A Spectral Approach Towards Analyzing Air Traffic Network Disruptions

Max Li, Karthik Gopalakrishnan, Hamsa Balakrishnan
Massachusetts Institute of Technology

Kristyn Pantoja
Texas A&M University
Motivation: Delays propagate on the air traffic network

Cancellations: Fire in Chicago Center

Delays: Power outage in Atlanta

Delays: Thunderstorms in the NE US
Large-scale disruptions are becoming more frequent
Large-scale disruptions are becoming more frequent

Need methodologies to analyze and characterize such large-scale disruptions
Historically, delays at Boston Logan International Airport (BOS) and New York LaGuardia Airport (LGA) are strongly correlated:

- Geographic proximity
- Large BOS ↔ LGA traffic flows (e.g. hourly American Airlines shuttles)

Consider two scenarios, both with a total delay of 180 min:

- \( r \approx 1 \)
  - BOS delays: 90 min
  - LGA delays: 90 min

- \( r \approx 1 \)
  - BOS delays: 10 min
  - LGA delays: 170 min
Imagine two scenarios, both with a total delay of 180 minutes:

- Some disruption, or off-nominal event, caused the delay ...
  - ... but one scenario is expected given historical correlations, while the other is unexpected
    - What traffic management initiatives should be deployed in each scenario?
    - What should the airline recovery strategies be?

Motivating example:
Spatial distribution of delays

\[ r \approx 1 \]

- LGA delays: 90 min
- BOS delays: 90 min

\[ r \approx 1 \]

- LGA delays: 170 min
- BOS delays: 10 min
• Now imagine $\binom{N}{2}$ scenarios with $N$ airports ...

• We have $\frac{N(N-1)}{2}$ scenarios similar to the prior (BOS-LGA) example
Motivating example: Characterizing specific types of disruptions

• Two disruptive events (an Atlantic hurricane vs. a nor’easter) result in similar levels of total delay
• Is the spatial distribution of delays caused by one type of disruption (e.g., a hurricane) different than the delay distribution caused by another type of disruption (e.g., a nor’easter)?
Spatial distribution of delays

• **Problem:** Is there a systematic and low-dimensional methodology for identifying unexpected spatial distributions of delay?

**Key idea in this paper**

Use *Graph Signal Processing* and *total variation* of graph signals to characterize expected vs. unexpected spatial delay distributions and relative differences.
Related work

• Models of air traffic delay dynamics
  • Queuing models [Pyrgiotis et al., 2013]
  • Network models [Gopalakrishnan et al., 2016]
  • Simulations [Ahmadbeygi et al., 2010]
  • Machine learning models [Kim et al., 2016]

• Fourier analysis applied to aviation
  • Final approach trajectory prediction (Gong & Sadovsky, 2010)
  • Aircraft trajectory analysis (Annoni & Forster, 2012)
  • Airport capacity & delay profiles; delay propagation (Diana, 2009; Yablonsky et al., 2014)

• Graph wavelet transform analysis of air traffic flows (Drew & Sheth, 2014; 2015)

• Spectral analysis of traffic networks (Crovella & Kolaczyk, 2003; Mohan et al., 2014)
Classical idea: Signals in time can be represented in terms of frequencies.
Modern interpretation: Signals on graphs can be represented in terms of “frequencies” via Laplacian eigenvalues

[Ortega 2016]
Spatial distribution of delays

• **Problem**: Is there a systematic and low-dimensional methodology for identifying unexpected spatial distributions of delay?

