airports (2018)."

[3] Operation Monitoring Center, the Civil Aviation Administration of China, "Civil flights operation performance reprot," Civil Aviation Administration of China, Tech. Rep., 2018.

[4] P. Ren and L. Li, "Characterizing air traffic networks via large-scale aircraft tracking data: A comparison between china and the us networks," *J AIR TRANSP MANAG*, vol. 67, pp. 181 – 196, 2018.

[5] X. Dong and M. S. Ryerson, "Increasing civil aviation capacity in china requires harmonizing the physical and human components of capacity: A review and investigation," *Transportation Research Interdisciplinary Perspectives*, vol. 1, 2019.

[6] J. P. Stevens, "Outliers and influential data points in regression analysis." *Psychological Bulletin*, vol. 95, no. 2, p. 334, 1984.

[7] T. Kanamori, T. Takenouchi, S. Eguchi, and N. Murata, "The most robust loss function for boosting," in *International Conference on Neural Information Processing*. Springer, 2004, pp. 496–501.

[8] G. Katz, C. Barrett, D. L. Dill, K. Julian, and M. J. Kochenderfer, "Reluplex: An efficient smt solver for verifying deep neural networks," in *International Conference on Computer Aided Verification*. Springer, 2017, pp. 97–117.

[9] P. Filzmoser, *A multivariate outlier detection method*, 2004.

[10] D. M. Rocke and D. L. Woodruff, "Identification of outliers in multivariate data," *Journal of the American Statistical Association*, vol. 91, no. 435, pp. 1047–1061, 1996.

[11] A. S. Hadi, "Identifying multiple outliers in multivariate data," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 54, no. 3, pp. 761–771, 1992.

[12] W. Eberle and L. Holder, "Discovering structural anomalies in graph-based data," in *Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007)*. IEEE, 2007, pp. 393–398.

[13] S. Shekhar, C.-T. Lu, and P. Zhang, "Detecting graph-based spatial outliers," *Intelligent Data Analysis*, vol. 6, no. 5, pp. 451–468, 2002.

[14] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," *IEEE transactions on signal processing*, vol. 61, no. 7, pp. 1644–1656, 2013.

[15] H. B. Ahmed, D. Dare, and A.-O. Boudraa, "Graph signals classification using total variation and graph energy informations," in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2017, pp. 667–671.

[16] M. Z. Li, K. Gopalakrishnan, H. Balakrishnan, and K. Pantoja, "A spectral approach towards analyzing air traffic network disruptions."

[17] K. Gopalakrishnan, M. Z. Li, and H. Balakrishnan, "Identification of outliers in graph signals," in *2019 IEEE Conference on Decision and Control (CDC)*, December 2019.

[18] A. Cook, S. Belkoura, and M. Zanin, "Atm performance measurement in europe, the us and china," *Chinese Journal of Aeronautics*, vol. 30, no. 2, pp. 479 – 490, 2017.

[19] S. Wandelt, X. Sun, and J. Zhang, "Evolution of domestic airport networks: a review and comparative analysis," *Transportmetrica B: Transport Dynamics*, vol. 7, no. 1, pp. 1–17, 2019.

[20] J. Wang, Y. Zhou, and G. Q. Huang, "Alternative pair in the airport network," *Transportation Research Part A: Policy and Practice*, vol. 124, pp. 408 – 418, 2019.

[21] M. C. R. Murça, R. J. Hansman, L. Li, and P. Ren, "Flight trajectory data analytics for characterization of air traffic flows: A comparative analysis of terminal area operations between new york, hong kong and sao paulo," *Transportation Research Part C: Emerging Technologies*, vol. 97, pp. 324 – 347, 2018.

[22] W.-B. Du, X.-L. Zhou, O. Lordan, Z. Wang, C. Zhao, and Y.-B. Zhu, "Analysis of the chinese airline network as multi-layer networks," *Transportation Research Part E: Logistics and Transportation Review*, vol. 89, pp. 108 – 116, 2016.

[23] O. Lordan, J. M. Sallan, P. Simo, and D. Gonzalez-Prieto, "Robustness of the air transport network," *Transportation Research Part E: Logistics and Transportation Review*, vol. 68, pp. 155–163, 2014.

[24] R. Guimera, S. Mossa, A. Turtschi, and L. N. Amaral, "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles," *Proceedings of the National Academy of Sciences*, vol. 102, no. 22, pp. 7794–7799, 2005.

[25] S. Kim and Y. Yoon, "On node criticality of the northeast asian air route network," *J AIR TRANSP MANAG*, vol. 80, p. 101693, 2019.

