Submitted to *Transportation Science*
manuscript TS-2019-0414.R1 (Accepted, September 2020)

Graph Signal Processing Techniques for Analyzing Aviation Disruptions

Max Z. Li*, Karthik Gopalakrishnan*, Hamsa Balakrishnan
Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, {maxli, karthikg, hamsa}@mit.edu

Kristyn Pantoja
Texas A&M University, College Station, TX, USA, kristynp@stat.tamu.edu

Understanding the characteristics of air traffic delays and disruptions is critical for developing ways to mitigate their significant economic and environmental impacts. Conventional delay performance metrics reflect only the magnitude of incurred flight delays at airports; in this work, we show that it is also important to characterize the *spatial distribution* of delays across a network of airports. We analyze graph-supported signals, leveraging techniques from spectral theory and graph signal processing to compute analytical and simulation-driven bounds for identifying outliers in spatial distribution. We then apply these methods to the case of airport delay networks, and demonstrate the applicability of our methods by analyzing US airport delays from 2008 through 2017. We also perform an airline-specific analysis, deriving insights into the delay dynamics of individual airline sub-networks. Through our analysis, we highlight key differences in delay dynamics between different types of disruptions, ranging from nor'easters and hurricanes to airport outages. We also examine delay interactions between airline sub-networks and the system-wide network, and compile an inventory of outlier days that could guide future aviation operations and research. In doing so, we demonstrate how our approach can provide operational insights in an air transportation setting. Our analysis provides a complementary metric to conventional aviation delay benchmarks, and aids airlines, traffic flow managers, and transportation system planners in quantifying off-nominal system performance.

Key words: airline networks; Graph Signal Processing; graph Fourier analysis; spectral methods; weather disruptions; airport operations; flight delays

1. Introduction

The air transportation system is a complex and highly interconnected infrastructure that is critical to enabling several aspects of modern society. In the US, airlines operate over 28,000 daily flights, transporting over 2.4 million passengers and 58,000 tons of cargo per day (Airlines for America

* Co-first authors

2018). Civil aviation accounts for 5.1% of the US Gross Domestic Product, and generates $1.6 trillion in revenues (Federal Aviation Administration 2017). The operational characteristics of air transportation renders it vulnerable to a variety of disruptions. In order to achieve a resilient and robust system, it is important to understand the impacts of, and recovery from, such disruptions.

The impacts of air traffic disruptions are most often manifested by canceled and delayed flights. Weather is the primary cause of air transportation disruption, accounting for almost 40% of incurred delays in 2017 (US Department of Transportation 2018). Low cloud ceilings, high winds, low visibility, and thunderstorms can significantly reduce the capacity of an airport runway or an airspace sector. Other causes of disruption include aircraft equipment issues, airport airside and landside abnormalities such as security threats and fires, as well as air traffic control related issues such as staffing shortages and strikes. These disruptions vary in their timings, intensities, and durations; furthermore, due to the highly-interconnected nature of the system, local perturbations have widespread impacts. In fact, *secondary* or *reactionary* delays account for approximately 40% of the domestic delay minutes in the US during 2018 (Bureau of Transportation Statistics 2018). Flight delays cost the US economy $32.9 billion annually (Ball et al. 2010), which is approximately $500 per minute of flight delay. Single instances of large-scale weather disruptions also cause severe economic and operational impacts: The January 22-24, 2016 winter storm resulted in the cancellation of over 11,000 flights, and an estimated financial losses of over $100 million. Airlines, air navigation service providers (ANSPs), and infrastructure planners will all benefit from the study of disruptions and the resultant impact patterns, in order to anticipate and mitigate their effects.

1.1. Problem motivation

Most operational performance benchmarks set by ANSPs and airlines measure, either directly or indirectly, the *magnitude* of incurred flight delays. However, there is a second, more subtle, performance metric: the *spatial distribution* of delays through the system. Two disruptive events could result in the same total delay, but affect very different airports spatially, resulting in different operational impacts. The overarching theme of our work is to identify and analyze days in terms of both the magnitude and the spatial distribution of delays.

The delays experienced by two airports may not be independent; they could be correlated due to geographic proximity, traffic volumes, delay propagation, airline schedules, and traffic flow management procedures. For example, nor'easter systems typically affect several East Coast hub airports, resulting in correlated delays. There is consequently the notion of a typical or *expected* spatial distribution of delays, reflected by the statistical correlations between delays at different airports. We illustrate this notion with the example of Boston Logan International (BOS) and New York LaGuardia (LGA) airports. The delays at these airports are positively correlated, due to the high

traffic volume between them, as well as their geographic proximity. When either airport experiences high delays, we expect high delays at the other airport as well. However, high delays at one of the airports but low delays at the other would be an unexpected occurrence. This unusual behavior would not be captured by the magnitude of delays alone: BOS and LGA may both experience average delays of 90 minutes on one day, while on another day, BOS may experience an average delay of 170 minutes and LGA only 10 minutes. The sum of the delays at the two airports will be 180 minutes on both days, even though the latter day experiences an unexpected distribution of delays. The size of the network in terms of the number of nodes (airports) also complicates a pairwise analysis of spatial delay distributions. The spectral and GSP-based method proposed in this paper addresses these challenges.

In this paper, the events that cause delays (e.g. extreme weather, airport outages, etc.) are referred to as *off-nominal* events or *disruptions*. The resultant spatial delay distributions can be either *expected* or *unexpected*. An expected spatial delay distribution is consistent with the historical delay correlations between the airports, whereas an unexpected spatial delay distribution refers to one that is not. A key objective of this work is to identify days that are unusual in terms of the magnitude or the spatial distribution of delays. We will formally define these *outlier days in scale* and *outlier days in distribution* in Section 3.

1.2. Problem description

We address three primary questions through the methods and results presented in this paper:

1. **Defining and characterizing outliers in graph signals:** We develop a mathematical framework to characterize days with unexpected spatial distributions of delays. Using this framework, we propose an outlier detection theory that identifies such days, using a combination of analytical and simulation-driven techniques.

2. **Operational insights from the outlier analysis:** We interpret the results of analyzing spatial distribution of airport delays from an operational perspective. In particular, we identify specific airports or groups of airports that contribute to unexpected delay dynamics on a given day, and observe differences in delay patterns between various types of disruptions.

3. **System-wide versus airline-specific dynamics:** In addition to a system-wide analysis, we consider the airline sub-networks and their influence on the whole system. We examine the question of expected versus unexpected spatial delay distributions from the perspective of individual airlines, compare these sub-networks, and evaluate the impacts of disruptions on different airlines.

1.3. Literature review

Several previous works have focused on modeling the dynamics of flight delays, using approaches ranging from queuing theory (Pyrgiotis, Malone, and Odoni 2013), network models (Gopalakrishnan, Balakrishnan, and Jordan 2016b,c), discrete event simulators (Ahmadbeygi, Cohn, and Lapp

2010), and machine learning (Kim et al. 2016). The insights developed by these prior works help explain the observed correlations between delays at different airports.

Prior efforts have also considered grouping weather phenomena in the US National Airspace System (NAS) (Sarkis and Talluri 2004, Grabbe, Sridhar, and Mukherjee 2013, Mukherjee, Grabbe, and Sridhar 2013), clustering similar airport arrival capacity profiles (Liu, Hansen, and Mukherjee 2008, Buxi and Hansen 2013) and traffic management initiatives such as Ground Delay Programs (GDPs) (Kuhn 2016, Gorripaty et al. 2017, Ren, Kim, and Kuhn 2018), identifying anomalous aircraft trajectories (Seah, Aligawesa, and Hwang 2010, Li et al. 2015, 2016), and more pertinently, clustering delay networks (Rebollo and Balakrishnan 2014). However, these previous works have two key limitations: They examine only particular sub-components of the system such as a specific airport, a small group of airports, or individual flight trajectories; and they only consider the magnitude, and not the spatial distribution, of disruptions or delays. As noted in Section 1.1, it is critical to examine both facets in order to gain a wider operational perspective.

The identification of unexpected spatial delay distributions is related to the broader problem of outlier detection. A common approach to this problem involves clustering, where data points that are far away from clusters are labeled as outliers (Hadi 1992). Prior work has considered k-means (Gan and Ng 2017), hierarchical clustering (Deb and Dey 2017), density-based clustering (DBSCAN (Abid, Kachouri, and Mahfoudhi 2017)), as well as graph similarity measures (Isufi, Mahabir, and Leus 2018) to identify outliers as data observations that do not belong to any cluster. Another approach involves fitting known distributions to the observed data, and using statistical tests to evaluate if the data point falls at the tails of such distributions (Filzmoser 2004, Rocke and Woodruff 1996). While the Gaussian (Filzmoser 2004, Peña and Prieto 2001) or Gaussian mixture model (Lam et al. 2017) assumption is the most common, recent works have even considered black-box deep neural networks (Kieu, Yang, and Jensen 2018) to empirically identify the tails of the distributions. Information-theoretic approaches have also been considered to identify structural outliers (Shekhar, Lu, and Zhang 2002, Eberle and Holder 2007). While these approaches are successful in detecting outliers, they do not identify outliers based on spatial distributions, they cannot decouple the detection of outliers based on magnitude versus spatial distribution, and there may not be an interpretable explanation for why a particular point was classified as an outlier. In particular, the interpretability of our results is critical to providing implementable insights.

Central to our methods is the GSP framework (Sandryhaila and Moura 2013, 2014a,b, Shuman, Ricaud, and Vandergheynst 2016, Shuman et al. 2013), which extends the notion of a Fourier decomposition to a graph setting, and provides a toolkit to analyze signals on graphs. Our focus on graph-supported signals differentiates our contributions from the large class of prior work that focuses on the *structure* of the graph (e.g., Laplacian embeddings from Belkin and Niyogi (2002))

without any notion of signals supported on such graph structures. Other researchers have applied graph spectral theory for classification (Ahmed, Dare, and Boudraa 2017), filtering or smoothing (Shuman, Ricaud, and Vandergheynst 2016), and extending convolutional neural networks for graphs (Bronstein et al. 2017). Spectral methods have been used to study surface traffic congestion (Crovella and Kolaczyk 2003, Mohan et al. 2014) as well as air traffic flows in Air Route Traffic Control Centers (ARTCCs) (Drew and Sheth 2014, 2015). In particular, Drew and Sheth (2014, 2015) note that spectral methods for air traffic flow management often produce results that are not clearly interpretable; our approach of combining GSP with an outlier detection framework overcomes this limitation.

The identification of outliers is critical for tasks such as data processing, ensuring data integrity, data-driven diagnostics, and anomaly detection. While several techniques exist for outlier detection in multivariate and graphical data sets, they typically focus on data points with deviations in magnitude, and not unexpected spatial distributions across the nodes of a graph. Finally, outside of the transportation context, GSP-based signal classification and anomalous signal detection have been studied using signal spectrum characteristics with no formalized outlier detection theory (Egilmez and Ortega 2014, Drayer and Routtenberg 2018, Sun et al. 2019).

1.4. Manuscript outline

The remainder of the manuscript is organized as follows. Section 2 outlines the main contributions of our work. Section 3 presents the main methodological contributions, including our GSP framework and data-driven approaches for identifying outliers. The analytical expressions for bounds that demarcate outliers can be found in the appendix (Sections A, B, and C). In Section 4, we apply our methods to the US system-wide aviation network. We then consider airline-specific sub-networks and compare them to their system-wide counterpart in Section 5. Section 6 presents concluding remarks and directions for future work.

2. Contributions of this paper

The major contributions of this work are as follows:

1. We formalize notions of, and develop tools to identify, outliers in graph signals. We show how the *total variation* metric can help identify graph-supported signals with an unexpected distribution across the nodes.

2. The proposed methods leverage GSP to provide interpretable explanations for why certain data points are classified as outliers in spatial delay distributions. We identify, via *eigenvector modes*, specific groups of airports whose delays on a particular day are unexpectedly distributed, and correlate them with known operational disruptions.

3. We identify, analyze, and interpret spatial delay patterns across the NAS, with a focus on specific types of disruptions such as nor'easters, hurricanes, airport outages, and thunderstorms. We characterize the differences in the impact of various types of disruptions.

4. We examine the spatial delay dynamics of airline sub-networks with different routing strategies (e.g., hub-and-spoke versus point-to-point), observe their interactions, and compare the impacts of disruptions on them and the system-wide network.

Our methodology identifies outlier NAS days containing disruptions with unexpected delay dynamics that could neither have been comprehensively detected using prior methods, nor operationally interpreted. With this new inventory of outlier NAS days, ANSPs will be able to diversify airspace scenario playbooks to include these rare but operationally important events. This in turn provides a more robust set of playbooks, reducing the need for tactical re-planning. Similar benefits specific to airline sub-networks can be derived from our airline-specific analysis.

The methods proposed in this paper can be applied to a number of networked systems, including other transportation systems, the Internet-of-Things, power systems, communication networks, and biological systems. A preliminary version of this work without a formalized outlier detection framework or airline-specific analysis was published in Li et al. (2019).

3. Methodological framework

In order to present our methodological framework, we first set up standard notations and terminologies from spectral theory and GSP in Sections 3.1 and 3.2. We develop the definitions and propositions directly related to our outlier detection theory in Section 3.3, with additional theoretical analyses in the appendix (Sections A, B, and C). We provide an empirical, data-driven approach to identifying outliers in Section 3.4 based off of the formal outlier theory.

In the context of transportation systems, our methodology identifies geo-spatial outlier events (e.g., geographical disruptions, irregular operations, sudden demand-capacity imbalances) by analyzing large-scale system performance metrics. Specifically, we apply our outlier detection theory to the air transportation network, using the performance metric of airport delays as the graph signals. The definitions, propositions, and data-driven outlier identification techniques formalized in this section form the basis for the analysis of the system-wide and airline-specific delay networks in Sections 4 and 5, respectively.

3.1. Setup and notations

Let us consider a multidimensional real signal $\mathbf{x} \in \mathbb{R}^{N\times 1}$. The N elements of the signal are not independent if they are observed at interconnected elements of a network. We can model the networked system as a graph $G = (V, E)$, where V is the set of $|V| = N$ nodes and $E \subseteq V \times V$ is the set of edges. We consider $\mathbf{x}$ to be supported on the nodes of G; i.e., there is a mapping $f : V \to \mathbb{R}$

from $i \in V$ to the ith element, x_i, of $\mathbf{x}$. There is a weight map $w : E \to \mathbb{R}$ that assigns a weight w_{ij} to edge $(i,j) \in E$. These weights can be represented using an adjacency matrix $A \in \mathbb{S}^{N\times N}$, where $[A]_{ij} = w_{ij}$. We restrict ourselves to undirected graphs, where $w_{ij} = w_{ji}$ and $A = A^{\intercal}$.

Suppose we are given a set of M data points, $\mathcal{O}_M = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}, \ldots, \mathbf{x}^{(M)}\}$, where each data point is $\mathbf{x}^{(k)} = \left(x_1^{(k)}, \ldots, x_N^{(k)}\right)^{\intercal} \in \mathbb{R}^{N\times 1}$. The empirical mean of the graph signal at node i is given by $\hat{\mu}_i = \frac{1}{M}\sum_{k=1}^{M} x_i^{(k)}$, and the sample Pearson correlation coefficient $r_{ij|\mathcal{O}_M}$ on edge (i,j) is

$$r_{ij|\mathcal{O}_M} = \frac{\sum_{k=1}^{M}\left(x_i^{(k)} - \hat{\mu}_i\right)\left(x_j^{(k)} - \hat{\mu}_j\right)}{\sqrt{\sum_{k=1}^{M}\left(x_i^{(k)} - \hat{\mu}_i\right)^2}\sqrt{\sum_{k=1}^{M}\left(x_j^{(k)} - \hat{\mu}_j\right)^2}}. \tag{1}$$

For the remainder of this paper, we set $[A]_{ij} = w_{ij} = r_{ij|\mathcal{O}_M}$. Our graph can be referred to as a *correlation network*, since the edge weights correspond to the correlation between the signals at two nodes. The signals $\mathbf{x}$ are assumed to be realizations of some non-negative random vector $\mathbf{X} = (X_1, \ldots, X_N)^{\intercal} \in \mathbb{R}_{\geq 0}^{N\times 1}$ drawn from some non-negative distribution with well-defined means $\boldsymbol{\mu} = (\mu_1, ..., \mu_N)^{\intercal} \in \mathbb{R}^{N\times 1}$ and covariance $\Sigma \in \mathbb{R}^{N\times N}$, $\Sigma \succeq 0$. The correlation matrix $\mathbf{C} = [\rho_{ij}] \in \mathbb{R}^{N\times N}$ is given by:

$$\rho_{ij} = \frac{\mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]}{\sqrt{\mathbb{E}[(X_i - \mu_i)^2]\mathbb{E}[(X_j - \mu_j)^2]}}. \tag{2}$$

While the value of $\mathbf{C}$ for the underlying distribution of the graph signal vectors may not be known, it can be estimated from $\mathcal{O}_M$ since $r_{ij|\mathcal{O}_M}$ is a consistent estimator of ρ_{ij}, meaning that $\lim_{M\to\infty}\left(r_{ij|\mathcal{O}_M}\right) = \rho_{ij}$. Consequently, A and $\mathbf{C}$ are interchangeable in this paper (see appendix for the partial information case where $A \approx \mathbf{C}$). Finally, we denote norms by $\|\star\| : \mathbb{R}^N \to \mathbb{R}_{\geq 0}$, and we use 1-norms $\|\mathbf{x}\| = \sum_{i=1}^{N}|x_i|$ unless otherwise specified.

3.2. Preliminaries and definitions

We define some preliminary concepts from spectral theory that we will use for outlier detection.

Definition 1 (Graph Laplacian) *The (combinatorial) graph Laplacian $\mathfrak{L}$ with respect to a graph with adjacency matrix A is $\mathfrak{L} = D - A$, where $D = [d_{ij}] \in \mathbb{R}^{N\times N}$ is the diagonal degree matrix of the graph, with $d_{ii} = \sum_{j=1}^{N} w_{ij}$.*

The graph Laplacian $\mathfrak{L}$ is a real symmetric matrix with a full set of orthogonal eigenvectors. The normalized eigenvectors are denoted by $v_i \in \mathbb{R}^{N\times 1}$, $i \in \{1, \ldots, N\}$, with $v_i^{\intercal} v_j = \delta_{ij}$, where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}. \tag{3}$$

All the eigenvalues satisfy $\mathfrak{L}v_i = \lambda_i v_i$. We sort the eigenvalues such that $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N$. Since the graph Laplacian has row sums of 0 (Definition 1), $v_1 = \mathbf{1}$ is the *constant* eigenvector corresponding

to the eigenvalue $\lambda_1 = 0$. Furthermore, the multiplicity of eigenvalues equal to 0 is the number of connected components in the underlying graph. Thus, if the correlation network is fully connected, then $0 = \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_N$, and $\text{span}(\{v_1, ..., v_N\}) \cong \mathbb{R}^{N\times 1}$. Any vector $\mathbf{x} \in \mathbb{R}^{N\times 1}$ can be written as a linear combination of $\{v_1, ..., v_N\}$; i.e., there exist scalars α_i such that $\mathbf{x} = \sum_{i=1}^{N} \alpha_i v_i$.

Definition 2 (Graph Fourier Transform (GFT)) *The Graph Fourier Transform (GFT) of a graph signal vector* $\mathbf{x}$ *is the set of scalars* $\{\alpha_1, ..., \alpha_N\}$ *where* $\alpha_i = v_i^\intercal \mathbf{x}$.

To draw an analogy to the classical Fourier transform, the eigenvectors are equivalent to sinusoids on graphs, and the eigenvalues correspond to discrete frequencies. The scalar $\alpha_i \in \{\alpha_1, ..., \alpha_N\}$ represents the magnitude of contribution of the ith eigenvector of "frequency" λ_i. Similar to the notion of spectral energies for the classical Fourier transform, larger eigenvalues are associated with eigenvectors having higher graph spectral energies, as follows.

Definition 3 (Spectral and total energy) *The spectral energy of* $\mathbf{x}$ *corresponding the* i^{th} *eigenvector is* α_i^2*, and the total energy of* $\mathbf{x}$ *is given by* $\|\mathbf{x}\|_2^2 = \sum_{i=1}^{N} \alpha_i^2$.

The graph Laplacian can also be used to compute a measure of the "smoothness" of a graph signal $\mathbf{x}$, called the *total variation.*

Definition 4 (Total variation (TV)) *The total variation (TV) of a graph signal* $\mathbf{x}$ *with respect to the graph Laplacian* $\mathfrak{L}$ *is defined as:*

$$TV(\mathfrak{L}, \mathbf{x}) = \frac{1}{2} \sum_{i \neq j} w_{ij} (x_i - x_j)^2 = \mathbf{x}^\intercal \mathfrak{L} \mathbf{x}. \tag{4}$$

For brevity, we write $TV(\mathbf{x})$*, dropping the reference to the graph Laplacian* $\mathfrak{L}$*. When the TV is computed with respect to the random vector* $\mathbf{X}$*, we denote it as* $TV(\mathbf{X})$.