**Key idea in this paper**

Use **Graph Signal Processing** and **total variation** of graph signals to characterize expected vs. unexpected spatial delay distributions and relative differences.
Airport delay graph signals and correlation graphs

- Airport delay graph signal on day $l$
  - Vector $\vec{x}_l = (x_{1,l}, x_{2,l}, \ldots, x_{|\mathcal{N}|,l})^T$ where
  - $x_{i,l}$ is total delay at airport $i$ on day $l$
  - $\mathcal{N}$ is set of airports (FAA Core 30, for example)
Airport delay graph signals and correlation graphs

- Airport delay graph signal on day \( l \)
  - Vector \( \mathbf{x}_l^T = (x_{1,l}, x_{2,l}, \ldots, x_{|\mathcal{N}|,l})^T \) where
  - \( x_{i,l} \) is total delay at airport \( i \) on day \( l \)
  - \( \mathcal{N} \) is set of airports (FAA Core 30, for example)
- Underlying correlation graph, \( G = (\mathcal{N}, \mathcal{E}, A) \)
  - Undirected, weighted graph
  - \( e_{i,j} \in \mathcal{E} \) connects \((n_i, n_j) \in \mathcal{N} \times \mathcal{N} \)
  - \( T \) is the number of days (samples)
  - \((i, j)^{th}\) element in adjacency matrix \( A \) is the sample Pearson correlation coefficient:

\[
r_{x^*_i x^*_j} = \max \left( 0, \frac{\sum_{l=1}^{T}(x_{i,l}-\bar{x}_i)(x_{j,l}-\bar{x}_j)}{\sqrt{\sum_{l=1}^{T}(x_{i,l}-\bar{x}_i)^2}\sqrt{\sum_{l=1}^{T}(x_{j,l}-\bar{x}_j)^2}} \right)
\]
Total variation of a graph signal

• Given a graph Laplacian \( \mathcal{L} \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{N}|} \) and nodal graph signal \( \vec{x} \in \mathbb{R}^{|\mathcal{N}| \times 1} \), the total variation of \( \vec{x} \) on the graph is:

\[
\vec{x}^T \mathcal{L} \vec{x} = \frac{1}{2} \sum_{i \neq j} a_{ij} (x_i - x_j)^2
\]

• In our setting:
  • \( \mathcal{L} = D - A \) is the constructed graph Laplacian, given a NAS correlation graph with adjacency matrix \( A \) and degree matrix \( D = \text{diag}(\sum_j a_{1,j}, \ldots, \sum_j a_{|\mathcal{N}|,j}) \)
  • \( a_{ij} \) is the sample Pearson correlation coefficient between delays at airport \( i \) and \( j \)
  • \( x_i \) and \( x_j \) are the delays at airport \( i \) and \( j \), respectively
Total variation and signal smoothness

- Total variation (TV) is considered to be a measure of “signal smoothness”
  - Smaller the TV, smoother the signal
- Illustrative example #1: $0 < a_{ij} \approx 1$

Contribution to TV:

$a_{\text{BOS,LGA}}(x_{\text{BOS}} - x_{\text{LGA}})^2$ will be large unless delays are similar at BOS and LGA

Implications: We expect that $(x_{\text{BOS}} - x_{\text{LGA}})$ is small; a large value would be unexpected
Total variation and signal smoothness

- Total variation (TV) is considered to be a measure of “signal smoothness”
  - Smaller the TV, smoother the signal
- **Illustrative example #2**: \(0 \approx a_{ij} < 1\)

\[a_{\text{SFO}, \text{MIA}} \approx 0\]

**Contribution to TV:**
\[a_{\text{SFO}, \text{MIA}} (x_{\text{SFO}} - x_{\text{MIA}})^2\] is small even if delays are significantly different at SFO and MIA

**Implications:** We expect that \((x_{\text{SFO}} - x_{\text{MIA}})\) could be small or large; neither is unexpected
Total variation and total delay

- Total delay: $\|\vec{x}\|_1 = \sum_{i=1}^{N} x_i$
- Total variation (TV) = $\frac{1}{2} \sum_{i,j} a_{ij} (x_i - x_j)^2$
- Expect quadratic relationship between total delay and total variation
- Total variation has nice properties for detecting outlier signals on graphs*
  - The variance of TV tends to zero as its mean tends to zero
  - Mean and variance of TV are bounded
- Projection onto a low-dimensional state space to compare scale (total delay) and spatial distribution (total variation)

*Identification of outliers in graph signals. Gopalakrishnan, Li & Balakrishnan, 2019
Total variation vs. total delay