[26] Y. Wang, J. Zhan, X. Xu, L. Li, P. Chen, and M. Hansen, "Measuring the resilience of an airport network," *Chinese Journal of Aeronautics*, 2019.

[27] W. Du, B. Liang, G. Yan, O. Lordan, and X. Cao, "Identifying vital edges in chinese air route network via memetic algorithm," *Chinese Journal of Aeronautics*, vol. 30, no. 1, pp. 330 – 336, 2017.

[28] J. Zhang, X.-B. Cao, W.-B. Du, and K.-Q. Cai, "Evolution of chinese airport network," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 18, pp. 3922 – 3931, 2010.

[29] M. Zhang, B. Liang, S. Wang, M. Perc, W. Du, and X. Cao, "Analysis of flight conflicts in the chinese air route network," *Chaos, Solitons & Fractals*, vol. 112, pp. 97 – 102, 2018.

[30] M. Zanin and F. Lillo, "Modelling the air transport with complex networks: A short review," *The European Physical Journal Special Topics*, vol. 215, no. 1, pp. 5–21, 2013.

[31] G. Brys, M. Hubert, and A. Struyf, "A robust measure of skewness," *Journal of Computational and Graphical Statistics*, vol. 13, no. 4, pp. 996–1017, 2004.

[32] M. Hubert and E. Vandervieren, "An adjusted boxplot for skewed distributions," *Computational Statistics & Data Analysis*, vol. 52, no. 12, pp. 5186 – 5201, 2008.

APPENDIX

IATA	Airport Name	IATA	Airport Name
CAN	Guangzhou	ATL	Atlanta
CGO	Zhengzhou	BOS	Boston
CKG	Chongqing	BWI	Baltimore
CSX	Changsha	CLT	Charlotte
CTU	Chengdu	DCA	Washington-National
DLC	Dalian	DEN	Denver
FOC	Fuzhou	DFW	Dallas-Fort Worth
HAK	Haikou	DTW	Detroit
HET	Hohhot	EWR	Newark
HGH	Hangzhou	FLL	Fort Lauderdale
HRB	Harbin	HNL	Honolulu
KMG	Kunming	IAD	Washington-Dulles
KWE	Guiyang	IAH	Houston-Intercontinental
NKG	Nanjing	JFK	New York-John F. Kennedy
NNG	Nanning	LAS	Las Vegas
PEK	Beijing	LAX	Los Angeles
PVG	Shanghai-Pudong	LGA	New York-LaGuardia
SHA	Shanghai-Hongqiao	MCO	Orlando
SHE	Shenyang	MDW	Chicago-Midway
SYX	Sanya	MIA	Miami
SZX	Shenzhen	MSP	Minneapolis
TAO	Qingdao	ORD	Chicago-O'Hare
TNA	Jinan	PDX	Portland
TSN	Tianjin	PHL	Philadelphia
TYN	Taiyuan	PHX	Phoenix
URC	Urumqi	SAN	San Diego
WUH	Wuhan	SEA	Seattle
XIY	Xi'an	SFO	San Francisco
XMN	Xiamen	SLC	Salt Lake City
ZGC	Lanzhou	TPA	Tampa

TABLE III
IATA THREE-LETTER CODE AND CORRESPONDING FULL AIRPORT NAME OF THE AIRPORTS WITHIN OUR GRAPH OF CHINA AND THE US.

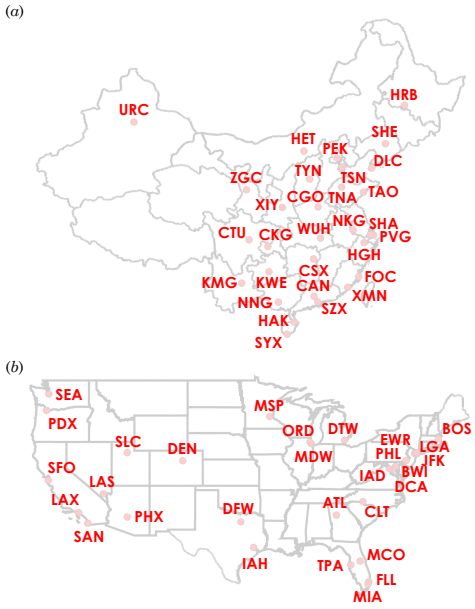


Fig. 10. Geographic locations of the airports (IATA code given) within our graph of China (*a*) and the US (*b*). Note that HNL is not shown in (*b*) for simplicity.