The total variation provides a metric map $\text{TV} : \mathbb{R}^{N\times N} \times \mathbb{R}^{N\times 1} \to \mathbb{R}$ that measures the smoothness of a graph signal. A low value of TV corresponds to a graph signal that is said to be *smooth.* The following proposition helps to interpret the GFT in terms of the TV and signal smoothness:

Proposition 1 *Suppose we have a data point* $\mathbf{x}^{(k)} \in \mathcal{O}_M$ *and its GFT* $\{\alpha_1^{(k)}, ..., \alpha_N^{(k)}\}$*. Then, the following two statements are equivalent:*

(i) $TV\left(\mathbf{x}^{(k)}\right) = \left(\mathbf{x}^{(k)}\right)^\intercal \mathfrak{L} \mathbf{x}^{(k)}$.

(ii) $TV\left(\mathbf{x}^{(k)}\right) = \sum_{i=1}^{N} \left(\alpha_i^{(k)}\right)^2 \lambda_i$.

Proof of Proposition 1. Starting with the definition for the TV of $\mathbf{x}^{(k)}$, we show that it is equivalent to $\sum_{i=1}^{N}\left(\alpha_i^{(k)}\right)^2\lambda_i$:

$$\mathrm{TV}\left(\mathbf{x}^{(k)}\right) = \left(\mathbf{x}^{(k)}\right)^{\intercal}\mathfrak{L}\mathbf{x}^{(k)} = \sum_{j=1}^{N}\alpha_j^{(k)}v_j^{\intercal}\ \mathfrak{L}\ \sum_{i=1}^{N}\alpha_i^{(k)}v_i = \sum_{i,j}\alpha_i^{(k)}\alpha_j^{(k)}\lambda_i v_j^{\intercal}v_i = \sum_{i=1}^{N}\left(\alpha_i^{(k)}\right)^2\lambda_i \quad (5)$$

The last equivalence comes from the fact that v_i and v_j are orthogonal eigenvectors, i.e. $\langle v_i, v_j\rangle = \delta_{ij}$. □

Proposition 1 formalizes the relationship between the GFT, the TV of a graph signal, and the spectral and total energies of a graph signal. Larger contributions of the eigenvector v_i to the GFT of $\mathbf{x}$ (i.e., larger values of α_i) result in a higher TV, translating to a less smooth graph signal. Similarly, the more energetic eigenvectors (i.e., larger values of λ_i) contribute to a higher TV, resulting in a less smooth graph signal. Since the eigenvalues are sorted in ascending order with respect to index $i \in \{1, ..., N\}$, we compare eigenvalue magnitudes using the index i.

3.3. Graph signal outliers

Recall that the edge weights of the correlation matrix are given by $r_{ij|\mathcal{O}_M}$. For a pair of nodes $i, j \in V$ connected by an edge with weight $r_{ij|\mathcal{O}_M}$, the contribution to the TV is $r_{ij|\mathcal{O}_M}\left(x_i - x_j\right)^2$. We consider the following possible scenarios:

Case 1: If the graph signals from i and j are highly correlated (i.e., $r_{ij|\mathcal{O}_M} \to 1$), one would then expect that the graph signal magnitudes change in a similar manner. One would expect both x_i and x_j to be large, or both to be small (i.e., $x_i \approx x_j$). However, a new observation may, or may not, conform to the expected behavior.

Case 1a: When the observed data point is as expected (i.e., it is consistent with historical trends), its contribution to the TV is small, since the second term in $r_{ij|\mathcal{O}_M}\left(x_i - x_j\right)^2$ is small.

Case 1b: When the observed data point differs significantly from what is expected, its contribution to the TV is *not* small, since the second term in $r_{ij|\mathcal{O}_M}\left(x_i - x_j\right)^2$ is large.

Case 2: If the graph signals from i and j are uncorrelated (i.e. $r_{ij|\mathcal{O}_M} \to 0$), then based on historical observations, we do not expect specific trends in the graph signal magnitudes. In this case, regardless of any realized signal magnitudes x_i and x_j, the contribution to the TV is small, since the first term of $r_{ij|\mathcal{O}_M}\left(x_i - x_j\right)^2$ is small.

The above reasoning is valid when the signs of the correlation coefficients are all the same. If only a small fraction of correlation coefficients have differing signs, a projective affine transformation can be applied to $r_{ij|\mathcal{O}_M}$, and the intuition still holds (see appendix for details). The analysis of networks with mixed-sign correlation coefficients is a direction for future research. Cases 1 and 2 motivate the use of TV as a metric for outlier detection in terms of a graph signal's spatial distribution.

Since Case 1b is the only case where a high TV may occur, such an occurrence is deemed to be *unexpected given historic observations*. The TV metric yields an aggregate representation of the behavior of $\mathbf{x}$ across the entire graph. We now define the notion of a *weak outlier in distribution*.

Definition 5 *An observation* $\mathbf{x}$ *is considered a* weak distribution outlier of level k *or a* weak outlier in distribution of level k *if*

$$TV(\mathbf{x}) \notin \left[\mathbb{E}\left[TV(\mathbf{X})\right] - k\sqrt{Var\left[TV(\mathbf{X})\right]}, \mathbb{E}\left[TV(\mathbf{X})\right] + k\sqrt{Var\left[TV(\mathbf{X})\right]}\right], \tag{6}$$

for some $k \geq 0$*. In other words, an observation is considered to be a weak outlier in distribution if its TV does not lie within* k *standard deviations of its expected value.*

The quantity $\mathrm{TV}(\mathbf{X})$ is a derived random variable with mean $\mathbb{E}\left[\mathrm{TV}(\mathbf{X})\right]$ and variance $\mathrm{Var}\left[\mathrm{TV}(\mathbf{X})\right]$. Although the definition of a weak outlier in distribution captures variations with respect to historical trends, it does not account for TV scaling quadratically with the graph signal's magnitude. An observation should not be labeled an outlier in distribution if it has a higher TV simply due to having a larger magnitude. We therefore designate it a *weak* outlier in distribution, and propose an alternative metric that captures outliers in magnitude, or scale. This metric corresponds to a classic definition of outliers in multidimensional data: observations with $\|\mathbf{x}\|$ differing significantly from $\mathbb{E}[\|\mathbf{x}\|]$.

Definition 6 *An observation* $\mathbf{x}$ *is considered to be a* scale outlier of level k *or an* outlier in scale of level k *if*

$$\|\mathbf{x}\| \notin \left[\mathbb{E}[\|\mathbf{X}\|] - k\sqrt{Var[\|\mathbf{X}\|]}, \mathbb{E}[\|\mathbf{X}\|] + k\sqrt{Var[\|\mathbf{X}\|]}\right],$$

for some $k \geq 0$*. In other words, an observation is considered to be an outlier in scale of level* k *if its norm does not lie within* k *standard deviations of its expected value.*

The notion of outliers in scale distinguishes the effects of a graph signal's magnitude from its spatial distribution. However, since we know that TV is a quadratic form on $\mathbf{x}$, the definition of weak outliers in distribution does not eliminate the dependence between $\mathrm{TV}(\mathbf{x})$ and its realized magnitude $\|\mathbf{x}\|$. We therefore condition the expectation and variance of $\mathrm{TV}(\mathbf{X})$ with respect to its realized norm $\|\mathbf{X}\| = \|\mathbf{x}\|$ in order to eliminate this dependence, giving rise to the following definition of a *strong* outlier in distribution:

Definition 7 *An observation* $\mathbf{x}$ *is considered to be a* strong distribution outlier of level k *or a* strong outlier in distribution of level k *if* $TV(\mathbf{x}) \notin [\mathcal{A}, \mathcal{B}]$*, where*

$$\mathcal{A} = \mathbb{E}\left[TV(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\|\right] - k\sqrt{Var\left[TV(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\|\right]}$$
$$\mathcal{B} = \mathbb{E}\left[TV(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\|\right] + k\sqrt{Var\left[TV(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\|\right]},$$

for some $k \geq 0$. In other words, an observation is considered to be a strong outlier in distribution if its TV does not lie within k standard deviations of its expected value, conditioned on the realized norm $\|\mathbf{X}\| = \|\mathbf{x}\|$.

The interpretation of our definitions in the context of airport delays is as follows: outliers in scale identify days where the *sum of all airport delays* is higher or lower than expected, whereas outliers in distribution identify days where the *geographical pattern of airport delays* are unexpected. To summarize our various definitions of outliers, Figure 1 presents a graphical description of these definitions. Note that the sum of all airport delays is precisely the 1-norm of $\mathbf{x}$ because all airport delays over a day are non-negative. This characteristic justifies the equality $\|\mathbf{x}\| = \sum_{i=1}^{N} x_i$, and is a commonality in many transportation networks with non-negative signals (e.g. number of bikes at a bike share station). Additionally, since strong outliers in distribution captures solely the geographical spread of airport delays, it will be our metric of choice for the remainder of this paper. While analytical expressions for strong outliers in distribution remain an open problem, our goal in the next subsection is to generate empirical bounds for strong outliers in distribution. For brevity, we relegate our analysis of outliers in scale and weak outliers in distribution to Sections A and B in the appendix.

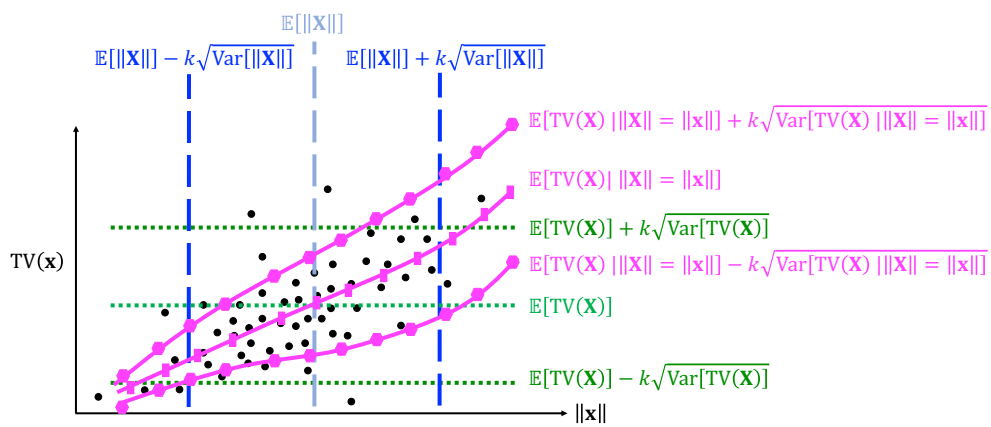

Figure 1 **Notional representation of bounds that we will derive analytically (outliers in scale and weak outliers in distribution) and empirically via simulation (strong outliers in distribution).**

3.4. Empirical bounds for strong outliers in distribution

Recall that while we would like to analytically evaluate the continuously conditional expectation and variance of $\mathrm{TV}(\mathbf{x})$ as utilized in Definition 7, a closed-form expression for the probability density function of the conditional derived random variable $\mathbb{E}\left[\mathrm{TV}(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\|\right]$ and $\mathrm{Var}\left[\mathrm{TV}(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\|\right]$ for all $\|\mathbf{X}\| = \|\mathbf{x}\|$ remains elusive. Thus, we present a modified, *empirical* version of Definition 7 below:

Definition 8 (Empirical strong outliers in distribution) *An observation* $\mathbf{x}$ *is considered to be an* empirical strong outlier in distribution of level k *if* $TV(\mathbf{x}) \notin \left[\widehat{\mathcal{A}_\ell}, \widehat{\mathcal{B}_\ell}\right]$, *where:*

$$\widehat{\mathcal{A}_\ell} = \widehat{\mathbb{E}}\left[TV(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\| \in \mathcal{U}_\ell\right] - k\sqrt{\widehat{Var}\left[TV(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\| \in \mathcal{U}_\ell\right]}$$
$$\widehat{\mathcal{B}_\ell} = \widehat{\mathbb{E}}\left[TV(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\| \in \mathcal{U}_\ell\right] + k\sqrt{\widehat{Var}\left[TV(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\| \in \mathcal{U}_\ell\right]},$$

for some $k \geq 0$ *and empirical bound interval* $\mathcal{U}_\ell$ *computed via Algorithm 1.*

Note that we utilize an interval-based estimation scheme to empirically estimate the mean and variance using $\delta < \infty$ pre-specified disjoint covers $\dot{\bigcup}_{\ell=1}^{\delta} \mathcal{U}_\ell \equiv [\min\{\|\mathbf{x}\|\}, \max\{\|\mathbf{x}\|\}] \subset \mathbb{R}_{\geq 0}$. We propose the following simulation-based method in Algorithm 1 that constructs the bounds from Definition 8. The workflow of Algorithm 1 is depicted in Figure 2.

Algorithm 1 Computing empirical bounds for strong outliers in distribution

Input: Observations $\mathcal{O}_M$; Number of intervals δ; Number of trials T

Output: $\mathcal{U}_\ell$, $\forall \ell \in \{1, \ldots, \delta\}$; $\widehat{\mathbb{E}}\left[\mathrm{TV}(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\| \in \mathcal{U}_\ell\right]$ and $\widehat{\mathrm{Var}}\left[\mathrm{TV}(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\| \in \mathcal{U}_\ell\right]$, $\forall \mathcal{U}_\ell$

Estimate $\hat{\boldsymbol{\mu}}, \hat{\Sigma}, \mathbf{C}$ from $\mathcal{O}_M$

$A \leftarrow \mathbf{C}$; $\mathfrak{L} \leftarrow D - A$

for *Trial i of 1:T* **do** `// Draw T graph signal vector realizations`

  $\mathbf{x} \leftarrow \mathbf{X} \overset{iid}{\sim} \mathcal{N}\left(\hat{\boldsymbol{\mu}}, \hat{\Sigma}\right)$

  $\mathbf{x} \leftarrow \max\{\mathbf{x}, \mathbf{0}\}$ `// Restrict to non-negative realizations`

  $\mathcal{V}_{\|\mathbf{x}\|,i} \leftarrow \sum_{j=1}^{N} x_j$; $\mathcal{V}_{\mathrm{TV}(\mathbf{x}),i} \leftarrow \mathbf{x}^\intercal \mathfrak{L} \mathbf{x}$

end

$\tilde{\Delta} \leftarrow \frac{\max\{\boldsymbol{\mathcal{V}}_{\|\mathbf{x}\|}\} - \min\{\boldsymbol{\mathcal{V}}_{\|\mathbf{x}\|}\}}{\delta}$ `// Width of intervals`

$\mathcal{U}_\ell = \left[\min\left\{\boldsymbol{\mathcal{V}}_{\|\mathbf{x}\|}\right\} + (\ell - 1)\tilde{\Delta}, \min\left\{\boldsymbol{\mathcal{V}}_{\|\mathbf{x}\|}\right\} + \ell\tilde{\Delta}\right]$, $\forall \ell \in \{1, .., \delta\}$

for *Interval* ℓ *of 1:*δ **do** `// Empirical interval-conditioned` $\mathbb{E}$`[TV(`$\mathbf{X}$`)] and Var[TV(`$\mathbf{X}$`)]`

  $\widehat{\mathbb{E}}\left[\mathrm{TV}(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\| \in \mathcal{U}_\ell\right] \leftarrow \mathrm{Mean}\left\{\mathcal{V}_{\mathrm{TV}(\mathbf{x}),i} \mid i \text{ s.t. } \mathcal{V}_{\|\mathbf{x}\|,i} \in \mathcal{U}_\ell\right\}$

  $\widehat{\mathrm{Var}}\left[\mathrm{TV}(\mathbf{X}) \mid \|\mathbf{X}\| = \|\mathbf{x}\| \in \mathcal{U}_\ell\right] \leftarrow \mathrm{Var}\left\{\mathcal{V}_{\mathrm{TV}(\mathbf{x}),i} \mid i \text{ s.t. } \mathcal{V}_{\|\mathbf{x}\|,i} \in \mathcal{U}_\ell\right\}$

end

It is important to note that we only assume that the underlying distribution for $\mathbf{X}$ has a well-defined mean and (co)variance. Thus, Algorithm 1 can be deployed as long as there are sufficient observations in $\mathcal{O}_M$ to estimate $\hat{\boldsymbol{\mu}}, \hat{\Sigma}$, and $\mathbf{C}$ reliably. For the remainder of this paper, the phrase "strong outliers in distribution" refers to the outliers computed via Algorithm 1. We now apply our methodological framework to examine the dynamics of airport delays, and highlight the importance of outlier detection in this context. Even though we focus on the air transportation network, these methods are broadly applicable to the analysis of data from networked systems.

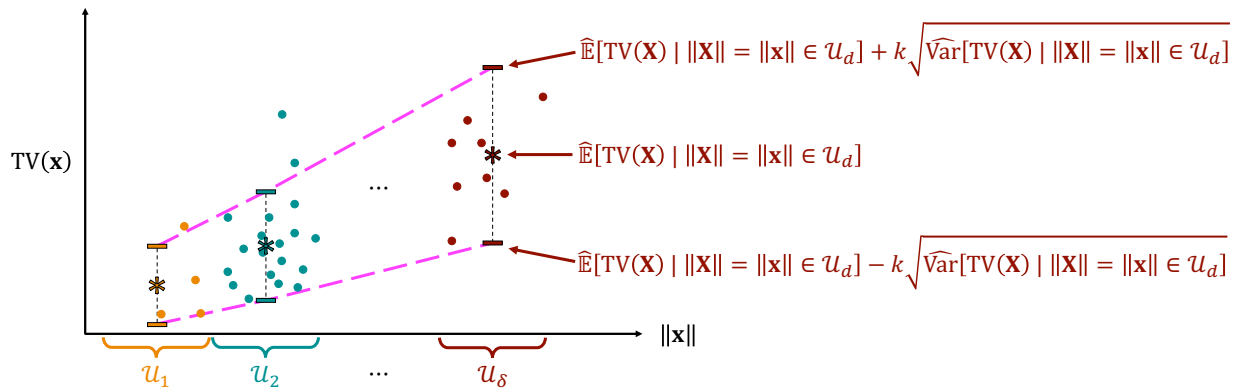

Figure 2 **Empirical strong outliers in distribution bound generating process given simulated observations** $(\|\mathbf{x}\|, \mathbf{TV}(\mathbf{x}))$ **generated from** $\left\{\hat{\boldsymbol{\mu}}, \hat{\Sigma}, \mathbf{C}\right\}_{\mathcal{O}_M}$ **partitioned via** $\dot{\bigcup}_{\ell=1}^{\delta} \mathcal{U}_\ell \equiv [\min\{\|\mathbf{x}\|\}, \max\{\|\mathbf{x}\|\}] \subset \mathbb{R}_{\geq 0}$. **This approximates the magenta bounds shown in Figure 1.**

4. System-wide analysis

In this section, we analyze delays in the aviation network, aggregated over all airlines, to study the system as a whole. We detail the data setup and processing in Section 4.1, then examine various spectral properties in Section 4.2. In Section 4.3, we discuss the projection of airport delays into a 2-dimensional subspace of total delay (TD) and total variation (TV), and discuss our results in Section 4.5.

4.1. Data setup and processing

We obtain delay statistics from the FAA Aviation System Performance Metrics (ASPM) database for the time period 2008 to 2017 (Federal Aviation Administration (FAA) accessed 2018). The analysis is limited to the busiest 30 airports in the US (FAA Core 30); we then compute the total delay at each of these airports during each day, defined as 0000Z to 2359Z. The total delay at an airport (i.e., a node in the graph) is the sum of the arrival and departure delays of all flights at that airport during the day. Consequently, we obtain 3,653 graph signal vectors (each of dimension 30), one corresponding to each day in the data set. The edge weights of the graph are the sample Pearson correlation coefficients based on the 10 years of data (Equation (1)). It is worth noting that all correlations estimated from data are strictly positive, and there is no need for a non-negative projection as discussed in Section 3. This process results in a graph with 30 nodes, $\binom{30}{2} = 435$ edges, and 3,653 instances of delay signal vectors on these nodes.

Figure 3 depicts the resultant correlation matrix as a heat map (Figure 3(a)) and as a geographical map (Figure 3(b)). Two distinct sub-networks demarcating major East Coast and West Coast airports can be seen in Figure 3(b), in addition to a smaller sub-network for the Midwest airports (MSP, ORD, MDW, and DTW) as well. Many airport pairs on the East Coast are connected by edges with high correlation coefficients. In other words, when the delay at one East Coast airport (e.g., IAD) is high, then it is likely that the delays at other East Coast airports (e.g., DCA,

BWI, PHL, etc.) will also be high. These relationships are due to heavy traffic connectivity, geographic proximity, and a higher likelihood of these airports being impacted by the same disruptions and traffic management initiatives (TMIs). By contrast, the two Chicago-area airports – O'Hare (ORD) and Midway (MDW) – are less than 20 miles apart with no commercial air traffic operating between them, and yet have highly correlated delays due to similar weather and TMI impacts. The correlations between delays are determined by a combination of geographical proximity, airline operations, scheduling practices, and traffic flows.