- 28 Dec 2010 [NYC holiday blizzard]
- 2 Feb 2015 [Widespread winter storms in ORD, NYC]
- 27 Jan 2011
- 3 Feb 2014
- 8 April 2016 [SFO, LAS fog + Augusta Masters]
- 17 Dec 2016 [Denver snowstorm]
Total variation vs. total delay

For a fixed value of total delay

- 27 Jan 2011
- 3 Feb 2014
- 8 April 2016 [SFO, LAS fog + Augusta Masters]
- 28 Dec 2010 [NYC holiday blizzard]
- 2 Feb 2015 [Widespread winter storms in ORD, NYC]
- 17 Dec 2016 [Denver snowstorm]
Do the impacts of different types of disruptions differ?

Can also look at the “modes” typically activated by different types of disruptions.

Outages found to activate more “typical” modes while nor’easters often activate more unusual modes.
Extensions

- Outlier identification in graph signals

*Identification of outliers in graph signals. Gopalakrishnan, Li & Balakrishnan, 2019*
<table>
<thead>
<tr>
<th>Extensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Outlier identification in graph signals</td>
</tr>
<tr>
<td>- Airline-specific spectral analyses</td>
</tr>
</tbody>
</table>

*Identification of outliers in graph signals. Gopalakrishnan, Li & Balakrishnan, 2019
Extensions

- Outlier identification in graph signals
- Airline-specific spectral and analyses
- Phase portrait analysis of recovery

Off-nominal recovery trajectories
Extensions

- Outlier identification in graph signals
- Airline-specific spectral and analyses
- Phase portrait analysis of recovery
- Expanding spectral analysis to global air traffic networks
- Learning sparser correlation networks from delay signals
Summary

- **Graph Signal Processing** and **total variation** of graph signals to characterize expected vs. unexpected spatial delay distributions and relative differences

- Results in a **projection onto a low-dimensional state space** to compare scale (total delay) and spatial distribution (total variation)

- Provides insights into how different types of disruptions differ in impact

- Several interesting extensions, including airline-specific network analysis, phase portrait analysis of post-disruption recovery, outlier identification for graph signals, etc.
Backup
### Airport outages

- Top 5 eigenvector modes of a particular day’s eigen-decomposition

<table>
<thead>
<tr>
<th>Date</th>
<th>Mode 1</th>
<th>Mode 2</th>
<th>Mode 3</th>
<th>Mode 4</th>
<th>Mode 5</th>
<th>Delay (× 10^4 min)</th>
<th>TV (× 10^6 min^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11/15/12</td>
<td>1 (93%)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.55</td>
<td>2.18</td>
</tr>
<tr>
<td>9/26/14</td>
<td>1 (73%)</td>
<td>18 (10%)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>2.33</td>
<td>25.71</td>
</tr>
<tr>
<td>9/17/15</td>
<td>1 (80%)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.73</td>
<td>8.02</td>
</tr>
<tr>
<td>7/20/16</td>
<td>1 (84%)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>2.05</td>
<td>11.26</td>
</tr>
<tr>
<td>7/21/16</td>
<td>1 (77%)</td>
<td>18 (9%)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>2.74</td>
<td>26.5</td>
</tr>
<tr>
<td>8/8/16</td>
<td>1 (70%)</td>
<td>13 (7%)</td>
<td>12 (5%)</td>
<td>–</td>
<td>–</td>
<td>2.72</td>
<td>31.35</td>
</tr>
<tr>
<td>8/9/16</td>
<td>1 (87%)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>2.23</td>
<td>6.69</td>
</tr>
<tr>
<td>1/22/17</td>
<td>1 (79%)</td>
<td>13 (8%)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>3.09</td>
<td>16.66</td>
</tr>
<tr>
<td>12/17/17</td>
<td>1 (65%)</td>
<td>13 (12%)</td>
<td>12 (10%)</td>
<td>–</td>
<td>–</td>
<td>1.64</td>
<td>14.86</td>
</tr>
</tbody>
</table>
Hurricanes and Nor’easters

- Top 5 eigenvector modes of a particular day’s eigen-decomposition