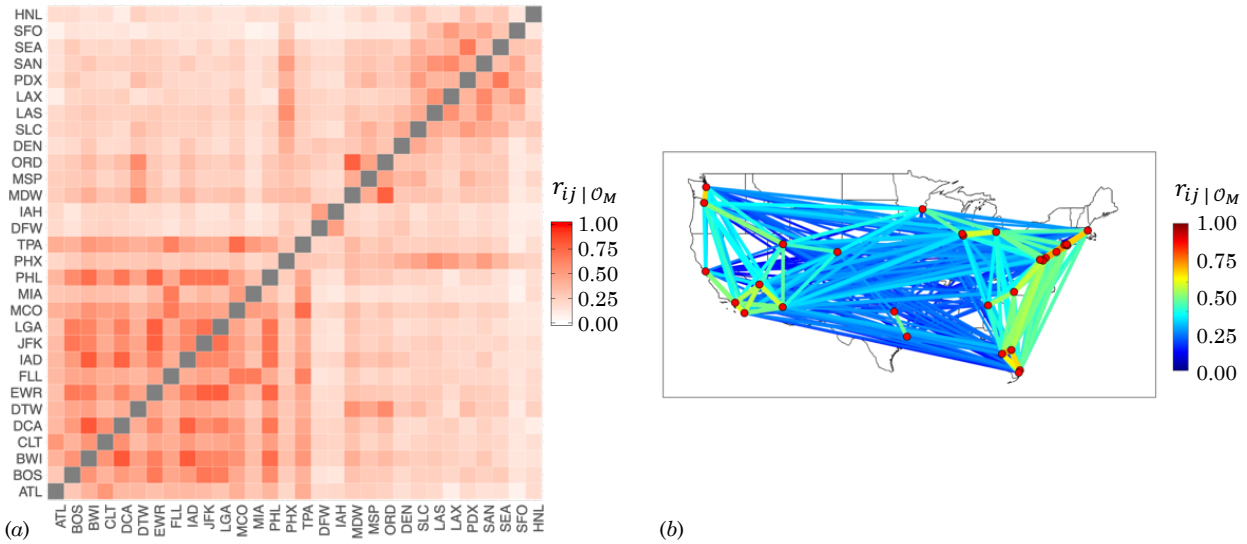

Figure 3 **(a) Heat map displaying the delay correlation between the top 30 airports; (b) Correlations shown with geographical context. Higher correlations are also emphasized with wider lines in (b). Note that HNL is not shown in (b) for simplicity.**

4.2. Spectral analysis

We compute the Laplacian $\mathfrak{L}$ from the adjacency matrix as per Definition 1, and compute its eigenvectors ($\lambda_1 < \cdots < \lambda_{30}$) and eigenvector modes ($v_1, \cdots v_{30}$). These eigenvector modes form the basis for the space of airport delay signals; the eigenvector modes v_i corresponding to higher indices i are said to be more energetic and have a higher TV (Section 3).

Table 1 presents a qualitative description of all the 30 eigenvector modes and their corresponding eigenvalues, while Figure 4 provides a visualization of the first and last three eigenvector modes. The key feature of interest in the modes is the sign of the mode at an airport. For a given eigenvector mode, airports with the same sign contribute in a similar way to the total delay signal. Airports with a positive component within the eigenvector mode move in the opposite way to airports with negative components of the eigenvector modes. For example, the v_2 mode encodes the delay dynamics where SFO delays are moving opposite to delays at DFW, IAH, ATL, PHL, and MIA.

In other words, if the delay at SFO is high, then the delays at the latter group of airports is low, and vice versa. Note that the most energetic eigenvector modes – the most unexpected modes – all involve East Coast airports with differing delay trends. As another example, consider the eigenvector mode v_{29}. It captures very energetic and unexpected delay dynamics where EWR delays are trending opposite to other New York-area airports (JFK, LGA), as well as other major East Coast airports (BOS, PHL, and IAD). Recalling the relationships between TV and spectral energy (Definition 3 and Proposition 1) as well as expected versus unexpected graph signal outliers (Section 3.3), a qualitative, operational interpretation is that eigenvector modes such as v_1, v_2, and v_3 are delay dynamics that are more expected, whereas v_{28}, v_{29}, and v_{30} are rarer, more unexpected delay dynamics.

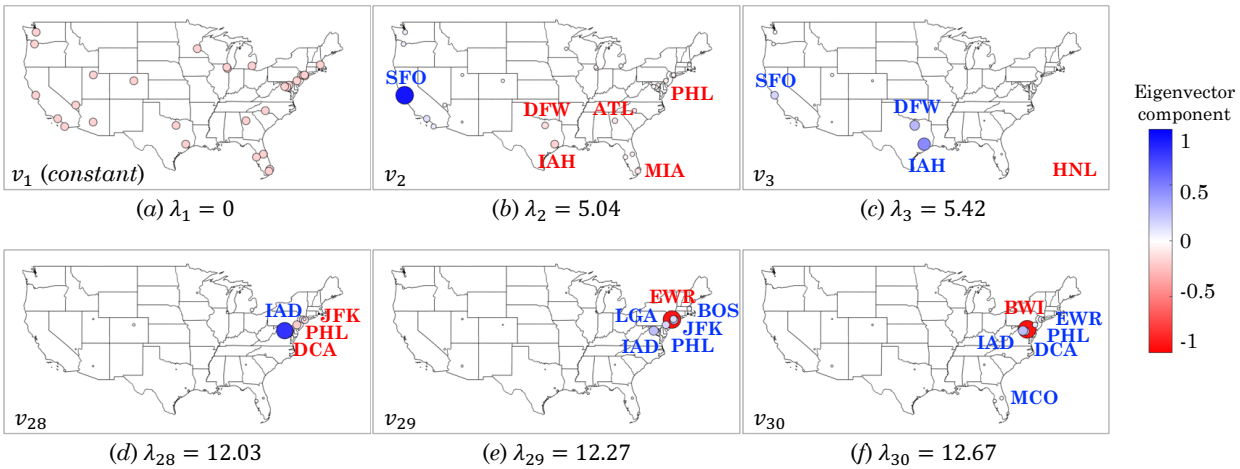

Figure 4 **Most (v_{30}, v_{29}, v_{28}) and least (v_1, v_2, v_3) energetic eigenvector modes of the system-wide graph Laplacian.**

As discussed in Section 3, an airport delay graph signal vector for any day can be decomposed into linear combinations of eigenvector modes v_1 through v_{30}. For each day in 2008 through 2017, we compute the spectral energy contributions of all 30 eigenvector modes, and plot the average contributions across the entire 10-year time frame in Figure 5. The first mode v_1, which is the constant mode, accounts for 80% of the energy, and we only plot the energy contribution of the remaining modes v_2 through v_{30}. A higher percentage of spectral energy contribution indicates that a particular eigenvector mode – and hence, a particular delay pattern – contributes more to the overall delay dynamics of a typical day in the NAS.

We interpret four eigenvector modes with relatively high spectral energy contributions based on Figure 5. Eigenvector mode v_{26} (contributing approximately 1.2% of the spectral energy) indicates that an interesting delay dynamic occurs when the two major New York City airports (JFK and

v_i	λ_i	Trend 1	Trend 2
v_1	$\lambda_1 = 0$	*Constant*	*Constant*
v_2	$\lambda_2 = 5.04$	ATL, MIA, PHL, DFW, IAH	SFO
v_3	$\lambda_3 = 5.42$	HNL	DFW, IAH, SFO
v_4	$\lambda_4 = 5.67$	ATL, BOS, BWI, CLT, DCA, DTW, EWR, FLL, IAD, JFK, LGA, MCO, MIA, PHL, TPA	DFW, IAH, HNL
v_5	$\lambda_5 = 6.29$	IAH	DFW
v_6	$\lambda_6 = 6.91$	DEN, SLC, LAX, PDX, SEA	ATL, FLL, MIA, SFO, HNL
v_7	$\lambda_7 = 7.65$	MIA, LAX, PDX, SEA	MSP, DEN
v_8	$\lambda_8 = 7.87$	ATL, DTW, MDW, MSP, ORD	MIA, DEN, LAX
v_9	$\lambda_9 = 7.97$	PHX, LAS, LAX, SAN	MIA, DEN, SEA
v_{10}	$\lambda_{10} = 8.27$	ATL, CLT, DEN, LAX, SEA	MIA, MDW, MSP, ORD
v_{11}	$\lambda_{11} = 8.48$	ATL, MIA, MSP, SLC, PDX	BOS, BWI, DCA, EWR, IAD, JFK, LGA, PHL, SEA
v_{12}	$\lambda_{12} = 8.71$	SLC, LAS, PDX, SAN	MSP, DEN, LAX, SEA
v_{13}	$\lambda_{13} = 8.91$	MDW, ORD, LAS, SAN	MSP, LAX, PDX
v_{14}	$\lambda_{14} = 8.95$	MDW, ORD, LAX, PDX	LGA, MSP, SLC, LAS, SAN, SEA
v_{15}	$\lambda_{15} = 9.03$	LAS, PDX, SAN	SLC, LAX
v_{16}	$\lambda_{16} = 9.20$	PHX, LAS	ORD, SLC, SAN
v_{17}	$\lambda_{17} = 9.84$	PHX	BOS, ORD, SLC, LAS, LAX, SAN
v_{18}	$\lambda_{18} = 9.89$	PHX, ORD	MDW
v_{19}	$\lambda_{19} = 10.20$	FLL, MCO, TPA	BOS, DCA, EWR, IAD, JFK, LGA, MIA, PHL
v_{20}	$\lambda_{20} = 10.33$	ATL, BOS, DTW, FLL, JFK, LGA	CLT
v_{21}	$\lambda_{21} = 10.83$	BOS, CLT, FLL, LGA, MDW, ORD	DTW, MCO, TPA
v_{22}	$\lambda_{22} = 10.85$	DCA, LGA, MCO, TPA	BOS, CLT, DTW, FLL
v_{23}	$\lambda_{23} = 11.01$	BWI, DCA, FLL, IAD, LGA, PHL	BOS, MCO, TPA
v_{24}	$\lambda_{24} = 11.30$	BWI, DCA, IAD, MCO, PHL	CLT, DTW, LGA, TPA
v_{25}	$\lambda_{25} = 11.35$	DCA, IAD, PHL, TPA	LGA, MCO
v_{26}	$\lambda_{26} = 11.57$	JFK, EWR, PHL	DCA, BOS, BWI, IAD, LGA
v_{27}	$\lambda_{27} = 11.71$	PHL, MCO	DCA, IAD, JFK
v_{28}	$\lambda_{28} = 12.03$	DCA, JFK, PHL	IAD
v_{29}	$\lambda_{29} = 12.27$	EWR	BOS, IAD, JFK, LGA, PHL
v_{30}	$\lambda_{30} = 12.67$	BWI	DCA, EWR, IAD, MCO, PHL

Table 1 **Description of eigenvector modes, delay trends (trends 1 and 2 move in opposite directions), and the airports involved.**

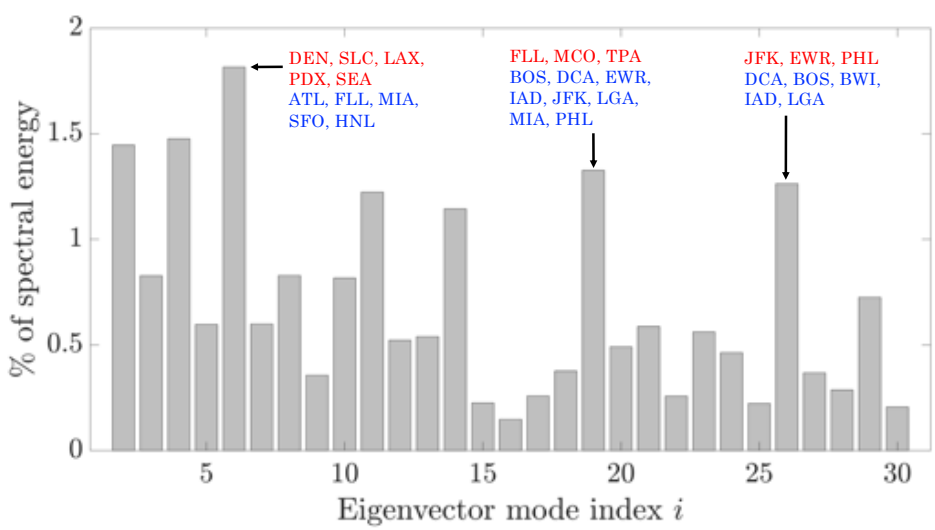

Figure 5 **Average spectral energy across each system-wide eigenvector mode; eigenvector modes v_2 through v_{30} is shown, with the constant mode v_1 removed for fair comparisons.**

EWR) along with the close-by PHL have significantly higher or lower delay magnitudes as compared to the remaining New York airport (LGA) and other major East Coast hubs (DCA, BWI, IAD and BOS). This disagreement in terms of airport delays not only occurs within a very localized level (LGA being different from EWR and JFK), but also at a regional level (New York City and Philadelphia being different from Boston and the Washington DC area). Another mode that accounts for a little over 1.3% of the average spectral energy is v_{19}. This mode describes a similar pattern as v_{26}, where there are local as well as region-wide disagreements in delay trends. Here, Florida airports (FLL, MCO, and TPA) have different delay magnitudes in comparison to several major East Coast airports and MIA, which is geographically close to FLL, MCO, and TPA.

Two other eigenvector modes that contribute around 1.2% of the average spectral energy are v_{14} and v_{11}. These modes highlight the need to perform an airline-specific analysis (in addition to the system-wide studies), since these two modes strongly suggest delay dynamics involving major Delta Air Lines (DL) hubs. Specifically, we see DL hubs such as LGA, MSP, SLC, and SEA grouped together in v_{14}, whereas v_{11} indicate delay dynamics where ATL, MSP, and SLC delays move opposite to BOS, JFK, LGA, and SEA.

4.3. Evaluation of outliers using total variation and total delay

Recall from Section 3 that TV and TD provide a low-dimensional projection for analyzing multivariate graph signals. Figure 6 plots the TV and TD for the airport delay graph signals for each day in the 2008-2017 data set (3,653 days). The bounds for outliers in scale as well as for the weak and strong outliers in distribution are computed for level $k = 4$ and plotted. It is worth noting that the lower bounds for the outliers in scale and the weak outlier in distribution are negative and not plotted. This highlights the significant level of conservatism in these bounds in relation to the thresholds for strong outliers in distribution. The primary factor that makes the weak bounds conservative is that they are not dependent on the TD, which does not allow them to capture the increasing variance in the TV for higher values of TD. For further discussions and empirical results on the gap between the strong and weak bounds, we refer the readers to Gopalakrishnan, Li, and Balakrishnan (2019). Additionally, the choice of k is important to consider when deploying our outlier detection methodology. One approach could be to decide on a reasonable value *a priori*, as done in other statistical procedures, and continue the analysis. Another approach could be to choose a k such that small perturbations in k does not result in large changes in outlier populations. This could be done through examining a plot of the percentage of outliers versus k, similar to an elbow plot in cluster analysis.

From Figure 6, we observe that 167 days (4.6 % of days) were classified as strong distribution outliers, 221 days (6.0% of days) were classified as weak outliers in distribution only, no days

were classified as outliers in scale only, and 14 days (0.4% of days) were classified as both weak outliers in distribution as well as outliers in scale. Another observation from Figure 6 is that the TV typically increases with an increase in TD, since the TV is related to individual airport delays via a non-negative quadratic polynomial. For all subsequent discussions and results in this paper, we only examine strong outliers in distribution, as it is the tightest in terms of its bounds.

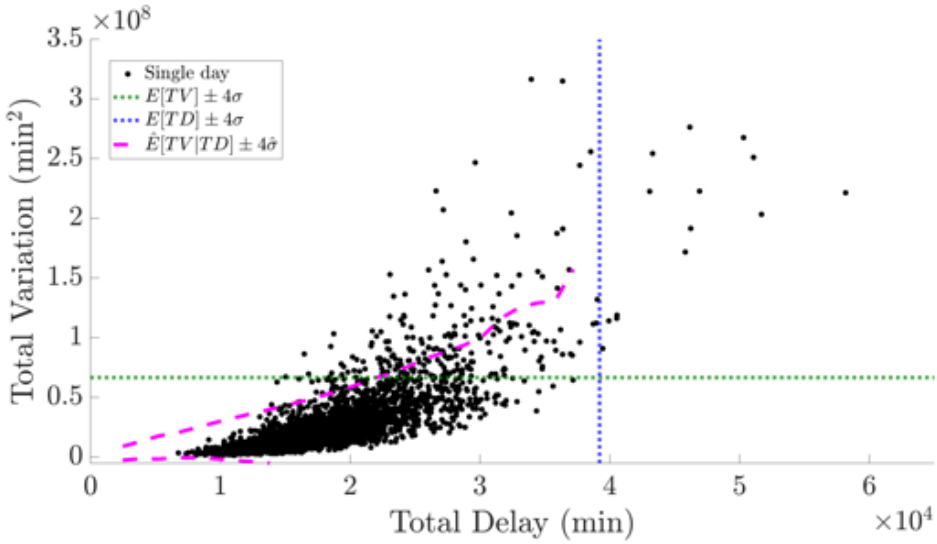

Figure 6 **TV versus TD for all days in 2008-2017 with level $k = 4$ weak and strong outlier bounds demarcated.**

4.4. Identifying disruptions for further analysis

A central motivation for our usage of GSP is to characterize differing delay patterns that result from aviation disruptions. To this end, we would like to analyze the subset of days (data points) in Figure 6 that experiences a particular type of disruption. Such an analysis would determine if certain types of disruptions are correlated with an unusually large number of outliers. We use two independent systems of categorizing disruption days and creating subsets for further analysis. The first categorization is based on external disruptions, whereas the second is based on delays and cancellations.

In the first categorization of days, we identified four specific disruptions: nor'easters, hurricanes, airport or airline outages, and thunderstorms. A total of 178 days out of the 10-year period was labeled with one of the four types of disruptions (see Table 12 in supplementary materials (SM) for a list of these dates), and the metrics used for identifying days are as follows:

Nor'easters: Nor'easters are large convective systems that typically impact the East Coast and are associated with heavy rain or snowfall. These disruptions typically occur between September and April, are well-predicted a few days in advance, and usually result in severe airport and airspace capacity reductions. We use the Regional Snowfall Index (RSI) metric (Squires et al. 2014), along

with an estimate of financial damage to identify 60 days in our data set which are affected by nor'easters (National Weather Service 2019).

Hurricanes: We consider Atlantic hurricanes that primarily impact the southern and southeastern coastal regions of the US, and the East Coast in rare circumstances. We considered three factors when selecting our list of 34 hurricane-type days: (1) the Saffir-Simpson hurricane wind scale (National Oceanic and Atmospheric Administration 2018b), (2) the geographic region of impact must include the contiguous US (National Oceanic and Atmospheric Administration 2018a), and (3) the resultant financial costs (Weinkle et al. 2018). Similar to nor'easters, they are well predicted storm systems, and impact air traffic operations for several consecutive days.

Airline and airport outages: Airline-specific and airport-specific outages typically occur due to equipment failure, and occasionally due to security-related incidents. Some examples of root causes include power outages that affect an airport, computer or hardware malfunctions affecting the flight dispatch system of one airline, and outages that affect third-party global distribution systems and computer reservation systems. These outages are typically localized to one specific airline or airport, or possibly a group of airlines using the same service provider. We used online news sources to identify 49 outage-type days (Yanofsky 2015).

Thunderstorms: Unlike the other three types of disruptions, thunderstorms are quite common, occur over very localized regions (on the order of a few hundred miles), are rapidly evolving, last only for a couple of hours, and very difficult to predict. Since there is no standardized way of locating significant thunderstorm days, we use Ground Delay Programs (GDPs), which is a procedure used to reduce the demand at affected airports, as a proxy of thunderstorm activities. Severe thunderstorm activity days in summer months are identified using a clustering procedure described in Gopalakrishnan, Balakrishnan, and Jordan (2016a), then cross-referenced with weather radar maps to confirm the presence of convective activity (National Oceanic and Atmospheric Administration 2019). Subsequently, 35 days with severe system-wide disruptions due to thunderstorms are identified.

In terms of operational performance measures, the amount of flight delay as well as cancellations are crucial metrics used by airport managers, airlines, as well as ANSPs. This forms the basis for the second categorization of days. To this end, we use the delay and cancellation clusters to assign a label indicating the delay and cancellation levels on a day-by-day basis. The clustering methodology and subsequent mapping into low or high delay and cancellation levels is discussed in Gopalakrishnan, Balakrishnan, and Jordan (2016a). The four labels that any day can be assigned to are:

- **Low delay; Low cancellation (DLCL):** This is the most common type of day (1044 days, 74.7% occurrence) with relatively normal operations throughout the system.

- **Low delay; High cancellation (DLCH):** These days (54 days, 3.9% occurrence) are typically indicative of proactive cancellations by airlines in anticipation of severe disruptions (e.g, a northeaster snowstorm). The huge reduction in flight volumes provides ample schedule buffer and results in low delays.
- **High delay; Low cancellation (DHCL):** Such days (170 days, 12.2% occurrence) may be indicative of an unplanned or poorly forecasted disruption such as pop-up thunderstorms, giving airlines little chance to proactively cancel.
- **High delay; High cancellation (DHCH):** The most severe unplanned disruption typically leads to significant delays and cancellations. These are the days (130 days, 9.3% occurrence) with the worst system impact.

Figure 7(a) depicts the 178 days classified as nor'easters, hurricanes, airport outages, or thunderstorms. Note that we use the same bounds for outliers in scale, as well as the weak and strong outliers in distribution as Figure 6, since we are still searching for outliers in the context of the entire system across the 10-year span. Figure 7(b) presents all the days from January 1, 2014 through October 31, 2017 partitioned into one of the four delay-cancellation groups. Because of limited availability of the complete cancellation data set used for clustering, we are restricted to a shorter time span. The coordinates of the centroids for each of the subset of the days is also plotted to provide a high-level overview. The counts for outliers in all these cases, as well as a discussion and interpretation of the results is presented in the next subsection.

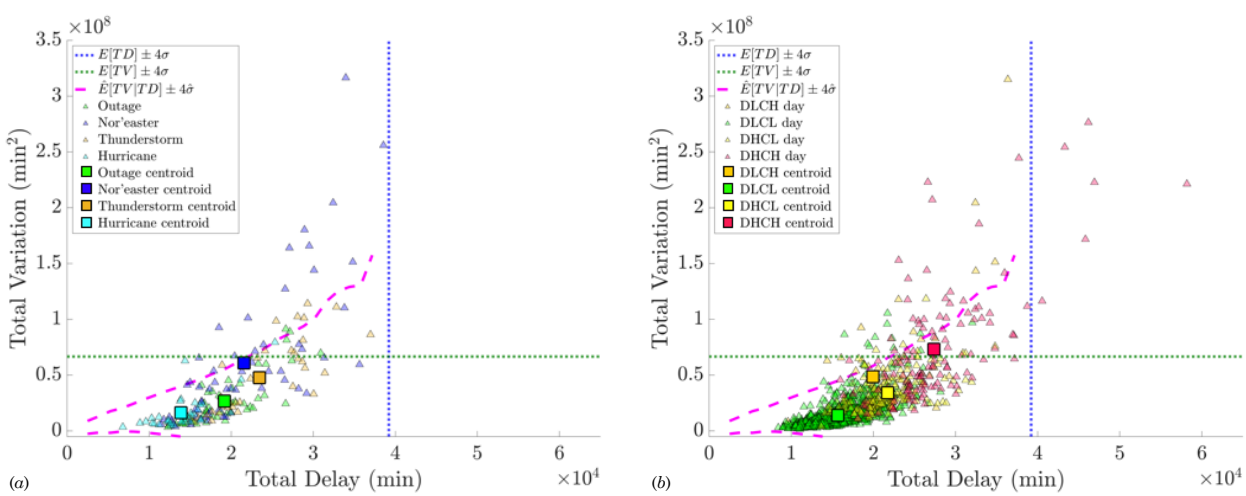

Figure 7 **TV versus TD plot for a subset of days; (a) 2008-2017 with four disruptions, and (b) 2014-2017 with four system-wide delay and cancellation conditions. The average value (centroids) for each category is also shown.**

Category	Outlier counts	%
Nor'easter	17 out of 60	28.3%
Hurricane	1 out of 34	2.9%
Thunderstorm	5 out of 35	14.3%
Outage	2 out of 49	4.1%

Table 2 Outliers for the four types of disruptions

4.5. The role of disruption in spatial delay distributions

We present the strong outliers in distribution statistics for the four types of disruptions (nor'easters, hurricanes, airline- or airport-specific outages, and thunderstorms) in Table 2, and for the delay-cancellation subsets (low delay with low cancellation, low delay with high cancellation, high delay with low cancellation, and high delay with high cancellation) in Table 4.

We see a clear distinction in the occurrence of strong distribution outlier days for the four disruption categories (Table 2). Taken together, the hurricanes and outages-type days only result in 3 days out of 83 being strong distribution outliers (3.6%). On the other hand, the system-wide impact of thunderstorms and nor'easters were higher in terms of unexpected spatial distribution of delays, and a total of 22 days out of 95, or 23.2% were classified as outliers. This is significantly higher than the outlier counts for hurricane- and outage-type disruptions. Thunderstorms and nor'easters are thus correlated with higher TV, higher TD, and more outliers, while airport outages and hurricanes are correlated with a lower occurrence of outliers.

The low TD and low TV characteristics of hurricane days are interesting and surprising, since hurricanes are extremely disruptive to the air transportation system. These results indicate that not only are hurricanes correlated with lower delays, but these delays are also distributed in an expected manner. This is in direct contrast with nor'easters, which are also very disruptive but result in higher delays, higher TV, and result in more unexpected distributions of delays. One could argue that cancellations (Bureau of Transportation Statistics 2015), which are not accounted for in our analysis, may offer an explanation. However, this is not the case. In Table 3, we list the average cancellation percentages across all days belonging to each of the four disruption categories, including the 10-year average, for the entire system as well as the four major airlines. We observe that hurricanes and nor'easter have comparable system-wide cancellation percentages, but nor'easters still result in higher delays and TV.

This difference in outlier occurrences may reflect differing operational philosophies when dealing with irregular operations (IROPS) stemming from each of the four disruption types. Specifically, hurricanes tend to be well-predicted in terms of its projected trajectory, giving airlines time to proactively cancel. Hence, we see hurricane-type days not only with low delay, but also expected distributions of delay. On the other hand, nor'easters may not be associated with airlines canceling

strategically and efficiently re-positioning aircraft to enable swift recovery, even though nor'easters may be well-predicted. This results in higher delays, higher TV, and unexpected distribution of delays. It is also possible that the regions typically affected by these nor'easters, i.e., the Mid-Atlantic and New England regions, involve highly congested airports within high-traffic density airspace which are already operating at their capacity limits, further exacerbating the problem. Our data-driven analysis clearly highlights the current challenges faced by airlines regarding proactive management of these nor'easters, and motivates the need to develop sophisticated tools for disruption recovery and management.

The more spontaneous nature of airport outages do not give airlines the luxury to proactively cancel, resulting in outage days having more incurred airport delays than hurricane days. However, since outages tend to be isolated to one particular airport or airline, its effects on the overall spatial distribution of airport delays within the entire system is limited, thus resulting in low levels of TV. Lastly, thunderstorms are geographically local and temporary phenomenon. These characteristics do not afford airlines a long prediction and planning horizon; thus, airlines typically try to operate through thunderstorms, preferring to incur moderate delays while avoiding cancellations. This explains the higher TD values associated with thunderstorms. However, since these events affect only a small fraction of the traffic at any instant, they do not lead to large-scale changes in the delay distribution, and hence are correlated with lower occurrences of outliers.

Interestingly, while it seems that nor'easters result in the largest impacts when it comes to the spatial distribution of airport delays at a system-wide level in comparison to airport outages or thunderstorms, we will see that this does not hold in the airline-specific analysis (Section 5).

Category	**System-wide**	**American (AA)**	**Delta (DL)**	**United (UA)**	**Southwest (WN)**
Nor'easter	7.9%	8.1%	8.0%	8.2%	5.1%
Hurricane	7.8%	7.3%	5.7%	9.6%	6.9%
Outage	2.3%	3.1%	2.5%	1.4%	2.2%
Thunderstorm	2.9%	3.1%	2.0%	1.9%	1.4%
10-year average	1.6%	1.8%	1.0%	1.4%	1.1%

Table 3 Percentage of flights canceled across the entire system as well as for each of the four airlines under different disruption categories.

We also observe an insightful relationship that links flight cancellations with strong outliers in distribution. From Table 4, it is apparent that high levels of cancellation, irrespective of whether it is associated with high or low delays, is correlated with higher outlier counts. Out of the 1,214 days with low flight cancellation levels, only 1.5% were outliers; but when the cancellations are high, almost 22% of the days are outliers. A possible explanation is that when flights are not cancelled,

Category	Outlier counts	%
Low delay, low cancellation (DLCL)	12 out of 1044	1.2%
Low delay, high cancellation (DLCH)	12 out of 54	22.2%
High delay, low cancellation (DHCL)	6 out of 170	3.5%
High delay, high cancellation (DHCH)	33 out of 130	25.4%

Table 4 Outliers for the delay-cancellation categories

they typically propagate delays based on their route structure and connectivity, and spread the delays across the system. While this increases the system-wide delay, it is more homogeneous, and hence decreases the TV. On the other hand, cancellations isolate parts of the network and prevent the propagation of delays. This results in significantly lower delays downstream in the schedule, since the cancelled aircraft cannot complete those routes. Although we caution that further work is required to ascertain the causal direction of this relationship between cancellations and high TV, this motivates the usage of flight cancellations as a control action to guide system-wide recovery towards more expected spatial delay distributions, if such a state is desired.

Note that our cancellation-based analysis is a proxy for various latent operational factors that could have resulted in or exacerbated an unexpected spatial delay distribution. These include localized disruptions such as pop-up thunderstorms and outages that do not have widespread system impacts, traffic management initiatives with an unexpected scope of coverage, or even sudden demand surges (e.g. additional scheduled flights for certain sporting events) or temporal shifts (e.g. unscheduled aircraft maintenance). One important way to further this analysis is to examine the situation from the perspective of individual *airline sub-networks*. Hence, in Section 5, we present the spectral analysis, outlier identification, and disruption impact assessment individually for four major US carrier. We also analyze the complex relationship between the system as a whole in comparison to the individual sub-networks of these carriers.

5. Airline-specific analysis

Several results in the previous section hint at the necessity to zoom in at an airline-specific level. First, some of the eigenvector modes focus on hub airports for specific airlines indicating a deeper, airline-specific effect. Second, disruptions such as thunderstorms or nor'easters tend to affect specific geographies, consequently impacting some airlines more than others. Finally, even though airlines may be affected individually, the system-wide view aggregates these variations and does not capture the subtleties and nuances of airline operations.

We detail the data setup in Section 5.1, the analysis of the spectral modes in Section 5.2, the identification of outliers using TV versus TD plots in Section 5.3.1, and the comparisons between system-wide versus airline-specific in Section 5.3.2.

5.1. Data setup and processing

Since ASPM does not provide airline-specific breakdowns of airport delays, we use publicly available on-time performance data retrieved from the Bureau of Transportation Statistics (BTS) for the time frame of January 1, 2008 through December 31, 2017 (Bureau of Transportation Statistics 2015). The data pre-processing involves filtering for flights arriving at, or departing from our Core 30 airport list, aggregating delays over the day, adjusting for multiple time zones, and eliminating canceled and diverted flights. Finally, we restrict our study to four airlines that all together account for approximately 79% of departed seats for all domestic US traffic (Bureau of Transportation Statistics 2019): American Airlines (AA), Delta Air Lines (DL), United Airlines (UA), and Southwest Airlines (WN).

For each of the four airlines, we have a corresponding unique non-negative correlation matrix that serves as the airline-specific adjacency matrix; for brevity, these airline-specific correlation heat maps are presented in Figures 11 and 12 in SM. Some of these four airlines do not serve all 30 airports during the time frame of our analysis; hence the graphs for WN had 24 nodes (no operations in ORD, MIA, JFK, DFW, IAH, and HNL), AA and UA have 29 nodes each (no AA or UA operations at MDW), while DL services all 30 airports. Thus, in our ordered indices i for eigenvector modes v_i and eigenvalues λ_i, the highest i for AA, DL, UA, and WN will be 29, 30, 29, and 24, respectively.

In AA's network, we see a fairly uniform distribution of strong correlations mostly focused on their East Coast hubs (e.g. CLT, DCA, LGA) as well as their hub at DFW. In contrast, the DL network reflects a much strong presence of airport delay correlations in the East Coast, and is more similar to that for the system-wide network. UA's network highlights correlations in the West Coast and Midwest, centered around SFO, DEN, and ORD. There are also noticeable airport pairs that have zero correlations, indicating little or non-existent UA operations between that specific airport pair (e.g. ATL and JFK). Finally, WN has a few airport pairs with high correlations (e.g. TPA-MCO, DCA-ATL), and no airports with a significant number of high correlation edges incident on them (except PHX and a few west coast cities to a mild extent). This emphasizes the intrinsically different network structure, routing strategies, and tail assignment by WN compared to the other three airlines.

5.2. Spectral analysis

We plot geographically the most energetic and the second most energetic eigenvector modes for each airline in Figure 8. Note that comparing eigenvalues across airlines is not meaningful as they originate from different graphs Laplacians, but within a particular airline, the eigenvalues retain the same interpretation of graph "frequencies" as discussed in Section 3.

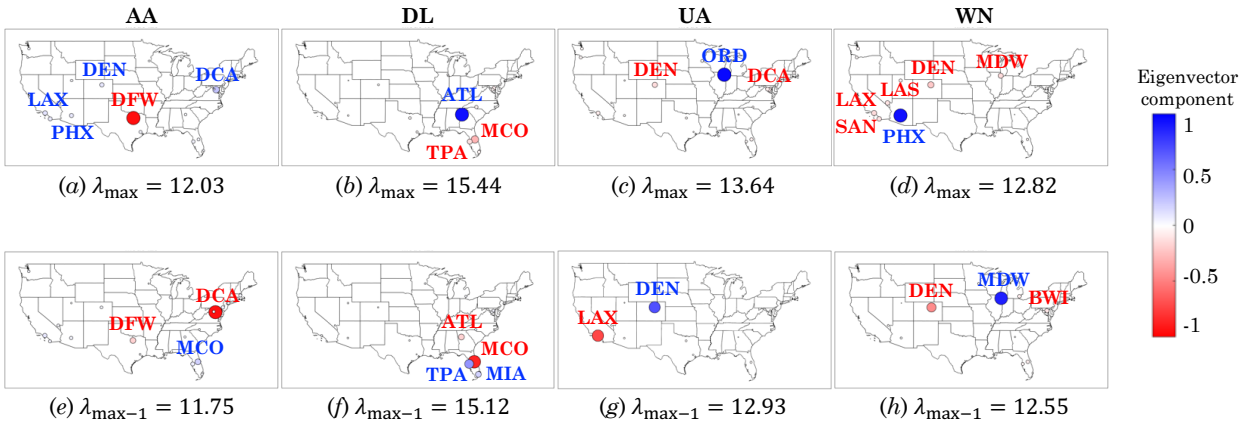

Figure 8 **Most ($\lambda_{\max}$) and second most ($\lambda_{\max-1}$) energetic eigenvector modes for AA, DL, UA and WN.**

At the airline-specific level, we see interesting patterns emerge within the top two highest-energy eigenvector modes that are not captured at the system-wide scale. For AA, DL, and UA, the most energetic eigenvector mode depicts a spatial distribution of delay where delays at the corresponding airline's largest hub (see Table 8 in SM for annotated table of airline hubs) are not in-sync with delays at other major hubs for that airline's sub-network. Additionally, while the eigenvectors for AA, UA, and WN target their hubs or focus cities, the two most energetic modes for DL involve only ATL and other geographically proximate Florida airports. This highlights the significant density of hub operations at ATL by DL, and the relatively small network presence of other DL hubs in comparison to ATL.

In Figure 9, we plot the distribution of spectral energy across each airline's eigenvector modes. Similar to the system-wide case, the constant eigenvector mode (v_1) accounts for a large portion of the average spectral energy (61.1%, 59.4%, 58.1%, and 66.1% for AA, DL, UA, and WN, respectively); we do not show this constant mode in order to highlight the subtleties of the other modes. The network legacy carriers (AA, DL, and UA) are similar to each other in the sense that their top eigenvector mode contributes significantly to the spectral energy. In other words, for these carriers, delays at their largest hubs move opposite to other airports sufficiently frequently so that $v_{\max}$ contributes to a high percentage of spectral energy. This is in contrast to WN which appears to loosely follow a power law decay in energy across higher modes, and has higher contributions from less-energetic modes such as v_2, v_3, v_4, and v_5.

For the three network legacy carriers, there are also some lower-energy modes that contain a high percentage of spectral energy. In particular, these include v_6 for AA, v_8 for DL, and v_{11} for UA. For these three eigenvector modes, they typically involve one or two airports that are hubs for the specific airline, but also many other airports that tend to be hubs for other network legacy carriers. Interestingly, we will see in Section 5.3.2 that these lower-energy modes play a dominant role on days when some airline sub-networks are outliers, but others are not.

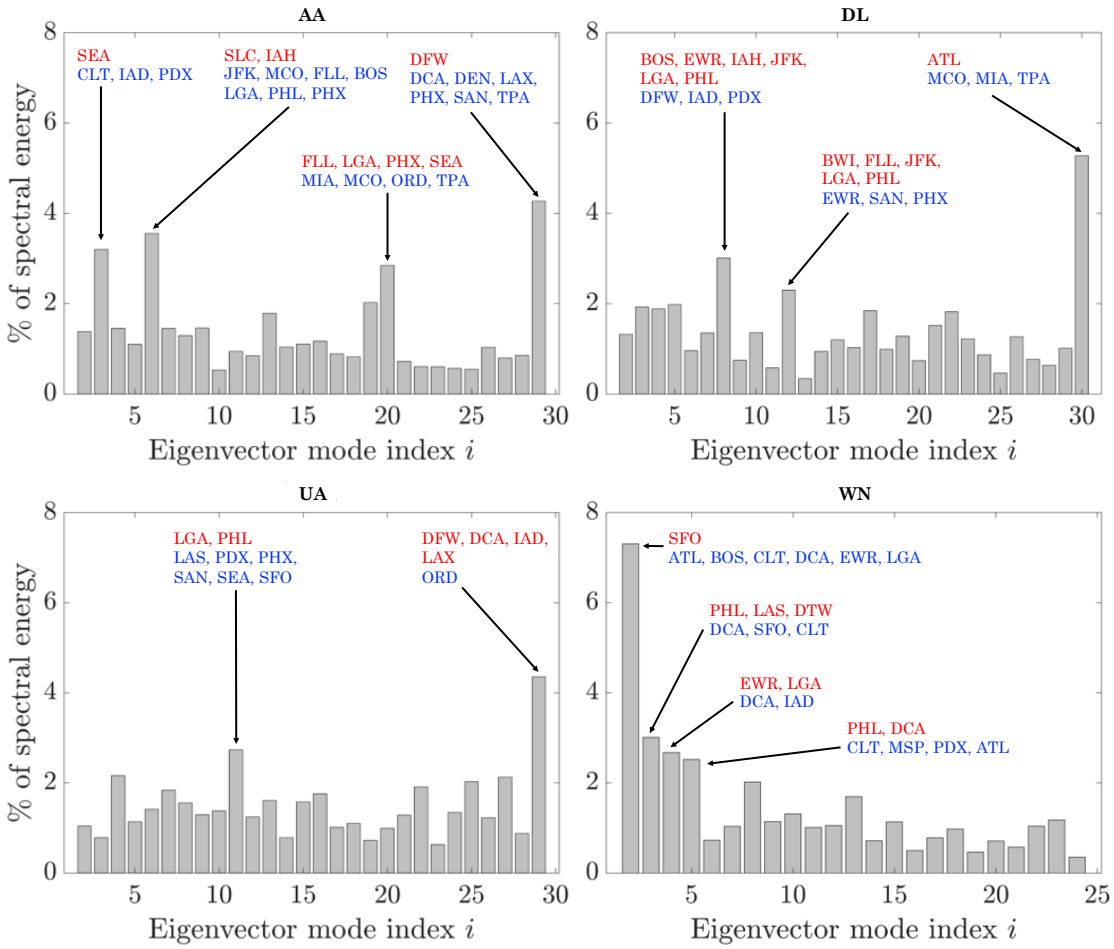

Figure 9 **Average spectral energy across each eigenvector mode for all four airlines. v_1 is removed for all airlines.**

5.3. Total variation and total delay

We analyze delay signals and graph Laplacians that are airline-specific to highlight the differences between delay dynamics, spatial delay distributions, and response to disruptions by individual airlines. Specifically, we identify the airline-specific outlier counts for each disruption (Section 5.3.1), and interpret the relationships between the airline-specific versus system-wide analysis (Section 5.3.2). For the disruption analysis, we consider the same set of 178 days from Section 4.4.

5.3.1. Discussion of airline-specific outliers We first compute airline-specific strong distribution outlier bounds; the TV versus TD plots with bounds for each airline can be found in the SM. The empirical strong outlier bounds for AA, UA, and WN are similar, with DL exhibiting significantly wider bounds. We consistently see that airline- or airport-specific outages and thunderstorms have greater effects on the spatial delay distribution of the airline sub-networks than the system-wide network. Furthermore, while nor'easters had the greatest effect on system-wide spatial delay distributions, their effect at the airline sub-network level is diminished. While AA,

UA, and WN disruption centroids were within the empirical strong outlier bounds, the DL centroids for thunderstorms and outages are outside the bounds. This indicates that even an "average" thunderstorm or outage event typically results in unexpected spatial delay distributions in DL's sub-network.

Airline-specific outlier statistics are compiled in Table 5, along with corresponding system-wide outlier statistics. Summary percentages we quote in this discussion can be computed straightforwardly from Table 5 by conditioning on the appropriate airline. We observe that disruptions affect airline sub-networks quite differently compared to the system-wide network. For example, an average of 39% of thunderstorm-type days were strong outliers in distribution for one of the three network legacy carriers, with DL having over half (51.4%) of its thunderstorm-type days classified as outliers, whereas only 14.3% of thunderstorm-type disruption days were classified as outliers in the system-wide analysis. The strong hub-and-spoke nature of these airline operations along with their routing strategies may contribute to the significant operational impact of transient disruptions such as pop-up thunderstorms. On the other hand, the more point-to-point nature of WN may explain why only 8.6% of their thunderstorm-type days are classified as outlier in distribution.

Category (***System-wide outlier %***)	**Airline**	**Outlier counts**	**%**
Nor'easter (*28.3%*)	AA	7 out of 60	11.7%
	DL	9 out of 60	15.0%
	UA	10 out of 60	16.7%
	WN	8 out of 60	13.3%
Hurricane (*2.9%*)	AA	0 out of 34	0.0%
	DL	1 out of 34	2.9%
	UA	0 out of 34	0.0%
	WN	2 out of 34	5.9%
Outage (*4.1%*)	AA	5 out of 49	10.2%
	DL	9 out of 49	18.4%
	UA	10 out of 49	20.4%
	WN	7 out of 49	14.3%
Thunderstorm (*14.3%*)	AA	10 out of 35	28.6%
	DL	18 out of 35	51.4%
	UA	13 out of 35	37.1%
	WN	3 out of 35	8.6%
10-year span (*4.6%*)	AA	292 out of 3,653	8.0%
	DL	301 out of 3,653	8.2%
	UA	288 out of 3,653	7.9%
	WN	355 out of 3,653	9.7%

Table 5 Outlier counts and percentages for each type of disruption and specific airline.

There are also differences in the spatial delay patterns caused by nor'easters at a system-wide level versus an airline-specific level. On an airline-specific level, we see that an average of 14% of

nor'easter-type days are classified as an outlier for at least one of the airline, compared to over 28% for the system-wide analysis. At the system-wide level, nor'easters often cause unexpected spatial delay distributions owing to their propensity to impact the highly correlated East Coast and Mid-Atlantic region. However, unlike the system-wide case, the eigenvector modes with the highest spectral energies for airlines (Figure 8) tend to be more geographically diverse. In particular, rather than concentrating on the East Coast, these energetic eigenvector modes tend to correspond to AA, DL, and UA's largest hubs, which are geographically spread out. Furthermore, as discussed in Figure 9, these high energy eigenvector modes also tend to be frequently triggered.

Finally, outages result in lower system-wide outlier occurrences (4.1%) in comparison to airline-specific outliers (average of 15.8% of the days). This is because outage events typically involve only one specific airport or airline that experiences most of the disruptions, with little diffusion to the system-wide network. In general, spatial delay distributions within airline sub-networks are more easily perturbed than system-wide spatial delay distributions; this can be seen in the 10-year span outlier statistics, where 4.6% of all 3,653 days were outliers in the system-wide analysis, but outlier percentages varied between 7.9% (for UA) and 9.7% (for WN) when analyzing individual airlines.

5.3.2. System-wide versus airline-specific outliers In this discussion, we connect the system-wide outlier results from Section 4.3 with airline-specific outlier results. For each day in the 10-year data set, we assign five labels that indicate whether or not the system-wide network and each airline's sub-network was classified as a strong distribution outlier. This information can be represented in the form of a tuple (System-wide, AA, DL, UA, WN), where each entry flags a "×" if the corresponding network is an outlier. For example, the tuple (×, , , ,) represents a day where the system-wide network was a strong outlier, but no airline-specific sub-networks were outliers. In our 10-year time frame, all $2^5 = 32$ possible combinations had at least one day labeled as such; we list the tuple statistics in Table 6.

Some of the day-types from Table 6 have interesting operational implications. The first day-type of interest denotes the case where only one airline's sub-network has unexpected spatial delay distributions, but no other airline's sub-network or the system-wide network is exhibiting unexpected spatial delay distributions. We see a total of 164 such days for WN, 131 for DL, 103 for AA, and 84 for UA, totaling 482 days out of 3,653 (13.2%). This particular subset of days may be of interest for airlines, as they represent spatial delay distributions that were likely triggered by, and remain confined, to their own sub-network.

Another day-type of interest is when exactly one airline's sub-network is exhibiting unexpected spatial delay distributions and the *entire system* is an outlier in distribution as well. These days can quantify system resilience, since the unexpected delay distributions were not quarantined to

System	AA	DL	UA	WN	Outlier Counts	%
					2817	77.1%
				×	164	4.5%
		×			131	3.6%
	×				103	2.8%
			×		84	2.3%
			×	×	37	1.0%
	×			×	29	0.8%
×					23	0.6%
		×	×		19	0.5%
	×		×		16	0.4%
	×	×			15	0.4%
×		×			15	0.4%
×	×	×	×	×	15	0.4%
	×		×	×	14	0.4%
	×	×	×		14	0.4%
		×		×	13	0.4%
×	×		×	×	13	0.4%
×	×	×	×		13	0.4%
		×	×	×	12	0.3%
×			×		11	0.3%
×			×	×	10	0.3%
×	×				10	0.3%
×	×	×			10	0.3%
	×	×		×	9	0.3%
	×	×	×	×	9	0.3%
×		×	×		9	0.3%
×	×		×		8	0.2%
×	×	×		×	8	0.2%
×				×	7	0.2%
×	×			×	6	0.2%
×		×		×	5	0.1%
×		×	×	×	4	0.1%

Table 6 **Counts of the number of days belonging to each of the 32 tuple types.**

the sub-network of one airline. On the other hand, days where the system is not an outlier, but only one particular airline is, are are also worth analyzing. As mentioned in Section 5.2, there is a pattern of certain low-energy eigenvector modes being triggered for non-outlier airlines during days when other airlines might be exhibiting unexpected spatial delay distributions. This is highlighted in Figure 10, where we plot the average spectral energy distribution for days where the system is an outlier and exactly one of the four airlines is an outlier. We note that when AA, DL, and UA are not outliers, there is a noticeable increase in the occurrence of eigenvector modes v_6 for AA, v_8 for DL, and v_{11} for UA. Furthermore, all three aforementioned modes for AA, DL, and UA are triggered when WN is an outlier. The activation of these low-energy eigenvector modes indicate

interactions between airline sub-networks occurring at shared hub airports. Specifically, unexpected spatial delay distributions in one airline sub-network partially impact other airline sub-networks as well.

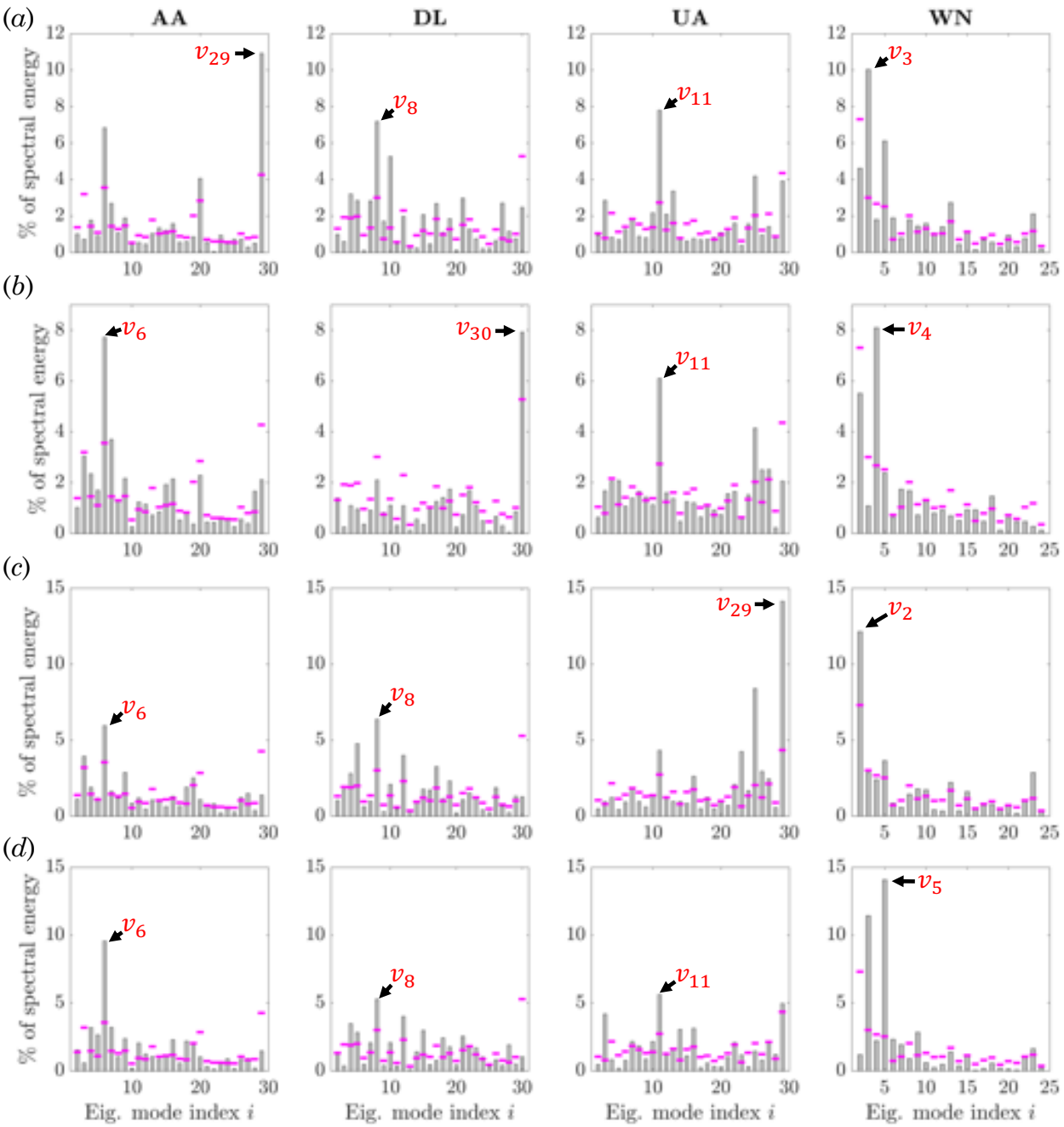

Figure 10 **Spectral energy averaged across days where only the system and one specific airline (AA, DL, UA, or WN in (a) through (d), respectively) has airport delays that are strong outliers in distribution (gray bars). The airline-specific average spectral energy across the 10-year time frame is shown for benchmarking purposes (magenta bars). Mode v_1 is not plotted in order to highlight the other modes with differing energy scales.**

We discuss a few case studies to illustrate the utility of our spectral analysis. Consider June 2, 2017, a day which was an outlier for AA and the entire system. A slow moving thunderstorm over DFW impacted operations out of the airport and caused delays. Since DFW is a major and influential AA hub, it is expected that delays at DFW would result in delays at other AA hubs. However, on this day, in spite of high delays and cancellations at DFW for AA, it did not spread to other airports. Thus, it was classified as an outlier for both AA and the system, but not for other airlines. Another example is September 11, 2017, where DL and the entire system were outliers. On this day, Hurricane Irma made landfall in the US southeast, resulting in the closure of all Florida airports as well as heavy flight delays and cancellations out of ATL. This resulted in a situation where ATL had high delays, but the Florida airports (MIA, MCO, TPA) had no delays due to airport closures. Given the historically strong correlations between ATL and the Florida airports, this was an extremely unusual distribution of delay, and thus classified as an outlier for DL. To add to the unexpected quality of this day, delays did not spread to other DL hubs such as MSP.

We emphasize the need to examine sub-network interactions, as we showed in Table 6 that the system-wide network is not simply a sum of the four airline sub-networks. This is further evident from the observation that there are 9 days in which the system is not an outlier but each of the four major carriers are, and 23 days in which the system is an outlier but none of the airlines are. This observation motivates future analysis to understand sub-network interactions and their emergent properties. Finally, the inventory of days belonging to each of the 32 tuple types from Table 6 can be found in Tables 10 and 11 in the SM.

6. Concluding remarks

In this work, we presented methods to analyze the spatial distribution of signals in networks, and applied these techniques to study airport delays in the US. Specifically, we formalized and defined notions of outliers in graph signals, then leveraged GSP to analytically as well as empirically identify these outliers. Outlier detection and spectral analysis were used to characterize and compare airport delays at a system-wide and airline-specific level in the US NAS. Our methods enable the automatic identification of outliers, providing airlines and ANSPs an inventory of days with unexpected delay distributions for further performance analysis. Such an inventory is essential for developing playbooks that will mitigate the element of surprise for controllers and flow managers due to unexpected delay distributions. Furthermore, we emphasize the contextual interpretability of outliers via the eigenvector modes. Our work presents the first network-wide spectral analysis of air traffic delays, outlier detection based on the spatial distribution of delays, and a quantification of the impacts of various disruptions on the system and airlines.

Our novel spectral-based approach for analyzing airport delays, as well as our outlier identification framework lead to several interesting future research directions, including applications of GSP

and spectral methods to the comparative analysis of different aviation systems, and the further development of the theoretical foundations of outlier detection for networked signals.

Acknowledgments

This work was supported by the National Science Foundation (NSF) through Award No.1739505 and an NSF Graduate Research Fellowship (Max Z. Li). We thank the anonymous reviewers whose comments and suggestions have greatly improved the clarity of our manuscript.

References

Abid A, Kachouri A, Mahfoudhi A, 2017 *Outlier detection for wireless sensor networks using density-based clustering approach. IET Wireless Sensor Systems* 7(4):83–90.

Ahmadbeygi S, Cohn A, Lapp M, 2010 *Decreasing airline delay propagation by re-allocating scheduled slack. IIE Transactions* 42(7):478–489.

Ahmed HB, Dare D, Boudraa AO, 2017 *Graph signals classification using total variation and graph energy informations. 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 667–671 (IEEE).

Airlines for America, 2018 *Industry Review and Outlook*. URL http://airlines.org/dataset/a4a-presentation-industry-review-and-outlook/.

Ball M, Barnhart C, Dresner M, Hansen M, Neels K, Odoni A, Peterson E, Sherry L, Trani A, Zou B, 2010 *Total delay impact study.*

Belkin M, Niyogi P, 2002 *Laplacian eigenmaps and spectral techniques for embedding and clustering. Advances in neural information processing systems*, 585–591.

Bronstein MM, Bruna J, LeCun Y, Szlam A, Vandergheynst P, 2017 *Geometric deep learning: going beyond euclidean data. IEEE Signal Processing Magazine* 34(4):18–42.

Bureau of Transportation Statistics, 2015 *Airline On-Time Statistics and Delay Causes.* Http://www.transtats.bts.gov/.

Bureau of Transportation Statistics, 2018 *Delay Cause by Year, as a Percent of Total Delay Minutes*. URL https://www.bts.gov/delay-cause-year-percent-total-delay-minutes.

Bureau of Transportation Statistics, 2019 *Air Carrier Financial Reports (Form 41 Financial Data)*. URL https://www.transtats.bts.gov/Tables.asp?DB_ID=135.

Buxi G, Hansen M, 2013 *Generating day-of-operation probabilistic capacity scenarios from weather forecasts. Transportation Research Part C: Emerging Technologies* 33:153 – 166.

Crovella M, Kolaczyk E, 2003 *Graph wavelets for spatial traffic analysis. IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 3, 1848–1857 vol.3.

Deb AB, Dey L, 2017 *Outlier detection and removal algorithm in k-means and hierarchical clustering. World Journal of Computer Application and Technology* 5(2):24–29.

Drayer E, Routtenberg T, 2018 *Detection of false data injection attacks in power systems with graph fourier transform. 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 890–894.

Drew M, Sheth K, 2014 *A Frequency Analysis Approach for Categorizing Air Traffic Behavior*, 1–11 (14th AIAA ATIO Conference).

Drew M, Sheth K, 2015 *A Wavelet Analysis Approach for Categorizing Air Traffic Behavior*, 1–14 (AIAA ATIO Conference).

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

Egilmez HE, Ortega A, 2014 *Spectral anomaly detection using graph-based filtering for wireless sensor networks. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1085–1089.

Federal Aviation Administration, 2017 *The Economic Impact of Civil Aviation on the U.S. Economy.* URL https://www.faa.gov/about/plans_reports/media/2017-economic-impact-report.pdf.

Federal Aviation Administration (FAA), accessed 2018 *Aviation System Performance Metrics (ASPM) website.* Http://aspm.faa.gov/.

Filzmoser P, 2004 *A multivariate outlier detection method. State University*, 18–22.

Gan G, Ng MKP, 2017 *K-means clustering with outlier removal. Pattern Recognition Letters* 90:8–14.

Gopalakrishnan K, Balakrishnan H, Jordan R, 2016a *Clusters and Communities in Air Traffic Delay Networks. American Control Conference.*

Gopalakrishnan K, Balakrishnan H, Jordan R, 2016b *Deconstructing Delay Dynamics: An air traffic network example. International Conference on Research in Air Transportation (ICRAT).*

Gopalakrishnan K, Balakrishnan H, Jordan R, 2016c *Stability of networked systems with switching topologies. Decision and Control (CDC), 2016 IEEE 55th Conference on*, 2601–2608 (IEEE).

Gopalakrishnan K, Li MZ, Balakrishnan H, 2019 *Identification of outliers in graph signals. 2019 IEEE 58th Conference on Decision and Control (CDC)*, 4769–4776 (IEEE).

Gorripaty S, Liu Y, Hansen M, Pozdnukhov A, 2017 *Identifying similar days for air traffic management. Journal of Air Transport Management* 65:144–155.

Grabbe SR, Sridhar B, Mukherjee A, 2013 *Similar days in the nas: an airport perspective. 2013 Aviation Technology, Integration, and Operations Conference*, 4222.

Hadi AS, 1992 *Identifying multiple outliers in multivariate data. Journal of the Royal Statistical Society: Series B (Methodological)* 54(3):761–771.

Isufi E, Mahabir AS, Leus G, 2018 *Blind graph topology change detection. IEEE Signal Processing Letters* 25(5):655–659.

Kan R, 2008 *From moments of sum to moments of product. Journal of Multivariate Analysis* 99(3):542–554.

Kieu T, Yang B, Jensen CS, 2018 *Outlier detection for multidimensional time series using deep neural networks. 2018 19th IEEE International Conference on Mobile Data Management (MDM)*, 125–134 (IEEE).

Kim YJ, Choi S, Briceno S, Mavris D, 2016 *A deep learning approach to flight delay prediction. 2016 IEEE/AIAA DASC.*

Kuhn KD, 2016 *A methodology for identifying similar days in air traffic flow management initiative planning. Transportation Research Part C: Emerging Technologies* 69:1–15.

Lam P, Wang L, Ngan HY, Yung NH, Yeh AG, 2017 *Outlier detection in large-scale traffic data by naïve bayes method and gaussian mixture model method. Electronic Imaging* 2017(9):73–78.

Li L, Das S, John Hansman R, Palacios R, Srivastava AN, 2015 *Analysis of flight data using clustering techniques for detecting abnormal operations. Journal of Aerospace Information Systems* 12(9):587–598.

Li L, Hansman RJ, Palacios R, Welsch R, 2016 *Anomaly detection via a gaussian mixture model for flight operation and safety monitoring. Transportation Research Part C: Emerging Technologies* 64:45 – 57.

Li MZ, Gopalakrishnan K, Pantoja K, Balakrishnan H, 2019 *A spectral approach towards analyzing airport performance and disruptions. 13th Air Traffic Management Research and Development Seminar* .

Liu PCB, Hansen M, Mukherjee A, 2008 *Scenario-based air traffic flow management: From theory to practice. Transportation Research Part B: Methodological* 42(7):685 – 702.

Mohan DM, Asif MT, Mitrovic N, Dauwels J, Jaillet P, 2014 *Wavelets on graphs with application to transportation networks. 17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 1707–1712.

Mukherjee A, Grabbe S, Sridhar B, 2013 *Classification of Days using Weather Impacted Traffic in the National Airspace System. AIAA Aviation Technology, Integration and Operations Conference.*

National Oceanic and Atmospheric Administration, 2018a *Atlantic Hurricane Seasons*. URL https://www.nhc.noaa.gov/data/tcr/.

National Oceanic and Atmospheric Administration, 2018b *Saffir-Simpson Hurricane Wind Scale*. URL https://www.nhc.noaa.gov/aboutsshws.php.

National Oceanic and Atmospheric Administration, 2019 *Radar Data Map*. URL https://gis.ncdc.noaa.gov/maps/ncei/radar.

National Weather Service, 2019 *What is a Nor'Easter?*

Peña D, Prieto FJ, 2001 *Multivariate outlier detection and robust covariance matrix estimation. Technometrics* 43(3):286–310.

Pyrgiotis N, Malone KM, Odoni A, 2013 *Modelling delay propagation within an airport network. Transportation Research Part C: Emerging Technologies* 27:60–75.

Rebollo JJ, Balakrishnan H, 2014 *Characterization and prediction of air traffic delays. Transportation Research Part C* 231–241.

Ren KM, Kim AM, Kuhn K, 2018 *Exploration of the evolution of airport ground delay programs. Transportation Research Record* 2672(23):71–81.

Rocke DM, Woodruff DL, 1996 *Identification of outliers in multivariate data. Journal of the American Statistical Association* 91(435):1047–1061.

Sandryhaila A, Moura JM, 2013 *Discrete signal processing on graphs. IEEE transactions on signal processing* 61(7):1644–1656.

Sandryhaila A, Moura JMF, 2014a *Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure. IEEE Signal Processing Magazine* 31(5):80–90.

Sandryhaila A, Moura JMF, 2014b *Discrete signal processing on graphs: Frequency analysis. IEEE Transactions on Signal Processing* 62(12):3042–3054.

Sarkis J, Talluri S, 2004 *Performance based clustering for benchmarking of us airports. Transportation Research Part A: Policy and Practice* 38(5):329–346.

Seah CE, Aligawesa A, Hwang I, 2010 *Algorithm for conformance monitoring in air traffic control. Journal of Guidance, Control, and Dynamics* 33(2):500–509.

Shekhar S, Lu CT, Zhang P, 2002 *Detecting graph-based spatial outliers. Intelligent Data Analysis* 6(5):451–468.

Shuman DI, Narang SK, Frossard P, Ortega A, Vandergheynst P, 2013 *The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Processing Magazine* 30(3):83–98.

Shuman DI, Ricaud B, Vandergheynst P, 2016 *Vertex-frequency analysis on graphs. Applied and Computational Harmonic Analysis* 40(2):260 – 291.

Squires MF, Lawrimore JH, Heim RR, Robinson DA, Gerbush MR, Estilow TW, 2014 *The regional snowfall index. Bulletin of the American Meteorological Society* 95(12):1835–1848.

Sun M, Isufi E, de Groot NMS, Hendriks RC, 2019 *A graph signal processing framework for atrial activity extraction. 2019 27th European Signal Processing Conference (EUSIPCO)*, 1–5.

US Department of Transportation, 2018 *Bureau of Transportation Statistics.*

Weinkle J, Landsea C, Collins D, Musulin R, Crompton RP, Klotzbach PJ, Pielke R, 2018 *Normalized hurricane damage in the continental united states 1900-2017. Nature Sustainability* 1(12):808–813.

Yanofsky D, 2015 *Tech glitches keep plaguing US airlines. This dashboard kept track of them all.* URL https://qz.com/535967/tech-glitches-keep-plaguing-us-airlines-this-dashboard-keeps-track-of-them-all/.

Appendix.

We organize the appendix as follows: we present the bounds for outliers in scale (Section A), followed by the bounds for weak outliers in distribution (Section B). Finally, we present a discussion on evaluating bounds for weak outliers in distribution where only partial information about the correlation coefficients (i.e., the weights associated with the edges of the graphs) is known, and derive analytical bounds for outliers in such a setting (Section C). These theoretical analyses and analytical expressions offer two critical advantages over simulations: Firstly, they allow for a parametric study (e.g. with respect to the mean and covariance parameters) to understand how outlier bounds behave with respect to the underlying probability distribution. Secondly, there may be an insufficient number of data observations in $\mathcal{O}_M$ to reliably estimate the bounds.

A. Bounds for outliers in scale

Our simplification of the 1-norm along with the consideration of only non-negative signals gives $\|\mathbf{X}\| = \sum_{i=1}^{N} X_i$. The expectation and variance of $\|\mathbf{X}\|$ are:

$$\mathbb{E}\left[\|\mathbf{X}\|\right] = \mathbb{E}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathbb{E}\left[X_i\right] = \sum_{i=1}^{N} \mu_i, \tag{7}$$

$$\mathrm{Var}\left[\|\mathbf{X}\|\right] = \mathrm{Var}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathrm{Var}\left[X_i\right] + \sum_{i\neq j} \rho_{ij}\sigma_i\sigma_j, \tag{8}$$

where σ_i and σ_j are the standard deviations of X_i and X_j, respectively. We can also write $\mathrm{Var}\left[\|\mathbf{X}\|\right]$ in terms of the covariance matrix Σ as $\mathrm{Var}\left[\|\mathbf{X}\|\right] = \mathbf{1}^\intercal \Sigma \mathbf{1}$.

B. Bounds for weak outliers in distribution

Since the TV is a random variable that is a function of $\mathbf{X}$, we can write the expression for its mean explicitly as:

$$\mathbb{E}[\mathrm{TV}(\mathbf{X})] = \mathbb{E}\left[\frac{1}{2}\sum_{i\neq j}\left\{\rho_{ij}(X_i - X_j)^2\right\}\right] = \frac{1}{2}\sum_{i\neq j}\left\{\rho_{ij}\left(\mathbb{E}\left[X_i^2\right] + \mathbb{E}\left[X_j^2\right] - 2\,\mathbb{E}\left[X_iX_j\right]\right)\right\}. \tag{9}$$

Using $\mathbb{E}\left[X_iX_j\right] = \mu_i\mu_j + \rho_{ij}\sigma_i\sigma_j$ (from (2)) and $\mathbb{E}\left[X_i^2\right] = \mu_i^2 + \sigma_i^2$, (9) simplifies to

$$\mathbb{E}[\mathrm{TV}(\mathbf{X})] = \frac{1}{2}\sum_{i\neq j}\left\{\rho_{ij}\left[(\mu_i - \mu_j)^2 + (\sigma_i^2 + \sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j)\right]\right\}. \tag{10}$$

We can examine a few special cases for the parameters in (10):

1. If the signals are not correlated, i.e., $\rho_{ij} = 0, \forall i, j \in V$, then the TV is zero.

2. If the signals are perfectly correlated, i.e., $\rho_{ij} = 1, \forall i, j \in V$, the expectation of the TV is determined by possible differences in the mean and variance of graph signals at adjacent nodes. Specifically, we have that $\mathbb{E}[\mathrm{TV}(\mathbf{x})] = \frac{1}{2}\sum_{i\neq j}\{(\mu_i - \mu_j)^2 + (\sigma_i - \sigma_j)^2\}$.

3. If the mean for all nodal signals are identical, i.e., $\mu_i = \mu_j, \forall i, j \in V$, the expectation of the TV is quadratic in the differences of the variances. Specifically, we have that $\mathbb{E}[\mathrm{TV}(\mathbf{X})] = \frac{1}{2}\sum_{i\neq j}\left\{\rho_{ij}\left[\sigma_i^2 + \sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j\right]\right\}$.

4. If the mean, variance, and pairwise correlation coefficient for all nodal signals are identical, i.e., $\mu_i = \mu_j$, $\sigma_i = \sigma_j = \sigma$, and $\rho_{ij} = \rho, \forall i, j \in V$, then the expectation of the TV is quadratic in the number of nodes N, the correlation coefficient ρ, and the variance σ. Specifically, we have that

$$\mathbb{E}[\mathrm{TV}(\mathbf{X})] = \sum_{i\neq j}\left\{\rho\sigma^2(1-\rho)\right\} = N(N-1)\rho\sigma^2(1-\rho). \tag{11}$$

We now outline the process for computing the variance of the TV for a random graph signal vector $\mathbf{X}$. Although we explicitly derive an analytical expression for $\mathrm{Var}\left[\mathrm{TV}(\mathbf{X})\right]$, we only symbolically evaluate it when needed since the number of terms is extremely large. We rewrite $\mathrm{Var}\left[\mathrm{TV}(\mathbf{X})\right]$ as:

$$\mathrm{Var}\left[\mathrm{TV}(\mathbf{X})\right] = \mathbb{E}\left[\mathrm{TV}(\mathbf{X})^2\right] - \mathbb{E}\left[\mathrm{TV}(\mathbf{X})\right]^2. \tag{12}$$

We have already essentially derived the expression for the second term, since from (10) we have that $\mathbb{E}\left[\mathrm{TV}(\mathbf{X})\right]^2 = \frac{1}{4}\left(\sum_{i\neq j}\left\{\rho_{ij}\left[(\mu_i-\mu_j)^2+(\sigma_i^2+\sigma_j^2-2\rho_{ij}\sigma_i\sigma_j)\right]\right\}\right)^2$. The first term representing the expectation of the square of the TV can be expanded and rewritten as:

$$\mathbb{E}\left[\mathrm{TV}(\mathbf{X})^2\right] = \frac{1}{4}\mathbb{E}\left[\left(\sum_{i\neq j}\left\{\rho_{ij}(X_i-X_j)^2\right\}\right)^2\right]. \tag{13}$$

Expanding (13) further will produce terms that depend on products of correlated random variables. Specifically, the expansion will introduce terms of the form $\mathbb{E}\left[X_i^4\right]$, $\mathbb{E}\left[X_i^3X_j\right]$, $\mathbb{E}\left[X_i^2X_j^2\right]$, $\mathbb{E}\left[X_i^3X_jX_l\right]$ and $\mathbb{E}\left[X_iX_jX_lX_m\right]$ for nodes $i,j,l,m\in V$. If $\mathbf{X}$ is a multivariate Gaussian random variable $\mathbf{X}=(X_1,\ldots,X_N)^\intercal\in\mathbb{R}^{N\times 1}$, where $\mathbf{X}\overset{iid}{\sim}\mathcal{N}(\boldsymbol{\mu},\Sigma)$ with mean $\boldsymbol{\mu}=(\mu_1,...,\mu_N)^\intercal\in\mathbb{R}^{N\times 1}$ and covariance $\Sigma\in\mathbb{R}^{N\times N}$, $\Sigma\succeq 0$, then (13) can be analytically evaluated through Proposition 2:

Proposition 2 (Isserlis (1918) and Kan (2008)) *Suppose $\mathbf{X}=(X_1,\ldots,X_N)^\intercal\sim\mathcal{N}(\boldsymbol{\mu},\Sigma)$, where Σ is an $N\times N$ positive semi-definite matrix. For non-negative integers s_1 to s_N, we have*

$$\mathbb{E}\left[\prod_{i=1}^{N}X_i^{s_i}\right] = \sum_{\nu_1=0}^{s_1}\cdots\sum_{\nu_N=0}^{s_N}\sum_{r=0}^{[s/2]}\binom{s_1}{\nu_1}\cdots\binom{s_N}{\nu_N}\times\left\{\frac{\left(\frac{\boldsymbol{h}^\intercal\Sigma\boldsymbol{h}}{2}\right)^r\left(\boldsymbol{h}^\intercal\boldsymbol{\mu}\right)^{s-2r}}{r!(s-2r)!}\right\}, \tag{14}$$

where $s=s_1+\cdots+s_N$ and $\boldsymbol{h}=\left(\frac{s_1}{2}-\nu_1,...,\frac{s_N}{2}-\nu_N\right)^\intercal$.

Proof of Proposition 2. See Kan (2008). □

As we have previously mentioned, due to the large number of terms present in the full expansion of (13), we do not attempt to analytically simplify it further, but emphasize that it can be symbolically evaluated using Proposition 2. On the other hand, for our subsequent analyses, we can numerically evaluate $\mathrm{Var}\left[\mathrm{TV}(\mathbf{X})\right]$ precisely using the analytical expression in (14).

Since we do not have a simplified analytical form for the variance of the TV, it is challenging to make qualitative comments on how the variance of the TV changes with parameters like Σ and $\boldsymbol{\mu}$. However, one sufficient condition for the TV to be equal to zero is as follows:

Proposition 3 *If $\mathbb{E}[TV(\mathbf{X})]=0$ and $\rho_{ij}\geq 0$, $\forall i,j$ (or $\rho_{ij}\leq 0$, $\forall i,j$), then $Var[TV(\mathbf{X})]=0$.*

Proof of Proposition 3. Since we have that $\rho_{ij}\geq 0$, we have that $\mathrm{TV}(\mathbf{X})\geq 0$ (or if all $\rho_{ij}\leq 0$, then $\mathrm{TV}(\mathbf{X})\leq 0$). Then, $\mathbb{E}[\mathrm{TV}(\mathbf{X})]=0 \implies \mathrm{TV}(\mathbf{X})=0$. Hence, $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]=0$. □

Note that $\mathbb{E}[\mathrm{TV}(\mathbf{X})]=0$ is not a necessary condition for the variance of the TV to be 0. Consider the following example: suppose $\rho_{ij}=1$, $\mu_i\neq\mu_j$, and $\sigma_i^2=\sigma_j^2,\forall i,j\in V$. Then, we have that in general $\mathbb{E}[\mathrm{TV}(\mathbf{X})]=\frac{1}{2}\sum_{i\neq j}(\mu_i-\mu_j)^2\neq 0$. However, since we have perfect correlation with differing means, X_i-X_j will always be constant $\forall i,j\in V$, and thus $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]=0$.

C. Bounds for weak outliers in distribution: Partial information case

In Section B, we analyzed the expectation and variance of the TV of a graph signal $\mathbf{x}$ assuming that we had perfect information regarding the strength of the nodal signal correlations ρ_{ij}. However, in reality it is possible that we do not know the exact value of ρ_{ij}, but we do know bounds ν_{ij} and ϵ_{ij} such that $0 \leq \nu_{ij} < \rho_{ij} < \epsilon_{ij} \leq 1$, for all nodes $i, j \in V$. This *partial information case* regarding correlations can happen in a variety of scenarios; we will describe a few such scenarios. First, due to privacy concerns, nodes in many physical systems act as independent agents that withhold information from other agents. Thus, each node may only report the mean and variance of its own signal. In this case, the inter-dependencies and correlations can only be partially estimated. Another example of the partial information case occurs when we do not know the underlying Gaussian distribution of the signal, implying that ρ_{ij} is an unknown parameter. Similar to the first example, statistical testing may only provide confidence intervals or bounds on pairwise correlations. Lastly, a small data set, i.e. small M for $\mathcal{O}_M$, would result in a gap between the sample correlation and the true correlation. This gap may be significant depending on M. In this scenario, it may be preferable to use the bounds on ρ_{ij} rather than an incorrect estimate to identify outliers.

For the rest of this derivation, we assume that the observations are drawn from a multivariate Gaussian distribution with a fixed mean vector $\boldsymbol{\mu} \in \mathbb{R}^{N\times 1}$ and covariance matrix $\Sigma \in \mathbb{S}_{\succeq 0}^{N\times N}$ (or equivalently the correlation matrix $\mathbf{C}$), but the precise value of ρ_{ij} is unknown. Due to the uncertainty in ρ_{ij}, we can only provide bounds on the values of $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$ and $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$. One could propose that given bounds $\rho_{ij} \in (\nu_{ij}, \epsilon_{ij}) \subseteq [0,1]$, we could use simulation to estimate $\widehat{\mathbb{E}}[\mathrm{TV}(\mathbf{X})]$ and $\widehat{\mathrm{Var}}[\mathrm{TV}(\mathbf{X})]$. However, we note that such an approach is computationally intractable in general for the two reasons: first, the number of intervals over which we need to simulate and evaluate the TV is exponentially large. Specifically, discretizing $\rho_{ij} \in (\nu_{ij}, \epsilon_{ij}) \subseteq [0,1]$ into N_ρ intervals for each edge leads to $N_\rho^{N\times(N-1)}$ evaluations of $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$ and $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$. A counterpoint may be that a more coarse discretization scheme might suffice, or a gradient-based optimization may be able to guide the exploration of this complex space, or considering just the extreme values of the bounds may suffice. This brings us to our second point: the non-monotonic behavior of $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$ and $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$ as a function of $\rho_{ij} \in (\nu_{ij}, \epsilon_{ij}) \subseteq [0,1]$. We provide a small-scale example in Gopalakrishnan, Li, and Balakrishnan (2019) that highlights the non-monotonicity in ρ_{ij} of $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$ and $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$. This behavior is apparent even in a relatively simple graph with 5 nodes.

Given the various difficulties with evaluating $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$ and $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$ in the case of partial information regarding ρ_{ij}, the tight analytical bounds we present in Propositions 4 and 5 for $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$ and $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$, respectively, offer an alternative to the computationally prohibitive exploration of the search space $\rho_{ij} \in (\nu_{ij}, \epsilon_{ij}) \subseteq [0,1]$ with no reliance on estimation intervals and discretizations. Specifically, our two propositions quantify the change in $\mathbb{E}[\mathrm{TV}(\mathbf{X})]$ (Proposition 4) and $\mathrm{Var}[\mathrm{TV}(\mathbf{X})]$ (Proposition 5) due to the uncertainty in ρ_{ij}.

For the propositions we construct in this section, we need all the correlation coefficients to have the same sign, *i.e.* all $\rho_{ij} \geq 0$ or all $\rho_{ij} \leq 0$, $\forall i, j \in V$. However, this is not a restrictive assumption; see the discussion in Gopalakrishnan, Li, and Balakrishnan (2019). We can now redefine the TV for an *unobserved* $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$

with respect to the Laplacian $\mathfrak{L} \in \mathbb{S}^{N\times N}$ constructed using $\mathcal{O}_M$. Observe that $\mathrm{TV}(\mathbf{X})$ is a derived random variable,

$$\mathrm{TV}(\mathbf{X}) = \frac{1}{2}\sum_{i\neq j}\left\{r^{+}_{ij|\mathcal{O}_M}(X_i - X_j)^2\right\}. \tag{15}$$

However, in the case where the correlations are only known within some interval, i.e $0 \le \nu_{ij} < \rho_{ij} < \epsilon_{ij} \le 1$, we have that:

$$\sum_{i=1}^{N}\sigma_i^2 + \sum_{i\neq j}\nu_{ij}\sigma_i\sigma_j < \mathrm{Var}\left[\|\mathbf{X}\|\right] = \mathrm{Var}\left[\sum_{i=1}^{N}X_i\right] < \sum_{i=1}^{N}\sigma_i^2 + \sum_{i\neq j}\epsilon_{ij}\sigma_i\sigma_j. \tag{16}$$

With these analytical expressions for $\mathbb{E}\left[\|\mathbf{X}\|\right]$ and $\mathrm{Var}\left[\|\mathbf{X}\|\right]$, or analytical bounds for $\mathrm{Var}\left[\|\mathbf{X}\|\right]$ in the case of partially-known correlations, we can substitute $\mathbb{E}\left[\|\mathbf{X}\|\right] = \sum_{i=1}^{N}\mu_i$ along with either $\mathrm{Var}\left[\|\mathbf{X}\|\right] = \mathbf{1}^{\intercal}\Sigma\mathbf{1} + \sum_{i\neq j}\rho_{ij}\sigma_i\sigma_j$ or $\mathrm{Var}\left[\|\mathbf{X}\|\right] \in \left(\sum_{i=1}^{N}\sigma_i^2 + \sum_{i\neq j}\nu_{ij}\sigma_i\sigma_j, \sum_{i=1}^{N}\sigma_i^2 + \sum_{i\neq j}\epsilon_{ij}\sigma_i\sigma_j\right)$ into Definition 6, transforming this definition to one that can be used to detect outliers in scale.

Proposition 4 *Suppose that $0 \le \nu_{ij} < \rho^{+}_{ij} < \varepsilon_{ij} \le 1$ for all unique pairs of nodes $i, j \in V$. Then, we can evaluate scalars δ_1 and δ_2, with $\delta_2 \ge 0$, such that* $\max\{0, \delta_1\} \le \mathbb{E}[TV(\mathbf{X})] < \delta_2$.

Proof of Proposition 4. See Gopalakrishnan, Li, and Balakrishnan (2019). □

Proposition 5 *Suppose $0 \le \nu_{ij} < \rho^{+}_{ij} < \varepsilon_{ij} \le 1$ for all unique pairs of nodes $i, j \in V$. Then, we can evaluate scalars δ_3 and δ_4, with $\delta_4 \ge 0$, such that* $\max\{0, \delta_3\} \le Var[TV(\mathbf{X})] < \delta_4$.

Proof of Proposition 5. See Gopalakrishnan, Li, and Balakrishnan (2019). □

Using these two propositions, we can modify Definition 5 for weak outliers in distribution of level k to include these more conservative bounds:

Definition 9 (Partial information case of Definition 5) *An observation* $\mathbf{x}$ *containing bounded partial information regarding all pairwise correlations, i.e. $\rho_{ij} \in (\nu_{ij}, \epsilon_{ij}) \subseteq [0,1], \forall i, j \in V$, is considered a* weak distribution outlier of level k *or a* weak outlier in distribution of level k *if*

$$TV(\mathbf{x}) \notin \left[\max\left\{0, \delta_1 - k\sqrt{\delta_4}\right\}, \delta_2 + k\sqrt{\delta_4}\right], \text{ for some } k \ge 0. \tag{17}$$

where δ_1, δ_2, δ_3, and δ_4 are as defined in Propositions 4 and 5.

The modified Definition 9 of weak outliers in distribution also shows how such bounds can be implemented in practice to detect weak outliers in distribution. We make some final remarks related to a well-known spectral bound (Rayleigh quotient) as well as the generalizability of our bounds to other underlying distributions. Denote $\lambda_{\max}$ as the largest eigenvalue of $\mathfrak{L}$. Then:

$$\mathrm{TV}(\mathbf{x}) = \mathbf{x}^{\intercal}\mathfrak{L}\mathbf{x} \le \lambda_{\max}\|\mathbf{x}\|_2^2 \le \lambda_{\max}\|\mathbf{x}\|_1^2. \tag{18}$$

While the Rayleigh quotient is indeed a valid upper bound for the TV of all data observations in $\mathcal{O}_M$, it is loose and does not provide further refinements on the various bounds we propose. Finally, these bounds only require the underlying distribution to have a finite expectation and variance; there is no explicit dependence on the underlying distribution being Gaussian.

D. Airline-specific correlation maps

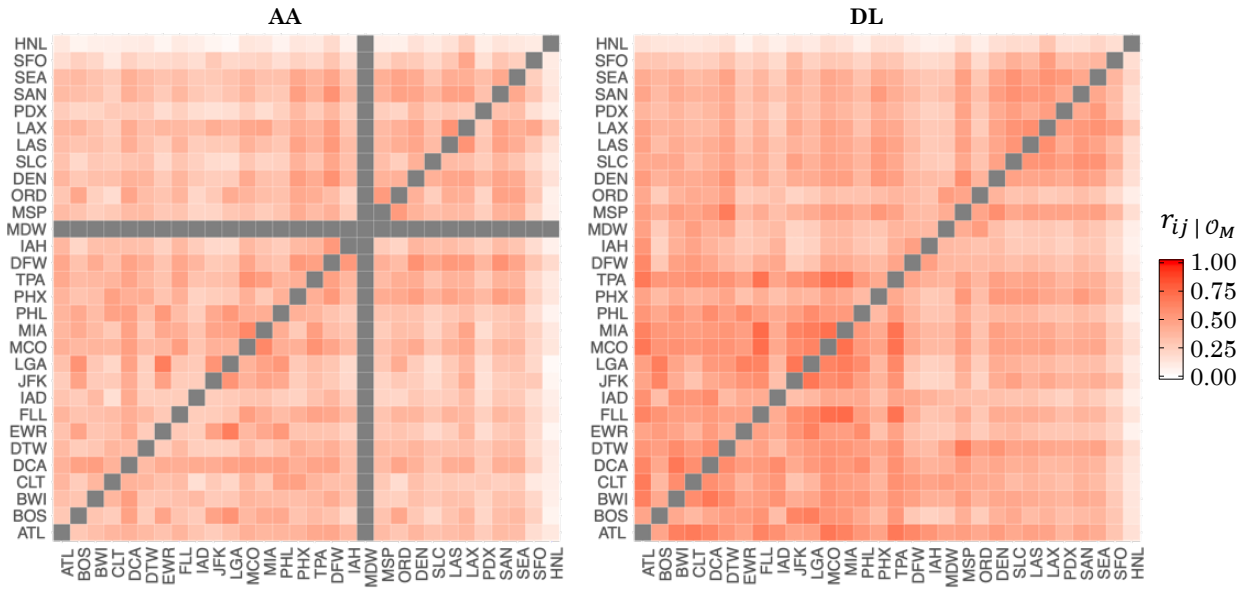

Figure 11 **Heat maps of the delay correlations between the top 30 airports for AA and DL.**

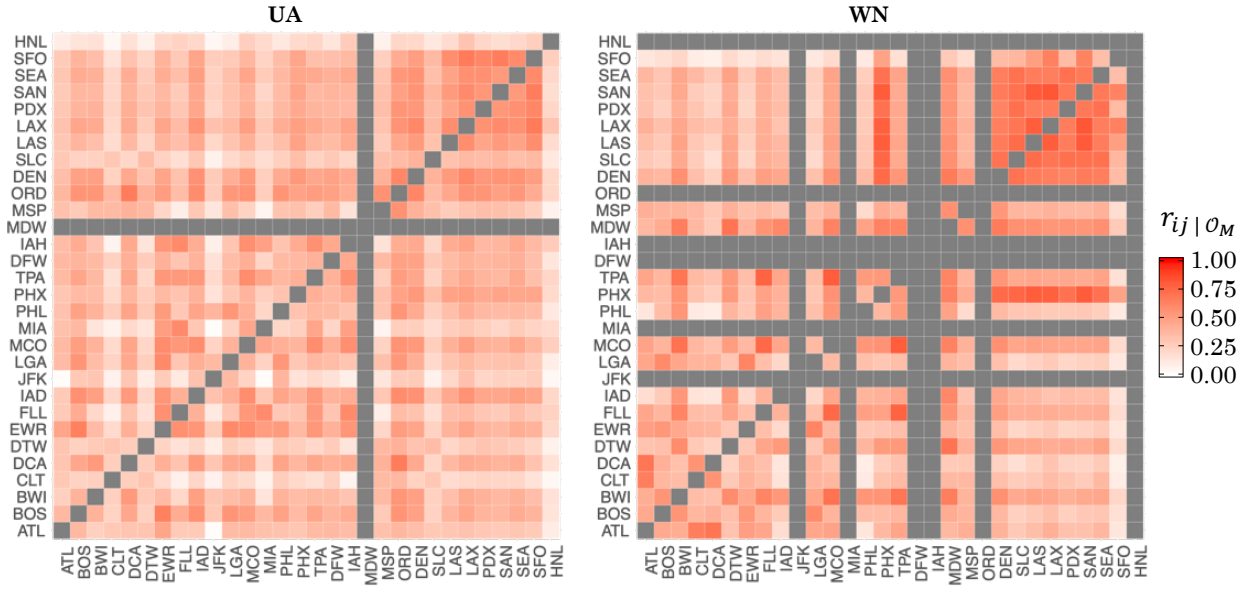

Figure 12 **Heat maps of the delay correlations between the top 30 airports for UA and WN.**

E. Outlier bounds for each airline

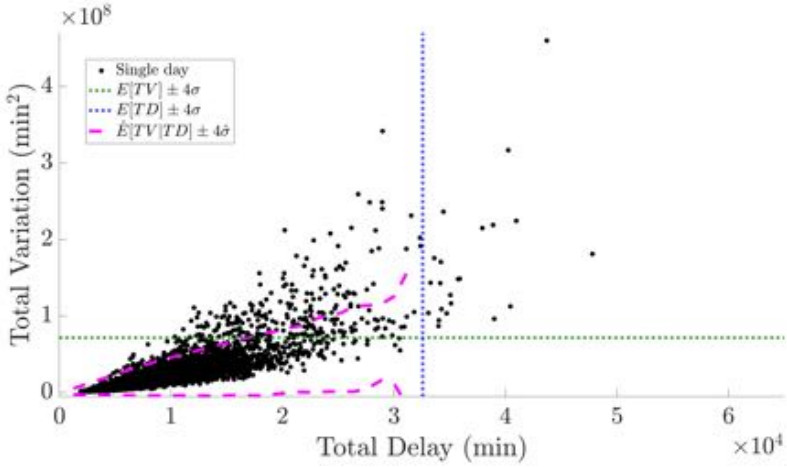

Figure 13 **AA sub-network: TV versus TD for all days in 2008-2017 with level $k = 4$ weak and strong outlier bounds demarcated.**

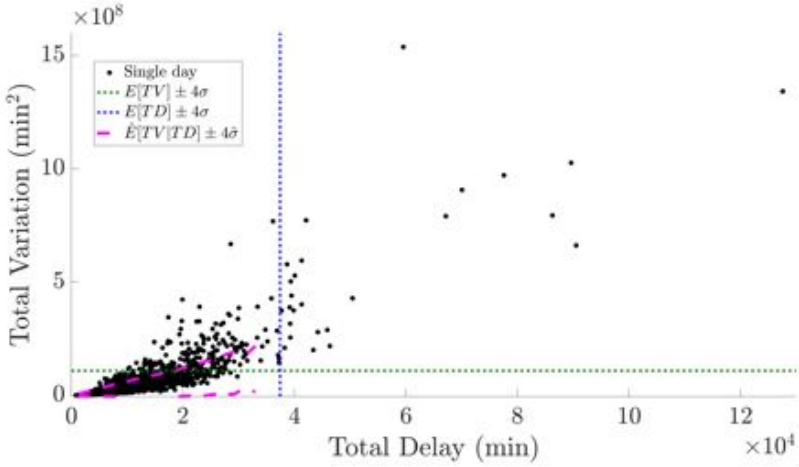

Figure 14 **DL sub-network: TV versus TD for all days in 2008-2017 with level $k = 4$ weak and strong outlier bounds demarcated.**

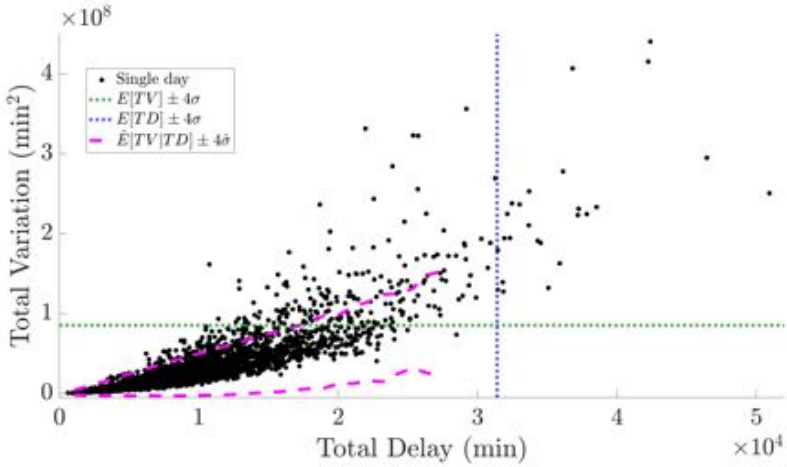

Figure 15 **UA sub-network: TV versus TD for all days in 2008-2017 with level $k = 4$ weak and strong outlier bounds demarcated.**

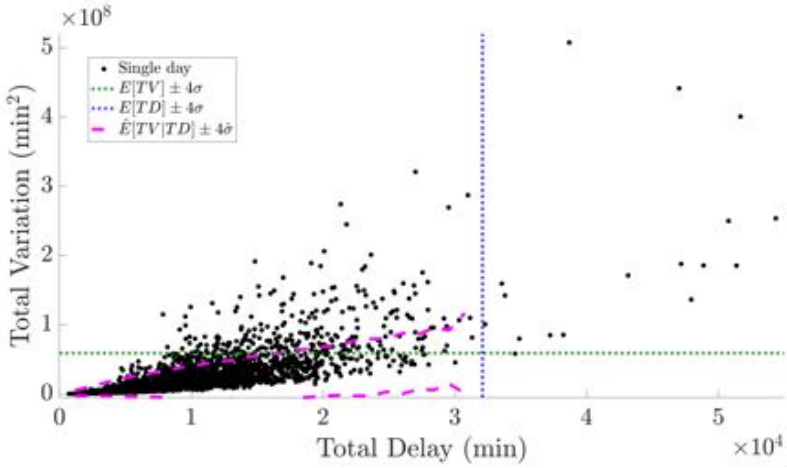

Figure 16 **WN sub-network: TV versus TD for all days in 2008-2017 with level $k = 4$ weak and strong outlier bounds demarcated.**

F. Airline-specific TV versus TD plots of disruptions

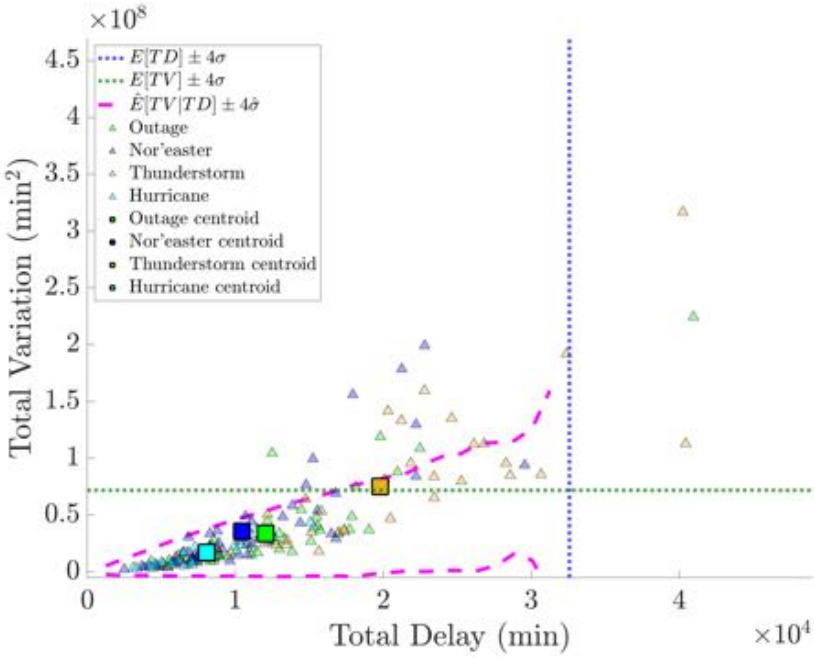

Figure 17 **TV versus total delay plot for American Airlines (AA) during 2008-2017 with specific disruptions and their average values annotated.**

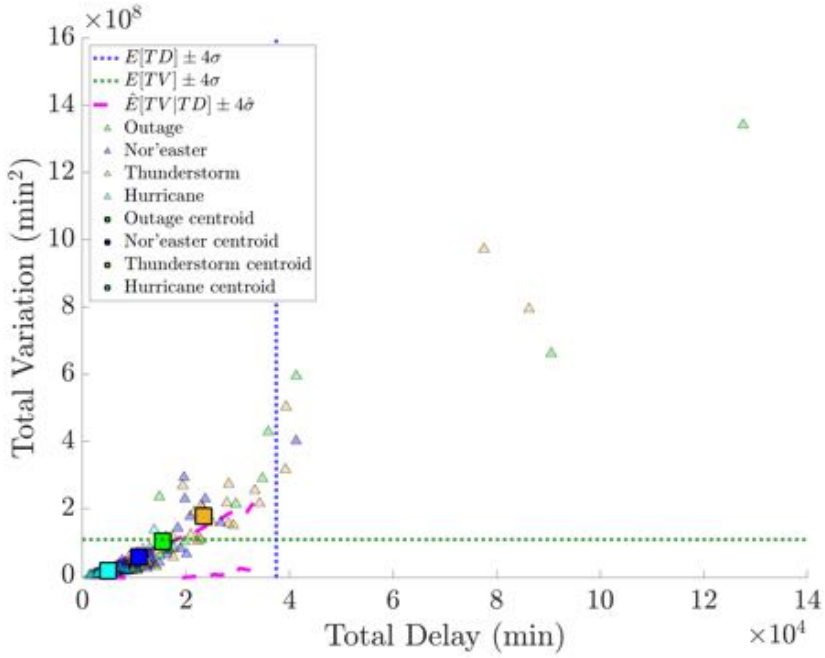

Figure 18 **TV versus total delay plot for Delta Air Lines (DL) during 2008-2017 with specific disruptions and their average values annotated.**

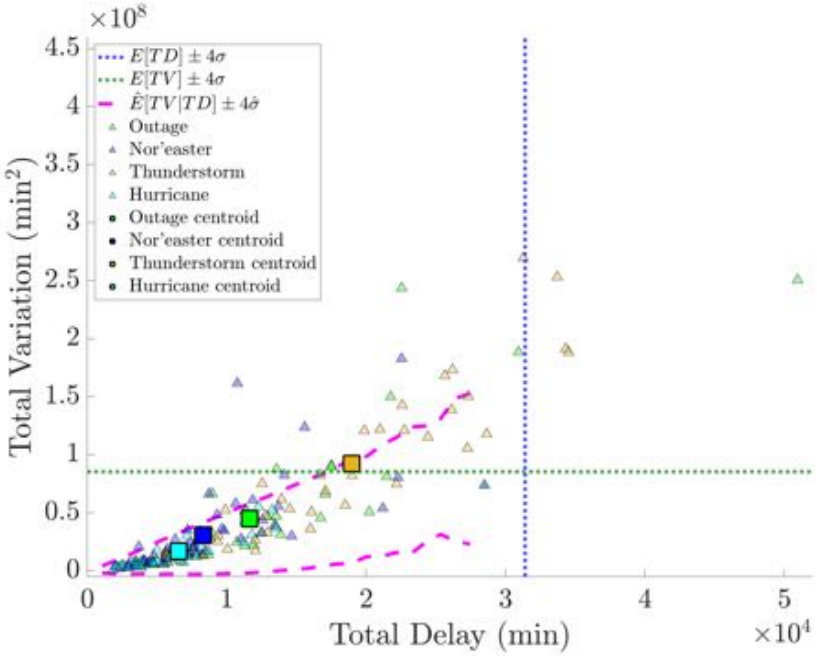

Figure 19 **TV versus total delay plot for United Airlines (UA) during 2008-2017 with specific disruptions and their average values annotated.**

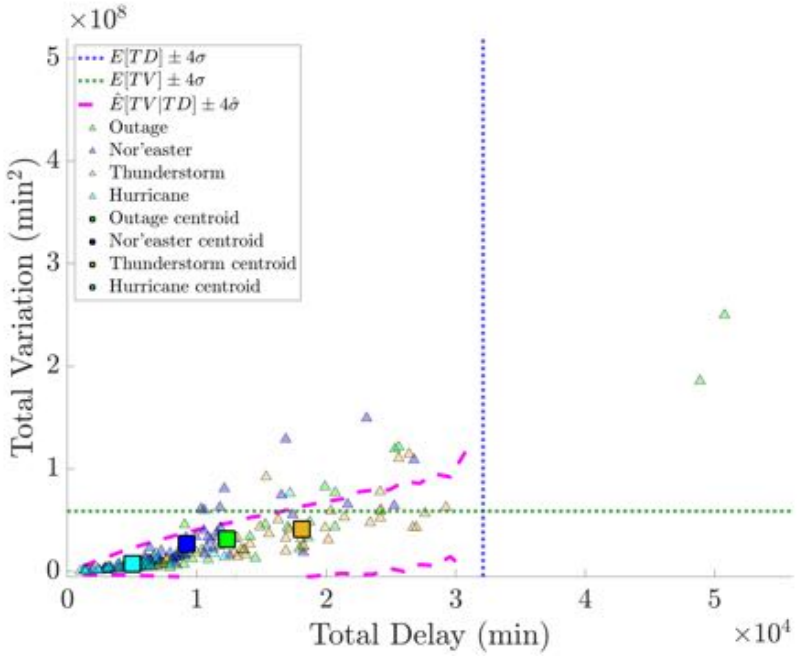

Figure 20 **TV versus total delay plot for Southwest Airlines (WN) during 2008-2017 with specific disruptions and their average values annotated.**

G. Supplementary tables

IATA Code	Airport Name	ARTCC (Center)
ATL	Hartsfield-Jackson Atlanta International Airport	ZTL (Atlanta)
BOS	Boston General Edward Lawrence Logan International Airport	ZBW (Boston)
BWI	Baltimore/Washington International Thurgood Marshall Airport	ZDC (Washington)
CLT	Charlotte Douglas International Airport	ZTL (Atlanta)
DCA	Ronald Reagan Washington National Airport	ZDC (Washington)
DEN	Denver International Airport	ZDV (Denver)
DFW	Dallas/Fort Worth International Airport	ZFW (Fort Worth)
DTW	Detroit Metropolitan Wayne County Airport	ZOB (Cleveland)
EWR	Newark Liberty International Airport	ZNY (New York)
FLL	Fort Lauderdale-Hollywood International Airport	ZMA (Miami)
HNL	Honolulu Daniel K. Inouye International Airport	ZHN (Honolulu)
IAD	Washington Dulles International Airport	ZDC (Washington)
IAH	Houston George Bush International Airport	ZHU (Houston)
JFK	New York John F. Kennedy International Airport	ZNY (New York)
LAS	Las Vegas McCarran International Airport	ZLA (Los Angeles)
LAX	Los Angeles International Airport	ZLA (Los Angeles)
LGA	New York LaGuardia Airport	ZNY (New York)
MCO	Orlando International Airport	ZJX (Jacksonville)
MDW	Chicago Midway International Airport	ZAU (Chicago)
MIA	Miami International Airport	ZMA (Miami)
MSP	Minneapolis-Saint Paul International Airport	ZMP (Minneapolis)
ORD	Chicago O'Hare International Airport	ZAU (Chicago)
PDX	Portland International Airport	ZSE (Seattle)
PHL	Philadelphia International Airport	ZNY (New York)
PHX	Phoenix Sky Harbor International Airport	ZAB (Albuquerque)
SAN	San Diego International Airport	ZLA (Los Angeles)
SEA	Seattle-Tacoma International Airport	ZSE (Seattle)
SFO	San Francisco International Airport	ZOA (Oakland)
SLC	Salt Lake City International Airport	ZLC (Salt Lake City)
TPA	Tampa International Airport	ZJX (Jacksonville)

Table 7: IATA three-letter code and corresponding full airport name; the ARTCC that each airport is located within is also listed.

Airline	AA, DL, UA Hubs and WN Operating Bases/Focus Cities
American Airlines (AA)	CLT, DCA, **DFW**, JFK*, LAX**, LGA*, MIA, ORD*, PHL, PHX*
Delta Air Lines (DL)	**ATL***, BOS, DTW, JFK*, LAX**, LGA*, MSP, SEA, SLC
United Airlines (UA)	DEN*, EWR, IAD, IAH, LAX**, **ORD***, SFO, (GUM)
Southwest Airlines (WN)	ATL*, BWI, DEN*, FLL, LAS, LAX**, MCO, **MDW**, PHX*, SAN, TPA, (AUS), (BNA), (DAL), (HOU), (OAK), (SJC), (SMF), (STL)

Table 8 List of airline hubs, operating bases, and focus cities. Boldface denotes an airline's largest hub, operating base, or focus city by number of departing seats in 2017. (*) and (**) denotes an airport that is shared as a hub, operating base, or focus city between 2 or 3+ airlines, respectively. Airport codes in parenthesis are airline-specific hubs, but not part of the Core 30, and thus are not included in our analysis nor defined in Table 7.

Event	Date	1st	2nd	3rd	4th	5th	Total Delay ($\times 10^4 min$)	Total Variation ($\times 10^6 min^2$)
Hurricane	10/28/12	1 (83%)	–	–	–	–	1.51	13.58
	10/29/12	1 (77%)	2 (3%)	–	–	–	0.89	7.02
	10/30/12	1 (77%)	16 (4%)	–	–	–	1.11	12.39
	10/31/12	1 (86%)	–	–	–	–	1.24	7.90
	11/1/12	1 (87%)	–	–	–	–	1.45	10.40
Hurricane	8/24/17	1 (87%)	–	–	–	–	1.63	10.79
	8/25/17	1 (88%)	–	–	–	–	1.63	9.32
	8/26/17	1 (58%)	4 (18%)	3 (12%)	–	–	1.39	28.29
	8/27/17	1 (74%)	3 (9%)	–	–	–	1.59	18.86
	8/28/17	1 (78%)	14 (6%)	–	–	–	1.65	23.57
	8/29/17	1 (91%)	–	–	–	–	1.50	7.74
	8/30/17	1 (91%)	–	–	–	–	1.26	4.43
Hurricane	9/9/17	1 (85%)	–	–	–	–	1.06	6.18
	9/10/17	1 (77%)	19 (6%)	–	–	–	0.96	8.26
	9/11/17	1 (47%)	10 (17%)	11 (15%)	8 (7%)	–	1.44	62.74
	9/12/17	1 (89%)	–	–	–	–	1.39	7.78
NAS-wide	1/2/14	1 (68%)	18 (8%)	14 (7%)	–	–	4.33	254.26
	1/3/14	1 (71%)	26 (9%)	–	–	–	4.62	276.30
NAS-wide	1/5/14	1 (78%)	26 (3%)	–	–	–	4.58	171.69
	1/6/14	1 (79%)	14 (3%)	–	–	–	3.87	111.53
NAS-wide	6/17/15	1 (67%)	4 (16%)	–	–	–	2.20	47.34
Nor'easter	2/25/10	1 (69%)	6 (5%)	24 (4%)	4 (4%)	–	2.29	71.36
	2/26/10	1 (60%)	26 (9%)	6 (6%)	29 (5%)	11 (5%)	2.89	180.21
	2/27/10	1 (87%)	–	–	–	–	1.69	12.45
Nor'easter	1/30/11	1 (87%)	–	–	–	–	1.14	5.77
	1/31/11	1 (74%)	7 (9%)	–	–	–	1.64	26.59
	2/1/11	1 (71%)	5 (9%)	–	–	–	2.83	77.64
	2/2/11	1 (72%)	29 (4%)	19 (3%)	26 (2%)	–	2.25	61.21
	2/3/11	1 (82%)	–	–	–	–	1.85	19.08
Nor'easter	2/7/13	1 (80%)	–	–	–	–	1.59	18.35
	2/8/13	1 (86%)	–	–	–	–	1.73	14.88
	2/9/13	1 (65%)	26 (12%)	29 (9%)	–	–	1.50	43.41
	2/10/13	1 (75%)	10 (5%)	–	–	–	1.64	25.52
	2/11/13	1 (72%)	6 (5%)	19 (4%)	–	–	2.13	53.19
Nor'easter	2/11/14	1 (89%)	–	–	–	–	1.63	10.34
	2/12/14	1 (76%)	11 (75%)	–	–	–	2.32	45.86
	2/13/14	1 (70%)	6 (5%)	11 (4%)	4 (4%)	–	3.48	151.25
	2/14/14	1 (86%)	–	–	–	–	2.82	37.67
Nor'easter	1/26/15	1 (74%)	4 (4%)	6 (4%)	–	–	1.94	37.12
	1/27/15	1 (83%)	–	–	–	–	1.18	10.19
	1/28/15	1 (84%)	–	–	–	–	1.28	9.25
	1/29/15	1 (89%)	–	–	–	–	1.50	8.96
	1/30/15	1 (79%)	19 (4%)	–	–	–	2.05	36.56
Nor'easter	1/21/16	1 (90%)	–	–	–	–	1.72	8.89
	1/22/16	1 (80%)	–	–	–	–	2.37	38.14
	1/23/16	1 (58%)	23 (24%)	–	–	–	1.61	65.48
	1/24/16	1 (56%)	26 (10%)	29 (7%)	27 (5%)	6 (3%)	1.85	92.61
	1/25/16	1 (68%)	29 (14%)	–	–	–	1.79	53.75
Outage	11/15/12	1 (93%)	–	–	–	–	1.55	5.68
Outage	9/26/14	1 (73%)	14 (12%)				2.33	60.16
Outage	9/17/15	1 (80%)	–	–	–	–	1.73	20.79
Outage	7/20/16	1 (84%)	–	–	–	–	2.05	26.13
	7/21/16	1 (77%)	14 (7%)	–	–	–	2.74	62.78
Outage	8/8/16	1 (70%)	11 (14%)	–	–	–	2.72	88.89
	8/9/16	1 (87%)	–	–	–	–	2.23	21.41
Outage	1/22/17	1 (79%)	10 (5%)	–	–	–	3.09	69.63
Outage	12/17/17	1 (65%)	11 (14%)	10 (13%)	–	–	1.64	40.85

Table 9 Different off-nominal events; columns "1st" through "5th" contain the highest-contributing eigenvectors and their energy contribution, in descending order (table modified and adapted from Li et al. (2019)).

Day-type	Dates							
0,0,0,0,1	1/2/08	12/25/08	8/21/09	7/12/10	4/16/11	12/25/11	11/30/12	8/1/15
	1/3/08	12/27/08	8/22/09	8/5/10	4/26/11	12/26/11	12/17/12	8/13/15
	1/12/08	1/5/09	10/19/09	8/12/10	5/11/11	12/29/11	12/26/12	9/4/15
	1/21/08	1/7/09	10/30/09	9/8/10	7/8/11	12/30/11	1/31/13	10/28/15
	2/1/08	1/28/09	11/19/09	9/27/10	7/11/11	1/8/12	8/3/13	11/19/15
	2/24/08	2/1/09	11/25/09	10/4/10	7/13/11	1/22/12	8/18/13	1/31/16
	2/26/08	2/2/09	12/5/09	10/5/10	7/14/11	1/23/12	10/27/13	3/29/16
	3/7/08	2/6/09	12/7/09	10/17/10	7/15/11	3/14/12	11/21/13	6/30/16
	5/20/08	2/10/09	12/10/09	10/26/10	7/25/11	3/16/12	1/18/14	7/20/16
	6/4/08	2/16/09	12/21/09	11/19/10	8/7/11	5/29/12	2/7/14	7/22/16
	6/10/08	3/22/09	1/22/10	11/22/10	8/15/11	6/1/12	4/30/14	9/29/16
	6/22/08	3/26/09	2/8/10	12/21/10	8/18/11	7/1/12	5/21/14	10/16/16
	7/8/08	4/9/09	3/11/10	1/2/11	8/21/11	8/12/12	8/3/14	1/5/17
	7/10/08	4/14/09	3/30/10	1/8/11	8/25/11	8/22/12	9/27/14	2/1/17
	7/21/08	5/1/09	4/25/10	2/5/11	8/28/11	8/26/12	10/23/14	5/11/17
	9/23/08	5/4/09	5/26/10	2/17/11	9/6/11	8/31/12	11/19/14	9/8/17
	11/13/08	5/26/09	6/11/10	2/18/11	9/11/11	9/8/12	11/25/14	9/18/17
	11/18/08	5/27/09	6/15/10	2/19/11	10/13/11	9/10/12	12/30/14	
	11/22/08	6/3/09	6/18/10	2/25/11	10/19/11	9/18/12	6/2/15	
	12/10/08	8/2/09	6/24/10	3/23/11	11/22/11	11/20/12	7/13/15	
	12/15/08	8/19/09	7/6/10	4/13/11	12/21/11	11/21/12	7/18/15	
0,0,1,0,0	1/16/08	7/30/12	12/17/13	3/21/15	5/10/16	2/12/17	12/9/17	
	1/19/08	8/9/12	1/7/14	3/27/15	5/28/16	3/2/17	12/13/17	
	2/17/08	10/22/12	1/28/14	4/6/15	6/21/16	3/10/17	12/17/17	
	2/27/08	12/24/12	1/30/14	5/26/15	6/24/16	3/24/17	12/23/17	
	7/31/08	1/13/13	3/26/14	6/9/15	7/2/16	3/28/17	12/25/17	
	2/7/09	1/24/13	4/7/14	6/24/15	7/18/16	4/3/17		
	6/30/09	1/30/13	5/27/14	7/21/15	7/26/16	4/7/17		
	11/3/09	2/11/13	6/10/14	9/10/15	8/9/16	4/8/17		
	11/27/09	2/26/13	6/14/14	10/24/15	8/10/16	4/9/17		
	1/16/10	3/13/13	7/27/14	10/29/15	8/11/16	4/15/17		
	2/23/10	3/18/13	9/23/14	10/30/15	8/14/16	4/20/17		
	2/24/10	5/17/13	9/30/14	12/24/15	9/19/16	5/6/17		
	7/13/10	5/26/13	10/14/14	1/6/16	11/15/16	5/21/17		
	7/23/10	6/5/13	10/22/14	1/7/16	11/16/16	5/22/17		
	5/26/11	6/8/13	1/6/15	1/9/16	12/14/16	6/3/17		
	6/15/11	6/13/13	2/3/15	2/4/16	1/8/17	7/11/17		
	11/25/11	6/28/13	2/9/15	3/1/16	1/9/17	7/14/17		
	12/22/11	7/17/13	2/16/15	3/4/16	1/14/17	7/17/17		
	4/9/12	10/18/13	2/20/15	3/10/16	1/21/17	10/12/17		
	6/16/12	11/1/13	3/1/15	3/24/16	1/29/17	10/29/17		
	7/15/12	12/11/13	3/15/15	4/17/16	2/10/17	12/6/17		
	1/22/08	6/12/10	5/22/12	5/12/14	7/31/15			
	3/21/08	9/15/10	5/30/12	6/9/14	10/23/15			
	3/27/08	9/28/10	6/6/12	6/19/14	3/8/16			
	4/8/08	10/23/10	6/21/12	7/14/14	4/7/16			
	4/10/08	1/17/11	8/16/12	8/6/14	4/29/16			
	6/19/08	2/20/11	4/9/13	8/17/14	5/26/16			
	7/2/08	4/8/11	4/16/13	9/5/14	5/31/16			
	10/24/08	4/19/11	5/2/13	10/2/14	6/12/16			
	12/12/08	4/20/11	5/3/13	10/30/14	6/18/16			
	1/14/09	5/1/11	6/17/13	12/16/14	6/23/16			
	5/2/09	5/14/11	8/13/13	12/20/14	7/29/16			
	6/5/09	5/23/11	9/5/13	12/28/14	7/30/16			
	6/6/09	9/16/11	9/20/13	12/31/14	9/14/16			
	6/12/09	9/18/11	10/22/13	1/17/15	10/21/16			
	7/27/09	1/28/12	12/21/13	2/26/15	1/15/17			

Day-type	Dates							
0,1,0,0,0	10/23/09	1/29/12	2/2/14	2/27/15	4/2/17			
	12/24/09	2/10/12	2/8/14	4/24/15	4/17/17			
	12/26/09	3/31/12	3/31/14	4/26/15	6/27/17			
	1/25/10	4/3/12	4/1/14	4/28/15	9/7/17			
	2/19/10	4/10/12	4/3/14	5/25/15				
	3/20/10	5/11/12	4/15/14	6/23/15				
0,0,0,1,0	2/23/08	11/2/10	6/25/13	6/16/15	7/23/16	5/7/17		
	9/21/08	11/3/10	7/10/13	6/18/15	9/6/16	5/20/17		
	1/2/09	5/18/11	7/29/13	7/8/15	10/25/16	5/31/17		
	1/8/09	6/10/11	9/12/13	8/20/15	10/28/16	6/13/17		
	2/11/09	6/17/11	10/4/13	10/31/15	11/2/16	6/14/17		
	2/13/09	6/18/11	7/3/14	11/10/15	11/17/16	6/17/17		
	6/24/09	8/1/11	8/9/14	11/11/15	12/15/16	7/20/17		
	10/1/09	12/9/11	9/26/14	12/30/15	1/10/17	7/21/17		
	1/4/10	7/13/12	12/2/14	2/11/16	2/2/17	7/28/17		
	3/23/10	8/5/12	12/5/14	4/3/16	2/8/17	8/7/17		
	5/27/10	9/4/12	1/5/15	5/25/16	3/3/17	8/22/17		
	8/24/10	12/29/12	1/9/15	6/2/16	4/4/17	9/16/17		
	9/5/10	5/22/13	2/6/15	6/16/16	4/13/17	10/23/17		
	10/15/10	6/23/13	4/16/15	7/8/16	4/23/17	12/15/17		
0,0,0,1,1	1/5/08	6/16/08	10/2/09	11/20/10	7/19/12	11/23/13	8/15/15	8/4/17
	1/25/08	12/11/08	10/15/09	5/16/11	8/10/12	2/28/14	4/2/16	11/26/17
	2/6/08	12/16/08	11/7/09	5/19/11	10/11/12	5/13/14	10/14/16	
	5/9/08	2/17/09	7/4/10	6/11/11	12/21/12	11/30/14	2/20/17	
	5/16/08	3/29/09	8/23/10	1/20/12	6/24/13	12/3/14	8/3/17	
0,1,0,0,1	1/29/08	4/28/08	2/22/09	6/11/09	4/11/10	9/30/10	3/18/11	12/4/14
	2/4/08	12/1/08	2/23/09	10/13/09	4/26/10	10/24/10	3/24/11	
	2/18/08	12/17/08	4/15/09	12/11/09	5/14/10	12/17/10	4/28/11	
	3/19/08	12/26/08	5/22/09	1/18/10	7/29/10	12/20/10	12/10/12	
1,0,0,0,0	6/14/08	1/18/09	1/30/10	5/8/10	7/28/12	2/1/15		
	6/15/08	2/26/09	2/10/10	7/19/10	1/27/13	1/16/16		
	8/2/08	3/2/09	2/11/10	1/10/11	2/9/13	1/24/16		
	10/25/08	4/13/09	2/26/10	9/29/11	2/17/14			
0,0,1,1,0	7/6/13	10/31/14	2/19/16	3/31/17	5/1/17	7/24/17	10/19/17	
	11/17/13	12/18/15	8/13/16	4/24/17	7/10/17	8/18/17		
	2/15/14	1/22/16	3/22/17	4/25/17	7/13/17	10/9/17		
0,1,0,1,0	1/4/08	6/6/08	3/11/09	4/25/11	3/12/14	6/8/15	12/27/15	12/8/16
	3/20/08	1/15/09	12/8/09	5/6/12	4/9/15	6/17/15	8/12/16	5/3/17

Table 10: Inventory of days belonging to the top 9 most frequently occurring day-type tuples (excluding the no-outlier case).

Day-type	**0,1,1,0,0**	**1,0,1,0,0**	**1,1,1,1,1**	**0,1,0,1,1**	**0,1,1,1,0**	**0,0,1,0,1**	**1,1,0,1,1**	**1,1,1,1,0**	**0,0,1,1,1**
Dates	6/26/09 12/12/10 11/29/11 12/23/12 5/23/13 6/13/14 10/13/14 11/10/14 2/22/15 11/18/15 7/28/16 8/19/16 12/11/16 5/5/17 12/8/17	3/1/09 1/24/10 2/25/10 8/19/11 11/7/12 9/2/13 12/19/13 1/21/14 1/29/14 5/16/14 7/2/14 2/2/16 8/8/16 2/9/17 9/11/17	12/19/08 12/20/08 12/21/08 4/3/09 9/11/09 3/31/11 3/8/13 4/17/13 6/18/14 3/23/15 12/15/15 12/17/16 4/5/17 4/6/17 6/6/17	1/27/08 2/3/08 5/2/08 11/2/08 11/9/08 2/15/09 3/8/09 1/21/10 5/17/11 6/9/11 4/18/13 6/26/13 12/15/14 12/14/15	6/22/09 11/10/11 7/18/12 7/26/12 4/10/13 9/19/13 5/18/15 7/31/16 8/20/16 12/18/16 12/21/16 3/14/17 6/19/17 10/24/17	7/29/09 3/12/10 3/24/13 12/9/14 12/10/14 12/19/14 2/8/15 12/23/15 4/9/16 11/21/16 12/23/16 3/30/17 5/24/17	2/12/08 8/4/08 9/26/08 11/6/08 11/14/08 12/23/08 6/19/09 12/20/09 3/15/10 6/23/10 10/1/10 5/25/11 1/2/14	8/15/08 7/2/09 12/28/10 2/3/14 2/21/15 11/21/15 12/28/15 2/5/16 4/4/16 7/1/16 7/12/17 8/2/17 10/14/17	11/22/13 1/4/14 12/21/15 12/29/15 1/5/16 2/16/16 3/11/16 1/22/17 2/3/17 2/17/17 6/7/17 10/13/17
Day-type	**1,0,0,1,0**	**1,0,0,1,1**	**1,1,0,0,0**	**1,1,1,0,0**	**0,1,1,0,1**	**0,1,1,1,1**	**1,0,1,1,0**	**1,1,0,1,0**	**1,1,1,0,1**
Dates	8/11/08 9/9/08 4/18/09 5/9/10 12/27/10 1/19/12 1/8/15 1/25/16 12/4/16 2/25/17 7/7/17	5/12/08 7/23/08 10/28/08 12/22/08 1/19/09 9/7/11 11/12/12 6/30/14 2/14/16 11/3/17	1/11/08 3/18/08 5/27/08 6/18/08 4/6/09 4/20/09 1/3/10 1/26/11 12/25/12 6/2/17	7/24/08 8/14/08 9/21/09 3/14/10 6/25/12 8/8/14 2/2/15 3/5/15 5/10/15 12/25/16	4/4/08 3/23/13 2/21/14 3/29/14 5/8/14 3/26/16 9/30/16 12/22/16 3/6/17	12/18/08 6/22/12 2/24/16 7/21/16 10/24/16 11/22/16 12/16/16 5/25/17 12/14/17	2/13/14 8/25/14 4/20/15 6/15/15 2/15/16 4/18/16 7/25/16 9/21/16 1/7/17	6/8/08 1/10/09 2/12/09 12/1/10 5/29/11 12/8/13 1/23/16 5/27/16	3/8/08 7/27/08 8/10/08 12/24/08 2/16/10 1/6/14 5/31/15 5/4/16
Day-type	**1,0,0,0,1**	**1,1,0,0,1**	**1,0,1,0,1**	**1,0,1,1,1**					
Dates	2/22/08 11/13/09 1/18/11 1/27/11 8/14/11 10/29/11 1/1/14	1/31/08 2/13/08 11/30/08 4/17/09 6/9/09 3/13/10	7/13/08 8/22/10 10/27/10 6/1/15 2/13/17	2/3/09 1/3/14 1/5/14 2/6/17					

Table 11 Inventory of days belonging to the day-type tuples not captured in Table 10.

Disruption	**Dates**								
Outage	9/28/17 12/17/17 5/26/16 7/2/15 5/21/11 9/26/14	9/8/17 1/2/17 3/17/16 4/28/15 3/26/11 9/27/14	8/21/17 11/4/16 2/9/16 3/30/15 7/5/11 11/24/14	4/3/17 10/13/16 12/2/15 8/26/08 8/28/12 4/9/08	3/20/17 8/8/16 10/29/15 11/19/09 2/21/12	2/22/17 8/9/16 10/11/15 7/2/09 4/16/13	2/8/17 7/24/16 9/17/15 1/4/10 6/21/13	1/29/17 7/20/16 8/15/15 6/17/11 8/6/13	1/22/17 7/21/16 7/8/15 6/18/11 9/13/13
Nor'easter	2/25/10 2/8/13 1/27/15 11/12/09 3/13/10 1/13/11 3/6/13	2/26/10 2/9/13 1/28/15 11/13/09 3/14/10 10/28/11 3/7/13	2/27/10 2/10/13 1/29/15 11/14/09 3/15/10 10/29/11 3/8/13	1/30/11 2/11/13 1/30/15 12/16/09 3/16/10 10/30/11 2/7/17	1/31/11 2/11/14 1/21/16 12/17/09 12/26/10 11/8/12 2/8/17	2/1/11 2/12/14 1/22/16 12/18/09 12/27/10 11/9/12 2/9/17	2/2/11 2/13/14 1/23/16 12/19/09 12/28/10 12/26/12	2/3/11 2/14/14 1/24/16 12/20/09 1/11/11 12/27/12	2/7/13 1/26/15 1/25/16 3/12/10 1/12/11 3/5/13
Thunderstorm	6/23/14 8/20/16 6/4/17 8/18/17	6/24/14 7/12/17 10/14/14 8/20/15	5/15/15 6/18/14 4/7/17 5/25/17	5/16/15 6/19/14 4/8/17 6/1/15	8/7/14 5/12/14 4/30/17 6/14/17	8/8/14 4/9/15 6/13/14 6/14/15	8/9/14 5/20/17 6/19/17 6/15/15	7/6/16 5/21/17 7/13/17 6/25/14	6/24/15 6/3/17 7/14/17
Hurricane	10/28/12 8/28/17 9/14/08 8/28/11	10/29/12 8/29/17 9/15/08 8/29/11	10/30/12 8/30/17 10/6/16 8/31/08	10/31/12 9/9/17 10/7/16 9/1/08	11/1/12 9/10/17 10/8/16 9/2/08	8/24/17 9/11/17 10/9/16 9/3/08	8/25/17 9/12/17 8/25/11 9/4/08	8/26/17 9/12/08 8/26/11	8/27/17 9/13/08 8/27/11

Table 12 List of 178 disruption days used in the system-wide and airline-specific analysis.